



MICHERLANIA DA SILVA NASCIMENTO

**MODELOS NÃO LINEARES NA DESCRIÇÃO DO
CRESCIMENTO DE FRUTOS DE CACAUEIROS**

LAVRAS - MG

2014

MICHERLANIA DA SILVA NASCIMENTO

**MODELOS NÃO LINEARES NA DESCRIÇÃO DO CRESCIMENTO DOS
FRUTOS DE CACAUEIROS**

Dissertação 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 Mestre.

Dr. Joel Augusto Muniz

Orientador

LAVRAS - MG

2014

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e
Serviços da Biblioteca Universitária da UFLA**

Nascimento, Micherlania da Silva.

Modelos não lineares na descrição do crescimento de frutos de
cacaueiros / Micherlania da Silva Nascimento. – Lavras : UFLA,
2014.

52p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2014.

Orientador: Joel Augusto Muniz.

Bibliografia.

1. Modelos de crescimento. 2. Medidas biométricas. 3.
Regressão não linear. 4. Aproximação linear. 5. Cacao. - Altitude. I.
Universidade Federal de Lavras. II. Título.

CDD – 519.536

MICHERLANIA DA SILVA NASCIMENTO

**MODELOS NÃO LINEARES NA DESCRIÇÃO DO
CRESCIMENTO DOS FRUTOS DE CACAUEIROS**

Dissertação 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 Mestre.

APROVADA em 26 de fevereiro de 2014.

Dr. Augusto Ramalho de Morais

UFLA

Dr. João Domingos Scalon

UFLA

Dr. Paulo Márcio Norberto

EPAMIG

Dr. Joel Augusto Muniz

Orientador

LAVRAS - MG

2014

Minha mãe, Cosma Pedro da Silva.

DEDICO.

AGRADECIMENTOS

Obrigada meu Deus, por seu amor, força e sabedoria em todos os momentos da minha vida.

A minha mãe, Cosma Pedro da Silva, por seu amor, amizade, confiança e respeito.

Aos meus irmãos, Rivânio da Silva e Regivaldo Batista da Costa.

Ao meu orientador, Joel Augusto Muniz, muito obrigada pela confiança, incentivo e disponibilidade durante os dois anos de convivência.

Aos amigos da UFLA, Allan, André, Andrezza, Carolina, Carlos, Carlos Muianga, Emanuelle, Érica, Fernando, Filipe, Joel, Laís, Leila Maria, Luciano, Marcela, Marcelo, Paulo, Rafael e Tamara pelas boas conversas, risadas, ajuda e palavras de incentivo.

Aos professores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária da Universidade Federal de Lavras, pelos conhecimentos transmitidos em cada disciplina.

À CAPES, pela concessão da bolsa de estudos.

RESUMO

A cultura do cacauzeiro é uma atividade agrícola de grande valor econômico, e o seu rendimento está diretamente ligado ao ponto de colheita. Conhecer o crescimento do fruto de cacauzeiro por meio de medidas biométricas ao longo do tempo é fundamental para compreender as fases de desenvolvimento e assim estabelecer formas adequadas de manipulação e armazenamento. Esse crescimento pode ser realizado utilizando-se modelos de regressão não linear por permitir interpretar biologicamente os fatores de desenvolvimento do vegetal. Assim, o objetivo neste trabalho é descrever e analisar o crescimento dos frutos de cacauzeiros do clone SIAL-105, por meio das medidas de comprimento, diâmetro e volume do fruto em função da idade, utilizando os modelos de regressão não lineares Logístico, Gompertz e Richards. A estimação dos parâmetros para os modelos foi realizada através do método dos mínimos quadrados e o processo iterativo de Gauss-Newton. O desvio padrão residual, critério de informação de Akaike corrigido e o coeficiente de determinação ajustado (R_{aj}^2) foram usados na comparação dos ajustes e as medidas de curvatura foram determinadas para avaliar a aproximação linear dos modelos. Dos modelos analisados o Richards apresentou ser o mais adequado para descrever o crescimento das medidas de comprimento e diâmetro, para as medidas de volume o modelo Logístico apresentou o melhor ajuste.

Palavras-chave: Modelo de crescimento. Medida biométrica. Regressão não linear. Aproximação linear. Cacau.

ABSTRACT

Cacao culture is an agricultural activity of great economic value and its yield is directly linked to the point of harvest. The knowledge of the cacao fruit growth by means of biometric measurements over time is fundamental to understand the development phases and, thus, establish adequate forms of manipulation and storage. This growth may be performed using non-linear regression models by allowing the biological interpretation of the plant's development factors. Thus, the objective in this work was to describe and analyze the growth of SIAL-105 clone cacao fruits, by means of the length, diameter and volume measurements of the fruit in relation to age, using the Logistic, Gompertz and Richards non-linear regression models. The parameter estimation for the models was performed by the least square method and the Gauss-Newton interactive process. The residual standard deviation, corrected Akaike information criteria and the adjusted coefficient of determination (R_{adj}^2) were used in the comparison of the adjustments and the curve measurements were determined to evaluate the linear approximation of the models. Of the analyzed models, the Richards showed to be the most adequate in describing the growth of the length and diameter measurements, for the volume measurements, the Logistic model presented the best adjustment.

Keywords: Growth model. Biometric measurement. Non-linear regression. Linear approximation. Cacao.

LISTA DE FIGURAS

Figura 1	Ajuste dos modelos Logístico, Gompertz e Richards aos dados de comprimento dos frutos de cacauzeiros.	33
Figura 2	Ajuste dos modelos Logístico, Gompertz e Richards aos dados dos diâmetro dos frutos de cacauzeiros.	36
Figura 3	Ajuste dos modelos Logístico e Richards aos dados de volumes dos frutos de cacauzeiros.	39

LISTA DE TABELAS

Tabela 1	Modelos de regressão não linear para descrever curvas de crescimento	23
Tabela 2	Valores de comprimento, diâmetro e volume dos frutos de cacaueteiro em função da idade do fruto.	25
Tabela 3	Estimativas para os parâmetros dos modelos Logístico, Gompertz e Richards para as médias da variável comprimento, com erros independentes, para o modelo Richards, AR(1), para os modelos Logístico e Gompertz, e variância constante.	32
Tabela 4	Medidas de curvatura e vieses de Box, em (%), para os modelos Logístico, Gompertz e Richards para a variável comprimento.	34
Tabela 5	Estimativas dos avaliadores de qualidade dos ajustes dos modelos Logístico, Gompertz e Richards para as médias da variável comprimento.	34
Tabela 6	Estimativas para os parâmetros dos modelos Logístico, Gompertz e Richards para as médias da variável diâmetro, com erros independentes, para o modelo Richards, AR(1), para os modelos Logístico e Gompertz, considerando variância constante.	35
Tabela 7	Medidas de curvatura e vieses de Box, em (%), para os modelos Logístico, Gompertz e Richards para a variável diâmetro.	37
Tabela 8	Estimativas dos avaliadores de qualidade dos ajustes dos modelos Logístico, Gompertz e Richards para as médias da variável diâmetro.	37
Tabela 9	Estimativas para os parâmetros dos modelos Logístico e Richards para as médias da variável volume, com erros independentes e variância constante.	38
Tabela 10	Medidas de curvatura e vieses de Box, em (%), para os modelos Logístico e Richards para a variável volume.	39
Tabela 11	Estimativas dos avaliadores de qualidade dos ajustes dos modelos Logístico e Richards para as médias da variável volume.	40
Tabela 12	Intervalos de 95% de confiança para os parâmetros do modelo Richards para as variáveis comprimento e diâmetro e para o modelo Logístico para a variável volume dos frutos de cacaueteiros.	40

SUMÁRIO

1	INTRODUÇÃO	11
2	REFERENCIAL TEÓRICO	12
2.1	A cultura do cacau	12
2.2	Modelos não lineares	14
2.3	Métodos de estimação	15
2.3.1	Mínimos quadrados ordinários	16
2.3.2	Mínimos quadrados generalizados	16
2.4	Métodos de iteração	18
2.5	Modelos de crescimento	19
2.6	Modelos de crescimento na fruticultura	24
3	MATERIAL E MÉTODOS	25
3.1	Ajustes e pressuposições dos modelos	26
3.2	Avaliadores da qualidade dos ajustes	27
3.2.1	Desvio padrão residual	27
3.2.2	Critério de informação de Akaike	28
3.2.3	Coefficiente de determinação ajustado (R_{aj}^2)	28
3.2.4	Medidas de curvatura	28
3.3	Viés de Box	30
3.4	Recursos computacionais	31
4	RESULTADOS E DISCUSSÃO	32
4.1	Ajuste dos modelos Logístico, Gompertz e Richards para as médias dos comprimentos	32
4.2	Ajuste dos modelos Logístico, Gompertz e Richards para as médias dos diâmetros	35
4.3	Ajuste dos modelos Logístico, Gompertz e Richards para as médias dos volumes	37
4.4	Algumas considerações sobre os ajustes dos modelos	40
5	CONCLUSÃO	42
	REFERÊNCIAS	43
	ANEXO	50

1 INTRODUÇÃO

Dentre os recursos genéticos vegetais voltados para a alimentação e agricultura, as fruteiras destacam-se pela grande diversidade e potencial que apresentam, sendo o Brasil, um dos mais importantes centros de diversidade genética de muitas espécies frutíferas tropicais (CARVALHO et al., 2009).

Na produção de um vegetal de interesse econômico como (caju, café, guaraná, cupuaçu, acerola, cacau, entre outras), de acordo com Santos et al. (2009) é importante conhecer as fases de crescimento, pois permite estabelecer formas adequadas do manejo da adubação, irrigação e planejar a época de plantio. Além disso, também permite estabelecer métodos de armazenamento para aumentar o período de conservação e a manutenção da qualidade do fruto.

Como exemplo, tem-se a cultura de cacau, que para obter frutos de qualidade é necessário cumprir as seguintes etapas: colheita, quebra, fermentação, secagem e armazenamento. Em que, o sucesso da etapa seguinte está diretamente relacionada com a eficiência da etapa anterior. Seguindo esse raciocínio, realizar a colheita dos frutos de cacueiro no estágio ideal de maturação ou o mais próximo possível é fundamental, pois a colheita do cacau sem atingir o ponto de maturação fisiológico causa uma perda no peso das amêndoas secas. Essa perda a cada 1.000 frutos colhidos verdes é de 20%, para frutos verdoengos é de 10% e não há perdas em frutos maduros. Além disso, ainda apresentam menor teor de açúcar, afetando a etapa de fermentação e conseqüentemente a qualidade do produto (MARTINS et al., 2011).

Estudos sobre o desenvolvimento dos frutos de cacauzeiros têm-se limitado a medidas biométricas sem levar em consideração a idade do fruto. Análises as medidas de crescimento do cacau considerando a idade, além de possibilitar conhecer o período de colheita, poderia fornecer meios ao estudo do peco fisiológico ou murchamento precoce dos frutos (BRITO; SILVA, 1983).

Diante do exposto, e usando a proposta dos modelos de crescimento, que usualmente são estudados por meio de regressão não linear por permitir interpretar os parâmetros biologicamente: maturidade, taxa de crescimento, por exemplo. Em modelos de regressão não linear, o interesse é relacionar uma variável resposta com outra variável, geralmente chamada explicativa. Esses modelos produzem uma curva, que inicialmente apresentam um crescimento lento, passando posteriormente a uma fase exponencial e em seguida uma estabilização do crescimento.

Assim, tem-se o objetivo neste trabalho de descrever e analisar o crescimento dos frutos de cacauzeiros do clone SIAL-105, através das medidas de comprimento, diâmetro e volume em função da idade do fruto em dias, por meio dos modelos de regressão não lineares Logístico, Gompertz e Richards.

2 REFERENCIAL TEÓRICO

Para esta seção será abordada uma breve introdução sobre os aspectos que envolvem a cultura de cacau, modelos de regressão não linear e métodos de estimação para os parâmetros que caracterizam esses modelos e uma revisão dos principais modelos de crescimento.

2.1 A cultura do cacau

O cacauzeiro é originário do continente americano e provavelmente das bacias dos rios Amazonas e Orinoco (ROSAND et al., 1975). Em que, o grupo do: *Criollo*, se desenvolveu em direção ao norte, para rio Orinoco, América Central e Sul do México, produz frutos grandes e com casca enrugada. E o *Forasteiro*, considerado o genuíno cacau brasileiro e que se diferencia por produzir frutos ovoides e de casca lisa (ROSÁRIO et al., 1978).

O cacauzeiro é uma planta umbrófila de porte arbóreo e perene, pertencente à família das Malváceas, gênero *Theobroma* (ALVERSON et al., 1999). Na literatura botânica, foi descrito inicialmente sob a denominação de *Cacao fructus*, por Charles de L'Ecluse, posteriormente designado por Carl Linnaeus como *Theobroma cacao* (ROSAND et al., 1975).

O cacauzeiro pode atingir entre 4 e 12 metros de altura (ALMEIDA; CHAVES, 2011). Apresenta boa adaptação em regiões com temperaturas médias superiores a 21° C, tolera por curto espaço de tempo, temperaturas mínimas próximas a 7° C; os solos devem ser profundos e bem drenados; a época de plantio para sementes em viveiro, setembro a abril, para mudas no campo, praticamente o ano todo; o espaçamento de plantio dependerá da fertilidade do solo e objetivos da exploração econômica, podendo variar entre 1.000 a 2.000 plantas/hectare (CENTRO DE PESQUISAS DO CACAU; COMISSÃO EXECUTIVA DO PLANO DA LAVOURA CACAUEIRA - CEPEC/CEPLAC, 2013).

Economicamente o cacau é mundialmente importante na produção de chocolate, com o processamento da polpa do fruto é possível produzir geleias, sucos, licores e sorvetes. O cacau é produzido principalmente na América Latina (Belize, México, Equador, Peru, Costa Rica e Brasil), África Ocidental (Costa do Marfim, Camarões, Gana, Nigéria e São Tomé) e Indonésia (FRANZEN; MULDER, 2007). De acordo com International Cocoa Organization - ICCO (2012), a produção mundial de cacau foi de 4,3 milhões de toneladas, com destaque para os dois maiores produtores: Costa do Marfim com 35% da produção e Gana com 24%. Em sequência estão Indonésia, Camarões e Nigéria.

O Brasil já foi um dos maiores produtores mundiais de cacau (ALMEIDA; CHAVES, 2011), atualmente sua produção é de aproximadamente 5%, o que corresponde em nível mundial, sexta posição, segundo dados da International Cocoa Organization - ICCO (2012). Em 2012, a produção brasileira foi estimada em 232.849 toneladas e os estados da Bahia (60,4%), Pará (26%) e Rondônia (7%) são os maiores produtores (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE, 2012). O Brasil começou a perder espaço em 1989, quando foi detectada a doença vassoura-de-bruxa na Bahia. Com isso o Brasil saiu da posição de produtor, com 15% da produção mundial, para de importador (MEINHARDT et al., 2008).

A vassoura-de-bruxa, causada pelo fungo *Crinipellis perniciosa*, é a principal doença do cacauzeiro (BRASIL, 1999). Recebeu esse nome porque deixa os ramos do cacauzeiro secos como uma vassoura velha (FUNDAÇÃO OSWALDO CRUZ - FIOCRUZ, 2014). O fungo contamina os novos lançamentos foliares, os frutos em desenvolvimento e as almofadas florais, podendo até provocar a morte da planta, quando afetada por sucessivos ciclos do patógeno associados a fatores abióticos (DANTAS NETO et al., 2005).

Desde a constatação da vassoura-de-bruxa algumas tentativas foram realizadas na busca de controlar o avanço da doença. Os métodos mais utilizados foram: cultural, químico, biológico e genético. O controle biológico é uma alternativa viável no manejo da doença. A utilização de micro-organismos antagonistas envolvendo diferentes mecanismos de ação como competição, antibiose e hiperparasitismo, poderia desempenhar um importante papel em um programa de manejo integrado da doença envolvendo, adicionalmente, o controle químico e cultural (OLIVEIRA; LUZ, 2005). Porém, Dantas Neto et al. (2005) afirmam que o controle químico e cultural mostraram-se ineficazes quando não executados rigorosamente de acordo com as normas técnicas da pesquisa, e antieconômicos em se tratando de lavouras formadas por variedades de alta suscetibilidade e de baixa produtividade.

Entre esses métodos o genético a longo prazo, no combate à vassoura-de-bruxa nos cacauzeiros, foi o que mostrou resultados de maior significância. Esse avanço só foi possível devido ao empenho de cacauicultores e programas de melhoramento genético. Depois de anos realizando coletas de cacauzeiros que deram origem ao Banco Ativo de Germoplasma (BAG) de cacauzeiros silvestres, que conta atualmente com 21.475 genótipos, introduzidos e mantidos em campos de pesquisa da CEPLAC (Comissão Executiva do Plano da Lavoura Cacaueira) na Amazônia brasileira e Bahia. Destes, 940 são acessos clonais e mais 877 progênies que somam 20.535 indivíduos (LUCCHESI, 1984).

Com essa medida foram desenvolvidos híbridos biclonais resultantes do

cruzamento dos clones selecionados localmente (SIC, SIAL e EEG) e seleções clonais introduzidas de outros países (SCA, UF, IMC, DR, dentre outros) (DIAS, 2001; DIAS et al., 2003). Com isso foi possível criar variedades de cacauzeiros resistentes à vassoura-de-bruxa e de alta produtividade. Em estudo elaborado por Dantas Neto et al. (2005), foi concluído que os clones de cacauzeiro SIC-864 e CNN-51 são resistentes, não apenas à vassoura-de-bruxa, como também à prodríção parda, outra doença que atinge a cultura de cacau.

2.2 Modelos não lineares

Conforme Seber e Wild (1988), em problemas de regressão, o interesse é relacionar uma variável resposta ou dependente, denotada por Y , com uma variável X geralmente chamada de variável explicativa, regressora ou independente. As variáveis explicativas podem ser aleatórias ou fixas (isto é, controlada), por exemplo, considere um experimento conduzido para avaliar a produção de trigo (Y) em diferentes níveis de densidade de plantação (X_1) e aplicação de fertilizantes (X_2). Neste caso, as variáveis X_1 e X_2 são fixas. Porém, se no momento da plantação, o pH do solo (X_3) também for medido em cada parcela, então a variável X_3 é aleatória.

A análise de regressão é uma técnica que pode ser utilizada para verificar situações em que a relação entre as variáveis de interesse é linear, porém há momentos que essa relação é de maior complexidade. Um exemplo dessa complexidade está em relacionar o crescimento de seres humanos, em que segundo Lourenço e Queiroz (2010) é caracterizado por um rápido crescimento após o nascimento, um crescimento acentuado durante a puberdade e uma estabilização em algum momento antes da idade adulta. Esse tipo de relação é de natureza não linear, com diferentes taxas de crescimento ao longo do ciclo de vida.

Modelos de regressão que descrevem processos naturais são geralmente modelos não lineares nos parâmetros (RATKOWSKY, 1993). Segundo Bates e Watts (1988) e Draper e Smith (1998) a linearidade está relacionada em verificar se as derivadas parciais são independentes de todos os parâmetros. E a não linearidade está em analisar se existe pelo menos uma das derivadas parciais dependente de algum parâmetro do modelo.

Um modelo de regressão, seja com características lineares ou não lineares, é constituído por duas partes, a primeira é determinística e a outra estocástica. A parte determinística representa a verdadeira relação entre a variável resposta e a variável explicativa, e a parte estocástica é composta pelo erro (RATKOWSKY, 1993). Para Box, Hunter e Hunter (2005), outras fontes como: a temperatura ambiente, habilidade ou agilidade das pessoas envolvidas no experimento e as boas

condições dos equipamentos utilizados podem contribuir para o erro, além dos erros de medição, análise e amostragem.

De acordo com Smyth (2002), um modelo de regressão não linear pode ser definido como:

$$Y_t = f(\mathbf{x}_t, \boldsymbol{\theta}) + \varepsilon_t, \quad t = 1, \dots, n \quad (1)$$

em que

1. Y_t representa os valores observados de cada variável resposta;
2. f é uma função conhecida do vetor das variáveis explicativas $\mathbf{x}_t = (x_{t1}, \dots, x_{tk})'$;
3. $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$ é o vetor de parâmetros desconhecidos;
4. ε_t são variáveis aleatórias não observáveis que usualmente são assumidas não correlacionadas com média zero e variância constante σ^2 .

Em modelo de regressão não linear, os estimadores de mínimos quadrados dos parâmetros não possuem as desejáveis propriedades dos modelos de regressão linear, ou seja, não-tendenciosidade, variância mínima e estimadores normalmente distribuídos. Além disso, os estimadores de parâmetros de modelos não lineares são altamente tendenciosos, possuem uma considerável flutuação sobre a variância mínima e uma distribuição significativamente assimétrica (RATKOWSKY, 1993).

Essas desejáveis propriedades, no caso não linear, somente são válidas assintoticamente, isto é, quando o tamanho da amostra é suficientemente grande. Geralmente, em pequenas amostras, essas propriedades são desconhecidas (JENN-RICH, 1969).

Como o processo de estimação em modelos não lineares consiste em uma aproximação linear, é possível, conforme relatam Bates e Watts (1988), quantificar o quanto é adequada essa aproximação, por meio de duas medidas de não linearidade, ao qual os autores denominaram como: curvatura devido ao efeito de parâmetros e curvatura intrínseca do modelo. Com isso, quanto mais próximo do linear for o comportamento de um modelo, mais precisos serão os resultados assintóticos e, conseqüentemente, mais confiáveis serão as inferências, mesmo em pequenas amostras (MAZUCHELI; ACHCAR, 2002; ZEVIANI et al., 2012).

2.3 Métodos de estimação

Em modelos lineares, para obter as estimativas dos parâmetros é utilizado o método dos mínimos quadrados, esse procedimento também pode ser utilizado

em modelos não lineares. No entanto, a minimização da soma de quadrado residual produz equações normais, que são não lineares nos parâmetros. Com isso torna-se impossível resolver as equações não lineares analiticamente. A alternativa é empregar métodos iterativos que permitem encontrar soluções analíticas aproximadas (PRAJNESHU, 1998). Dependendo da natureza do tipo da distribuição dos erros, algumas variações do método dos mínimos quadrados podem ser encontradas, como: mínimos quadrados generalizados ou ponderados, mínimos quadrados alternados e mínimos quadrados parciais (ABDI, 2007; MAZZINI et al., 2003, 2005; PEREIRA; MUNIZ; SILVA, 2005).

2.3.1 Mínimos quadrados ordinários

De acordo com Seber e Wild (1988), para realizar uma análise pelo método dos mínimos quadrados, assume-se que os ε_i são independentes e identicamente distribuídos com média zero e variância constante σ^2 . Se além disso, forem normalmente distribuídos, a estimativa de mínimos quadrados de θ , também, é o estimador de máxima verossimilhança. As estimativas de mínimos quadrados para as soluções do modelo (1) são determinadas obtendo-se a soma de quadrados dos resíduos:

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [Y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]^2. \quad (2)$$

Com a finalidade de obter os valores de θ que minimizam (2) e supondo que cada $f(\mathbf{x}_i, \boldsymbol{\theta})$ é diferenciável em relação a $\boldsymbol{\theta}$, tem-se que

$$\sum_i [Y_i - f(\mathbf{x}_i, \boldsymbol{\theta})] \frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_r} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0, \quad r = 1, \dots, p \quad (3)$$

A expressão (3) é conhecida como equações normais não lineares e deve ser resolvida para $\hat{\boldsymbol{\theta}}$.

2.3.2 Mínimos quadrados generalizados

Quando as suposições de homocedasticidade e independência dos erros forem violadas, utiliza-se o método dos mínimos quadrados generalizados. Em outras palavras, isso significa que a matriz de variância e covariância dos erros $Var(\boldsymbol{\varepsilon})$ não é uma matriz do tipo identidade ($\mathbf{I}\sigma^2$), assim os elementos fora da

diagonal principal dessa matriz são diferentes de zero.

Para casos desse tipo, Draper e Smith (1998) e Seber e Wild (1988), sugerem o modelo:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad (4)$$

em que

1. $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V}\sigma^2)$;
2. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$;
3. $Var(\boldsymbol{\varepsilon}) = \mathbf{V}\sigma^2$.

A matriz \mathbf{V} é simétrica e positiva definida e conhecida como matriz de variância e covariância. Portanto, é possível mostrar que existe uma matriz simétrica e não singular \mathbf{P} , tal que

$$\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P} = \mathbf{P}^2 = \mathbf{V}. \quad (5)$$

Pré-multiplicando o modelo (4) por \mathbf{P}^{-1} , o modelo a seguir é adquirido:

$$\mathbf{Z} = \mathbf{k}(\boldsymbol{\theta}) + \boldsymbol{\eta}, \quad (6)$$

em que $\mathbf{Z} = \mathbf{P}^{-1}\mathbf{Y}$, $\mathbf{k}(\boldsymbol{\theta}) = \mathbf{P}^{-1}\mathbf{X}\boldsymbol{\theta}$ e $\boldsymbol{\eta} = \mathbf{P}^{-1}\boldsymbol{\varepsilon}$. Logo, é fato que, se $\boldsymbol{\eta}$ é um vetor de variáveis aleatórias, então

$$\begin{aligned} E(\boldsymbol{\eta}) &= E(\mathbf{P}^{-1}\boldsymbol{\varepsilon}) \\ &= \mathbf{P}^{-1}E(\boldsymbol{\varepsilon}) \\ &= \mathbf{P}^{-1}\mathbf{0}, \end{aligned}$$

e

$$\begin{aligned} Var(\boldsymbol{\eta}) &= E(\boldsymbol{\eta}\boldsymbol{\eta}') \\ &= E(\mathbf{P}^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{P}^{-1}) \\ &= \mathbf{P}^{-1}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{P}^{-1} \\ &= \mathbf{P}^{-1}\mathbf{P}\mathbf{P}^{-1}\sigma^2 \\ &= \mathbf{I}\sigma^2. \end{aligned}$$

Sendo que, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = Var(\boldsymbol{\varepsilon}) \Rightarrow E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{V}\sigma^2 = \mathbf{P}\mathbf{P}\sigma^2$. Também é verdade que $\boldsymbol{\eta} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$, uma vez que os elementos de $\boldsymbol{\eta}$ são combinações lineares dos elementos dos $\boldsymbol{\varepsilon}$, que é normalmente distribuído.

Com essa transformação, o modelo de mínimos quadrados generalizados torna-se semelhante ao modelo de mínimos quadrados ordinários. Assim, as estimativas de mínimos quadrados generalizados que minimizam a soma dos quadrados dos erros do modelo (4) são encontradas por:

$$S(\boldsymbol{\theta}) = [\mathbf{Z} - \mathbf{k}(\boldsymbol{\theta})]' [\mathbf{Z} - \mathbf{k}(\boldsymbol{\theta})]$$

Se uma análise for realizada por métodos dos mínimos quadrados, quando deveria ser adotado os mínimos quadrados generalizados, as estimativas obtidas ainda seriam não tendenciosas, porém não teriam variância mínima (DRAPER; SMITH, 1998).

Quando a pressuposição de independência residual não for atendida, por exemplo, o modelo (4) será escrito por:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + u, \quad (7)$$

em que $u \sim N(\mathbf{0}, \mathbf{V}\sigma^2)$.

Assumindo que os erros são autocorrelacionados na forma de um processo autoregressivo estacionário de primeira ordem AR(1), tem-se que

$$u_t = \phi_1 u_{t-1} + \varepsilon_t,$$

sendo $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, $E(\varepsilon_t \varepsilon_{t-h}) = 0$, se $h \neq 0$. E o modelo u_t será considerado estacionário se o parâmetro de autocorrelação, ϕ , estiver no intervalo $-1 \leq \phi \leq 1$; $t = 2, \dots, n$, em que n é o número de observações. Assim, $\sigma_u^2 = \frac{\sigma_\varepsilon^2}{1-\phi_1^2}$ e $cov_u = \frac{\sigma_\varepsilon^2}{1-\phi_1^2} \phi_1^h$, sendo ϕ_1 o parâmetro autorregressivo de primeira ordem.

2.4 Métodos de iteração

Como foi escrito na seção anterior, métodos iterativos são necessários para alcançar soluções aproximadas das somas de quadrados residuais em modelos não lineares. Os métodos mais utilizados para determinar a estimativa do parâmetro $\boldsymbol{\theta}$ que minimize a $S(\boldsymbol{\theta})$ são: (i) Gauss-Newton, (ii) Steepest Descent e (iii) Marquardt's (DRAPER; SMITH, 1998).

O método de Gauss-Newton usa resultados da teoria dos mínimos quadrados lineares em uma sucessão de etapas. Esse método converge rapidamente, desde que, os valores iniciais tenham uma boa aproximação aos verdadeiros valores dos parâmetros. Caso contrário, a convergência pode não ocorrer (PRAJ-NESHU, 1998).

O Método de Marquardt's é o mais utilizado para encontrar estimativas de mínimos quadrados não lineares (MOTULSKY; RANSNAS, 1987). De acordo com Marquardt (1963), esse método é uma intercalação entre as teorias do método de Gauss-Newton e o método Steepest Descent. Em outras palavras, o método combina a capacidade do método Steepest Descent em convergir, a partir de uma estimativa inicial, que pode estar fora da região de convergência em outros métodos, e a qualidade do método de Gauss-Newton em convergir rapidamente quando aproxima-se dos verdadeiros valores a serem estimados. Assim, o método reuni as melhores características dos seus antecessores, evitando as limitações mais graves.

Para que os métodos iterativos alcancem a convergência, é necessário estabelecer valores iniciais para os parâmetros o mais preciso possível. Prajneshu (1998), geralmente, recomenda utilizar informações prévias, sempre que possível, de experimentos semelhantes realizados no passado e executar uma análise gráfica que também pode contribuir para obter os valores iniciais.

2.5 Modelos de crescimento

Modelos de crescimento geralmente referem-se à tentativa de descrever o comportamento de uma série de medidas de variáveis de interesse (peso, comprimento, volume, entre outras) ao longo do tempo (SANDLAND; MCGILCHRIST, 1979).

Vários modelos foram propostos para descrever o crescimento de plantas, animais e população humana. Inicialmente esses modelos foram elaborados com uma abordagem biológica. No entanto, atualmente são aplicados em diferentes áreas do conhecimento e as interpretações para os parâmetros podem variar de acordo com o objetivo em estudo. Em seguida, será descrito alguns modelos que historicamente desempenharam papéis importantes no desenvolvimento da teoria dos modelos de crescimento.

Modelo Malthus

Malthus (1798) afirmou que a população humana tinha uma tendência natural de crescer geometricamente, enquanto que a produção de alimentos cresceria em progressão aritmética, de modo que a capacidade de crescimento da população era indefinidamente maior do que a capacidade da terra de produzir meios de subsistência para o homem.

As profecias de Malthus ainda estão longe de se concretizarem, pois mesmo havendo aumento populacional, também houve aumento na produção de alimen-

tos, devido aos avanços na tecnologia e na medicina nos últimos dois séculos (HENRIQUES, 2007).

O modelo sugerido por Malthus (1798) foi desenvolvido com base na equação diferencial:

$$\frac{dN}{dt} = rN,$$

e sua solução fornece a seguinte curva

$$N(t) = N_0 e^{rt}. \quad (8)$$

em que N_0 é o valor inicial da população, r é a taxa de crescimento intrínseco. Quando $r > 0$, o modelo apresenta uma curva exponencial crescente; para $r < 0$ há uma curva exponencial decrescente e para $r = 0$ o modelo caracteriza uma reta, diante dessa circunstância a população tem taxa de natalidade e mortalidade iguais.

Modelo Logístico

Conforme Savageau (1980), o modelo logístico foi proposto pelo matemático P. F. Verhulst em 1838, e tornou-se um dos modelos mais populares para descrever o crescimento de animais, plantas e também em economia. Inicialmente foi utilizado para estimar a população americana em 1920 (TSOULARIS; WALLACE, 2001).

O modelo logístico é frequentemente escrito pela equação diferencial:

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

que apresenta a seguinte solução

$$N(t) = \frac{K}{1 - \left(1 - \frac{K}{N_0} \right) e^{(-rt)}}. \quad (9)$$

em que N_0 representa o tamanho no instante $t = 0$; K é o valor máximo atingido e r é a taxa de crescimento intrínseco. Para $t \rightarrow \infty$, $N(t) = K$, que representa o crescimento máximo. O ponto de inflexão da curva logística (no qual a taxa de crescimento é máxima), é exatamente $N = \frac{K}{2}$. Ou seja, nesse ponto já foi atingido metade do crescimento, e posteriormente, o crescimento diminuirá à medida que se aproximar de K . Esse comportamento em torno do ponto de inflexão caracteriza uma curva simétrica.

Modelo Gompertz

Este modelo foi publicado por Benjamin Gompertz em 1825, e foi designado para descrever a faixa etária da população humana (ZEIDE, 1993). Segundo Savageau (1980), é bastante utilizado para descrever o crescimento de animais, plantas, tumores, e em estudos de envelhecimento e mortalidade.

O modelo de Gompertz sugerido em 1825, foi desenvolvido com base na equação diferencial:

$$\frac{dN}{dt} = rN \left[\ln \left(\frac{K}{N} \right) \right],$$

que tem solução

$$N(t) = K e^{\ln\left(\frac{N_0}{K}\right)e^{-rt}}. \quad (10)$$

em que a curva de Gompertz apresenta os mesmos parâmetros da curva logística e as interpretações são equivalentes. Contudo, o ponto de inflexão da curva de Gompertz é $N = K e^{-1}$, significando que a taxa de crescimento é máxima antes da metade do crescimento máximo, mudando rapidamente para um crescimento mais lento. O ponto de inflexão da curva de Gompertz é menor quando comparado com a curva logística, isso a caracteriza como uma curva assimétrica.

Modelo Von Bertalanffy

Este é um dos modelos mais bem sucedidos para descrever o crescimento animal (SAVAGEAU, 1980). Para Zeide (1993), o modelo de Von Bertalanffy reside no rigor pretendido na sua fundamentação teórica. Essa teoria considera que o crescimento animal é o resultado da ação combinada de dois processos opostos, anabolismo e catabolismo. Com essas suposições, Von Bertalanffy propôs que a taxa de anabolismo é proporcional à área de superfície corporal, enquanto que o catabolismo é proporcional ao peso corporal.

O modelo sugerido por Bertalanffy (1957) é baseado na equação diferencial

$$\frac{dW}{dt} = \eta W^{\frac{2}{3}} - \kappa W.$$

em que W é o peso corporal, η e κ são as constantes de anabolismo e catabolismo respectivamente.

A equação diferencial acima descreve que as mudanças do peso corporal é

determinada pela diferença entre os processos de construção (η) e degradação (κ). A potência $2/3$, expressa que o metabolismo é proporcional à área da superfície corporal.

O modelo acima apresenta a seguinte solução

$$W(t) = \left\{ \frac{\eta}{\kappa} - \left(\frac{\eta}{\kappa} - W_0^{1/3} \right) e^{-\kappa t/3} \right\}^3, \quad (11)$$

em que W_0 é o peso corporal no instante $t = 0$.

Modelo Richards

O modelo proposto por Richards (1959) é valorizado pela sua precisão e é usado mais do que qualquer outro modelo em estudo de crescimento de plantas. Foi derivado do modelo de Von Bertalanffy, porém a interpretação biológica proposta por Von Bertalanffy foi dispensada no modelo de Richards (ZEIDE, 1993). Com as modificações, o modelo que antes era restrito pelo valor 3, no caso Von Bertalanffy, pode assumir qualquer valor na curva de Richards.

O modelo de Richards (1959) foi desenvolvido com base na equação diferencial:

$$\frac{dN}{dt} = rN \left[1 - \left(\frac{N}{K} \right)^\beta \right],$$

que tem solução

$$N(t) = \frac{K}{\left\{ 1 + \left[\left(\frac{K}{N_0} \right)^\beta - 1 \right] e^{-\beta r t} \right\}^{1/\beta}}. \quad (12)$$

em que, com exceção de β , que é o parâmetro de inflexão, os demais parâmetros apresentam a mesma interpretação biológica dos modelos Logístico e Gompertz.

Segundo Draper e Smith (1998) modelos não lineares são bastante utilizados para descrever o comportamento do crescimento ao longo do tempo. Para Sandland e McGilchrist (1979), vários dos modelos desenvolvidos para descrever o crescimento possui fundamentação biológica e são baseados em equações diferenciais determinísticas. Porém, ao estudar dados de crescimento deve-se levar em consideração que existe uma estrutura de dependência entre as observações, influência dos fatores ambientais, bem como a capacidade intrínseca de cada indivíduo (plantas, animais, entre outros) em adaptar-se ao ambiente inserido.

Levando essas hipóteses em consideração, Prajneshu (1998) afirma que

modelos na forma determinística não representam de maneira eficiente estudos de dados de crescimento, sendo necessário substituí-los por modelos estatísticos, possibilitando assim adquirir resultados mais precisos.

Seguindo esse raciocínio, os modelos Logístico, Gompertz e Richards podem ser reescritos como modelos estatísticos adicionando-se ao lado direito de cada expressão o termo aleatório (ε_t). Com isso, modelos antes determinísticos passam a ter uma abordagem estatística, conforme apresentados na tabela (1).

Tabela 1 Modelos de regressão não linear para descrever curvas de crescimento

Modelos	Modelos Estatísticos
Logístico	$N(t) = \frac{K}{1 - \left(1 - \frac{K}{N_0}\right) e^{(-rt)}} + \varepsilon_t$
Gompertz	$N(t) = K e^{\ln\left(\frac{N_0}{K}\right) e^{(-rt)}} + \varepsilon_t$
Richards	$N(t) = \frac{K}{\left\{1 + \left[\left(\frac{K}{N_0}\right)^\beta - 1\right] e^{-\beta rt}\right\}^{1/\beta}} + \varepsilon_t$

Um problema que ocorre nos ajustes de modelos não lineares, como os apresentados na tabela (1), é a provável presença de autocorrelação entre os erros. Além desse problema, também é possível que os erros não possuam variância constante (GUEDES et al., 2004; MENDES et al., 2008; MUNIZ; SAVIAN; SCALON, 2008; SAVIAN; MUNIZ, 2007).

Uma aplicação que leva em consideração os erros autocorrelacionados, pode ser vista em estudos de Prado, Savian e Muniz (2013), os quais compararam os modelos Logístico e Gompertz no desenvolvimento de frutos de coqueiro da variedade anão verde, baseados nos dados de diâmetro externos longitudinal e transversal. Os autores concluíram que o modelo logístico foi o mais adequado na descrição do crescimento do fruto e ao ajustar as variáveis diâmetro externo longitudinal, segundo um processo autorregressivo de primeira ordem, e diâmetro externo transversal, considerando um processo autorregressivo de segunda ordem, as estimativas dos parâmetros foram mais precisas.

Ao analisar dados de crescimento dos frutos do cafeeiro, cultivar Obatã IAC 1669-20, Fernandes (2012) comparou os modelos logístico e gompertz, considerando o acúmulo de massa fresca em diferentes alinhamentos de plantio em diferentes faces de incidência solar da planta (norte e sul). E concluiu que o modelo Gompertz adotando erros autorregressivos de primeira ordem e ponderação

da variância, foi o melhor para descrever o crescimento dos frutos do cafeeiro cultivar Obatã IAC 1669-20.

2.6 Modelos de crescimento na fruticultura

Vários estudos foram realizados utilizando modelos de crescimento para descrever crescimento de frutos, a importância dos modelos de crescimento dar-se por permitir quantificar o desenvolvimento da planta, estimar o ponto de colheita e com isso é possível aplicar práticas culturais adequadas: semeadura, espaçamento, irrigação, fertilização, seleção de variedades, clima, dentre outros (BRUNA, 2007; CARVALHO et al., 2009; SENHOR et al., 2009).

Em estudos de crescimento dos frutos de tamareira-anã, Terra, Muniz e Savian (2010) compararam os modelos Logístico e Gompertz, e concluíram que o modelo Logístico foi o mais adequado para descrever as medidas de comprimento e diâmetro dos frutos de tamareira-anã. Em crescimento do fruto e sementes dos clones de cacau: SIAL-169, SIAL-505, SIC-328, ICS-8 e o cultivar Catongo, Almeida e Valle (1995) utilizaram o modelo Richards para determinar as variações nas taxas de acúmulo de matéria seca.

Maia et al. (2009), ao estudar duas variedades de bananeiras submetidas a cinco doses de um regulador de crescimento ao longo do tempo, comparou os modelos Brody, Gompertz, Logístico e Bertalanffy e concluiu que dentre os modelos, o Logístico foi o melhor para descrever curvas de crescimento de bananeiras. Ao estimar o desempenho germinativo de sementes de três lotes comerciais de milho híbrido OC 705, Gazola et al. (2011) utilizou o modelo Logístico para ajustar os dados e para avaliar o diagnóstico do modelo usaram o viés de Box e as medidas de curvatura de Bates e Watts (1988).

3 MATERIAL E MÉTODOS

Os dados utilizados neste trabalho foram extraídos de Brito e Silva (1983). Os frutos em estudo procederam de cacauzeiros sombreados do clone SIAL-105, que estavam dispostos em fileiras no Jardim Clonal da Quadra E, do Centro de Pesquisa do Cacau, em Ilhéus, Bahia, de coordenadas geográficas $14^{\circ}47'S$, $39^{\circ}16'W$, a uma altitude média de 38 m. Os cacauzeiros vegetavam em terrenos de encosta tipo alfisol da série germoplasma, considerado um bom solo para o cultivo do cacau.

Para a obtenção dos frutos com idade conhecida, as 23 árvores da fileira foram polinizadas artificialmente, tomando-se o pólen do clone EEG-9, da fileira vizinha.

A primeira coleta dos frutos iniciou-se com 31 dias após a polinização, e repetiu-se a cada 15 dias, até completar 180 dias. Como a princípio os frutos eram pequenos, na primeira coleta foram tomados 50 frutos. Posteriormente, com o aumento do tamanho dos frutos e por escassez de material, tomaram-se 40 frutos na segunda coleta, 30 frutos, na terceira e na quarta, 20, na quinta, sexta, sétima e oitava coletas, 15 frutos, na nona e décima e 16 frutos na décima primeira. Os frutos, após coleta, eram colocados em sacos plásticos e imediatamente levados ao laboratório, onde eram realizadas as medidas de crescimento. O comprimento e diâmetro foram tomados com o auxílio de um paquímetro e o volume, através do deslocamento do nível de água da coluna líquida. Após adquiridas as medidas de comprimento, diâmetro e volume de cada coleta, foram obtidas as médias, totalizando 11 observações, como mostra a tabela a seguir.

Tabela 2 Valores de comprimento, diâmetro e volume dos frutos de cacauzeiro em função da idade do fruto.

Comprimento (cm)	Diâmetro (cm)	Volume (cm ³)	Idade do Fruto (dias)
1,9227	0,8227	0,8620	31
3,0407	1,3252	2,8187	45
4,4977	1,9405	7,9067	60
5,5155	2,5157	17,7699	75
8,0607	4,0267	69,7000	90
10,0102	6,0032	191,5000	105
10,1610	7,1187	288,0000	120
10,6085	8,2632	395,7500	134
10,6086	8,5640	421,0000	150
10,6440	8,4937	410,0000	165
10,3681	8,2975	365,0800	180

3.1 Ajustes e pressuposições dos modelos

Para as medidas biométricas (comprimento, diâmetro e volume) dos caueiros do clone SIAL-105, foram ajustados os modelos não lineares Logístico, Gompertz e Richards, com tempos, $t = 31, 45, 60, \dots, 180$ dias. Os modelos com erros independentes estão descritos na tabela 1. A seguir será apresentado as expressões considerando erros autorregressivos de primeira ordem.

Modelo Logístico com erros autorregressivos de primeira ordem

$$N(t) = \frac{K}{1 - \left(1 - \frac{K}{N_0}\right) e^{(-rt)}} + \phi u_{t-1} + \varepsilon_t$$

Modelo Gompertz com erros autorregressivos de primeira ordem

$$N(t) = K e^{\ln\left(\frac{N_0}{K}\right) e^{-rt}} + \phi u_{t-1} + \varepsilon_t$$

Modelo Richards com erros autorregressivos de primeira ordem

$$N(t) = \frac{K}{\left\{1 + \left[\left(\frac{K}{N_0}\right)^\beta - 1\right] e^{-\beta rt}\right\}^{\frac{1}{\beta}}} + \phi u_{t-1} + \varepsilon_t$$

em que:

1. $N(t)$ expressa os valores do comprimento, diâmetro e volume no tempo t ;
2. K corresponde ao valor máximo atingido da variável dependente ao longo do tempo;
3. N_0 corresponde aos valores iniciais do comprimento, diâmetro e volume;
4. r é a taxa de crescimento;
5. β parâmetro sem interpretação biológica;
6. u_{t-1} e ε_t correspondem ao resíduo no tempo $t - 1$ e o ruído branco, respectivamente;

7. ϕ é o parâmetro autorregressivo de primeira ordem.

Para verificar a presença de heterogeneidade, foi aplicado o teste de Breusch e Pagan (1979), em que a hipótese nula considera os erros homocedásticos. Utilizando o valor-p como critério de decisão, tem-se que para valores menores ou iguais ao nível de significância adotado, rejeita-se a hipótese nula. Quando a presença de heterogeneidade é comprovada pelo teste, deve-se ajustar os modelos aos dados usando ponderação de variância, por meio dos mínimos quadrados generalizados.

A presença de autocorrelação residual foi comprovada pelo teste de Durbin e Watson (1950), em que sua hipótese nula equivale a não existência de correlação serial dos resíduos, e para valores menores ou iguais ao nível de significância, rejeita-se a hipótese nula. Para constatar a pressuposição de normalidade dos resíduos foi utilizado o teste de Shapiro-Wilk (SHAPIRO; WILK, 1965).

3.2 Avaliadores da qualidade dos ajustes

Ao realizar o ajuste será necessário determinar dentre os modelos em estudo, o mais adequado para descrever o crescimento dos frutos em análise. Essa verificação é possível através de procedimentos gráficos, porém utilizar medidas que quantifiquem a qualidade do ajuste torna-se mais informativo. Logo, para selecionar o modelo foi considerado a interpretação biológica dos parâmetros, e a precisão dos ajustes de acordo com os critérios a seguir.

3.2.1 Desvio padrão residual

O desvio padrão residual é obtido pela seguinte expressão:

$$DPR = \sqrt{\frac{QME}{(n - p)}}$$

em que, QME é o quadrado médio do erro, n o número de observações e p o número de parâmetros do modelo.

Quanto menor o valor do desvio padrão residual, melhor será o modelo ajustado.

3.2.2 Critério de informação de Akaike

O critério de informação de Akaike (AIC) foi proposto por Akaike (1974). Essa medida é frequentemente utilizada para selecionar modelos em diversas áreas e de acordo com esse critério, o melhor modelo será aquele que apresentar o menor valor de AIC. Segundo Burnham e Anderson (2002), para pequenas amostras deve-se utilizar o AIC Corrigido (AICc).

O critério de informação de Akaike corrigido é dado por:

$$AICc = -2\log(L(\hat{\theta})) + 2p + \frac{2p(p+1)}{n-p-1},$$

em que

1. $L(\hat{\theta})$ é a função de máxima verossimilhança do modelo;
2. o número de parâmetros do modelo;
3. o número de observações.

3.2.3 Coeficiente de determinação ajustado (R_{aj}^2)

$$R_{aj}^2 = 1 - \left[\frac{(1 - R^2)(n - 1)}{n - p} \right]$$

em que, $R^2 = 1 - \frac{SQR}{SQT}$ é o coeficiente de determinação, SQR a soma de quadrado de resíduos e SQT a soma de quadrado total.

3.2.4 Medidas de curvatura

Usando os conceitos de geometria diferencial, Bates e Watts (1988) desenvolveram medidas de não linearidade baseadas na noção de curvatura. Essas medidas independem de mudanças de escala dos parâmetros ou dos dados, podendo ser usadas para comparar diferentes conjuntos de dados como também diferentes parametrizações (SEBER; WILD, 1988; SOUZA et al., 2010).

Por expansão de série de Taylor, numa vizinhança de $\hat{\theta}$, tem-se

$$f(x_t, \boldsymbol{\theta}) = f(x_t, \hat{\boldsymbol{\theta}}) + \underbrace{\frac{\partial f(x_t, \hat{\boldsymbol{\theta}})}{\partial \theta_r}}_{\mathbf{F}} \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}' \underbrace{\frac{\partial^2 f(x_t, \hat{\boldsymbol{\theta}})}{\partial \theta_r \theta_s}}_{\mathbf{H}} \boldsymbol{\delta}; \quad r, s = 1, \dots, p \quad (13)$$

em que, $\boldsymbol{\delta} = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$, \mathbf{F} é a matriz de derivadas parciais de primeira ordem de dimensões $n \times p$ e \mathbf{H} é a matriz $n \times p \times p$, sendo que cada face de \mathbf{H} é uma matriz de derivadas parciais de segunda ordem de dimensões $p \times p$.

Reescrevendo (13), segue que

$$f(\boldsymbol{\theta}) = f(\hat{\boldsymbol{\theta}}) + \mathbf{F}\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}'\mathbf{H}\boldsymbol{\delta}. \quad (14)$$

O grau de linearização de $f(\boldsymbol{\theta})$ é medido comparando o tamanho do vetor $\boldsymbol{\delta}'\mathbf{H}\boldsymbol{\delta}_{n \times 1}$ relativamente ao termo linear $\mathbf{F}\boldsymbol{\delta}$ (BATES; WATTS, 1988; SOUZA, 1998).

Para realizar essa comparação Bates e Watts (1988) dividiram o termo $\boldsymbol{\delta}'\mathbf{H}\boldsymbol{\delta}$ em duas componentes ortogonais, projeção sobre o plano tangente e normal ao plano tangente. Essa decomposição pode ser realizada utilizando-se a matriz de projeção

$$\mathbf{P} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'.$$

Assim, as duas componentes são dadas por:

$$\mathbf{H} = \mathbf{H}^T + \mathbf{H}^N,$$

em que $\mathbf{H}^T = \mathbf{P}\boldsymbol{\delta}'\mathbf{H}\boldsymbol{\delta}$ é a componente tangencial e $\mathbf{H}^N = (\mathbf{I} - \mathbf{P})\boldsymbol{\delta}'\mathbf{H}\boldsymbol{\delta}$ é a componente normal. Com base nessa decomposição, Watts e Bates (1980) definiram as medidas como:

$$K^T = \frac{\|\mathbf{H}^T\|}{\|\mathbf{F}\boldsymbol{\delta}\|^2} \quad (15)$$

e

$$K^N = \frac{\|\mathbf{H}^N\|}{\|\mathbf{F}\boldsymbol{\delta}\|^2} \quad (16)$$

em que K^T é o efeito de curvatura de parâmetros e K^N é o efeito de curvatura intrínseca, com o propósito de tornar as medidas invariantes de escala, os autores definiram as medidas:

$$\gamma^T = \rho K^T \quad e \quad \gamma^N = \rho K^N$$

em que $\rho = s\sqrt{p}$, com p o número de parâmetros e $s = \left\{ \frac{\sum_{i=1}^n [Y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]^2}{n-p} \right\}^{1/2}$.

Como critério para determinar o quanto o modelo ajustado apresenta curvatura significativa, Watts e Bates (1980) sugere usar

$$\gamma^T > 2F^{-1/2} \quad e \quad \gamma^N > 2F^{-1/2}$$

em que F é o quantil $(1 - \alpha)$ de uma distribuição F com p e $n - p$ graus de liberdade. Segundo Souza (1998), quanto mais próximos de zero forem os valores das medidas de curvatura γ^T e γ^N , mais validade terão os resultados inferenciais associados ao modelo não linear.

3.3 Viés de Box

Na maioria dos modelos não lineares de interesse prático, a medida de curvatura intrínseca geralmente é pequena. Isso significa que, se um modelo não linear apresentar um comportamento distante do linear, essa não linearidade é devida, principalmente, ao efeito de parâmetros (MAZUCHELI; ACHCAR, 2002).

Uma maneira de verificar quais os parâmetros que apresentam comportamento distante do linear, é utilizar a medida proposta por Box (1971), conhecida como viés de Box, que é estimado por:

$$Viés(\hat{\boldsymbol{\theta}}) = -\frac{\sigma^2}{2} \left(\sum_{i=1}^n \mathbf{F}\mathbf{F}' \right)^{-1} \sum_{i=1}^n \mathbf{F} \operatorname{tr} \left[\left(\sum_{i=1}^n \mathbf{F}\mathbf{F}' \right)^{-1} \right] \mathbf{H}$$

em que, \mathbf{F} é o vetor $p \times 1$ de primeiras derivadas do modelo e \mathbf{H} é a matriz $p \times p$ de segundas derivadas, ambos com relação a cada elemento de $\boldsymbol{\theta}$. Geralmente, o valor da estimativa do viés de Box é expresso em porcentagem, isto é,

$$\%Viés(\hat{\boldsymbol{\theta}}) = \frac{100Viés(\hat{\boldsymbol{\theta}})}{\hat{\boldsymbol{\theta}}}$$

Como critério de decisão Box (1971) e Seber e Wild (1988) sugerem que vieses acima de 1%, em valor absoluto, indicam comportamento não linear. Uma vez conhecidos os parâmetros responsáveis pelo comportamento distante do linear, pode-se buscar por uma reparametrização que possa reduzir a não linearidade (MAZUCHELI; ACHCAR, 2002).

Os intervalos de confiança para os parâmetros dos modelos em estudo foram construídos usando a aproximação linear e a pressuposição de normalidade assintótica, segundo Benšić (1997) e Seber e Wild (1988), tem-se que uma aproximação da matriz de variância e covariância é dada por

$$\hat{V} = s^2 [\mathbf{F}'(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta})]^{-1}$$

em que $s^2 = \frac{\sum_{i=1}^n [Y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]^2}{n-p}$, \mathbf{F} é matriz de derivadas parciais do modelo. Assim, o intervalo de confiança para θ_j , $j = 1, \dots, p$, é determinado por:

$$IC(\boldsymbol{\theta}) = \hat{\theta}_j \pm t_{\frac{\alpha}{2}, (n-p)} \sqrt{\hat{V}_{jj}}$$

em que, \hat{V}_{jj} é o j -ésimo elemento da diagonal de \hat{V} e $t_{\frac{\alpha}{2}, (n-p)}$ é o quantil superior da distribuição t de Student com $n - p$ graus de liberdade.

3.4 Recursos computacionais

Todas as análises, para o presente trabalho, foram realizadas utilizando o programa R (R DEVELOPMENT CORE TEAM, 2013). Para estimar os parâmetros dos modelos, foram empregadas as funções *nls* do pacote *stats* que utiliza o método dos mínimos quadrados ordinários, e *gnls* do pacote *nlme*, que utiliza o método dos mínimos quadrados generalizados.

Os valores iniciais necessários para determinar as estimativas dos parâmetros, foram obtidos através de informações do conjunto de dados, uma vez que a parametrização dos modelos permite conhecer o valor inicial das variáveis comprimento, diâmetro e volume, parâmetro N_0 , bem como o valor máximo dessas variáveis, parâmetro K . Para os parâmetros r e β , os valores iniciais foram obtidos por meio da função *manipulate*, um método gráfico iterativo, disponível no RStudio, editor para *script R*. O método de iteração, Gauss-Newton, foi utilizado para determinar as soluções aproximadas das somas de quadrados dos modelos em estudo.

4 RESULTADOS E DISCUSSÃO

Foram ajustados os modelos Logístico, Gompertz e Richards aos dados de comprimento, diâmetro e volume dos frutos de cacauzeiros, em seguida foram extraídos os resíduos de cada modelo e aplicado os testes de Shapiro-Wilk, Breusch-Pagan e Durbin-Watson para verificar as pressuposições de normalidade, variância constante e independência, respectivamente. Como critério de decisão foi utilizado o valor-p ao nível de $\alpha = 5\%$ de significância.

4.1 Ajuste dos modelos Logístico, Gompertz e Richards para as médias dos comprimentos

Os modelos Logístico, Gompertz e Richards foram ajustados aos dados de comprimento dos frutos de cacauzeiros, e de acordo com a tabela 2, o comprimento do fruto aumentou até os 105 dias, passando para um crescimento mais lento até os 150 dias, o fruto atinge o comprimento máximo aos 165 dias, a partir desse período começa a decrescer.

Aplicando o teste Shapiro-Wilk, verificou-se que para modelos Logístico, Gompertz e Richards a pressuposição de normalidade residual foi atendida, com valor-p iguais a 0,565, 0,991 e 0,208, respectivamente. Por meio do teste Breusch-Pagan, foi confirmada a hipótese de variâncias constantes, com valor-p 0,287, 0,414 e 0,448. A autocorrelação residual foi analisada com base no teste Durbin-Watson, e os resultados indicam adotar processo autorregressivo de primeira ordem AR(1), para os modelos Logístico e Gompertz, com valor-p iguais a 0,0455 e 0,0058.

Na tabela 3 estão apresentadas as estimativas dos parâmetros dos modelos, com erros independentes, para o modelo Richards, e erros autorregressivos de primeira ordem, para os modelos Logístico e Gompertz, considerando variância constante.

Tabela 3 Estimativas para os parâmetros dos modelos Logístico, Gompertz e Richards para as médias da variável comprimento, com erros independentes, para o modelo Richards, AR(1), para os modelos Logístico e Gompertz, e variância constante.

Modelos	Parâmetros				
	K	N_0	r	β	ϕ
Logístico	10,8597	0,4307	0,0470	–	0,1566
Gompertz	11,2000	0,0640	0,0300	–	0,3619
Richards	10,4986	1,0728	0,0225	7,8590	–

Com exceção das estimativas dos parâmetros N_0 para o modelo Gompertz e β para o modelo Richards, todas as demais foram significativas pelo teste t ao nível de 5% de significância. O ponto de inflexão para o modelo logístico ocorre quando o fruto atinge metade do seu crescimento, ou seja, $\frac{K}{2} \approx 5,432$ no tempo $\frac{\ln(\frac{K}{N_0}-1)}{r} \approx 67,82$ dias. Para o modelo Gompertz, o ponto de inflexão ocorreu em $\frac{K}{e} \approx 4,12$ no tempo $\frac{\ln(-\ln(\frac{N_0}{K}))}{r} \approx 55$ dias. E o modelo Richards atingiu ponto de inflexão em $K(\frac{1}{1+\beta})^{\frac{1}{\beta}} \approx 7,95$ entre 75 e 90 dias.

Considerando os modelos ajustados aos dados de comprimento dos frutos de cacueiro, na figura 1, e os valores das estimativas dos parâmetros da tabela 3, há evidências que o modelo Richards é o melhor para descrever os dados observados.

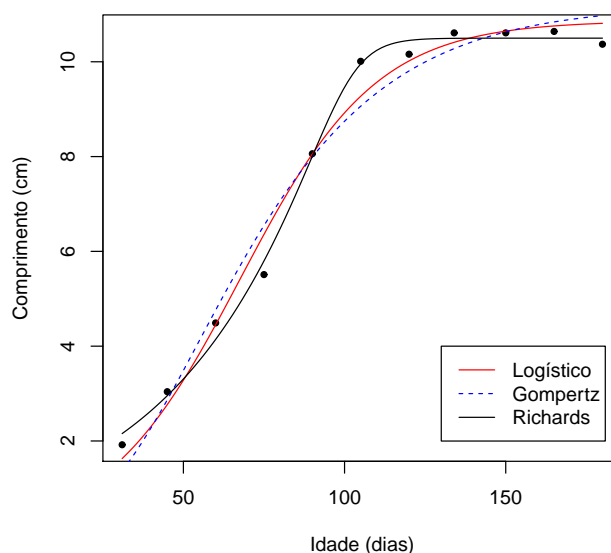


Figura 1 Ajuste dos modelos Logístico, Gompertz e Richards aos dados de comprimento dos frutos de cacueiros.

Para avaliar a qualidade das inferências, foram utilizadas as medidas de não linearidade e a medida do viés de Box, como pode ser verificado na Tabela a seguir.

Tabela 4 Medidas de curvatura e vieses de Box, em (%), para os modelos Logístico, Gompertz e Richards para a variável comprimento.

Modelos	Medidas de Curvatura		Viés de Box			
	γ^T	γ^N	K	N_0	r	β
Logístico	1,976	0,1540	0,328	5,643	1,745	–
Gompertz	7,023	0,1719	0,742	123,286	2,095	–
Richards	3,483	0,5000	0,263	3,325	5,116	79,361

Utilizando o critério sugerido por Bates e Watts (1988), para determinar o quanto os modelos ajustados apresentam curvatura significativa, tem-se o valor de $2F^{-1/2}$ ao nível de 5% de significância e com 3 e 8 graus de liberdade, calculado para os modelos Logístico e Gompertz foi de 1,098 e para o modelo Richards de 1,053, com 4 e 7 graus de liberdade. Com isso, observa-se que o efeito devido a não linearidade intrínseca foram não significativos, para os três modelos, enquanto que a não linearidade devido ao efeito de parâmetros foi significativa, destaque para o modelo Gompertz.

Quando o efeito de curvatura de parâmetro é significativo, de acordo com Mazucheli e Achcar (2002), a medida de vício de Box pode ser utilizada para revelar quais parâmetros são os maiores responsáveis pelo comportamento distante do comportamento linear. Como verificado na tabela 4, as medidas dos vieses de Box para o modelo Gompertz, apresentou um valor muito alto para o parâmetro N_0 , possivelmente uma explicação para o mal ajuste apresentado pelo modelo aos dados, o parâmetro β , para o modelo Richards, também indicou valor alto. Portanto, possivelmente uma reparametrização nesses parâmetros possa reduzir o viés.

Para selecionar o modelo mais adequado na descrição da variável comprimento dos frutos de cacueiro, foi utilizado o critério de informação de Akaike corrigido, desvio padrão residual e coeficiente de determinação ajustado. Os resultados encontram-se na tabela 5.

Tabela 5 Estimativas dos avaliadores de qualidade dos ajustes dos modelos Logístico, Gompertz e Richards para as médias da variável comprimento.

Modelos	AIC_c	DPR	(R_{aj}^2)
Logístico	23,8136	0,4600	0,9825
Gompertz	29,1726	0,6186	0,9669
Richards	11,9776	0,2451	0,9947

Com base nos critérios da tabela acima, o modelo Richards é o mais adequado para descrever os dados de comprimento dos frutos de cacueiros.

4.2 Ajuste dos modelos Logístico, Gompertz e Richards para as médias dos diâmetros

Para descrever o comportamento da variável diâmetro, os modelos Logístico, Gompertz e Richards foram ajustados aos dados da tabela 2. Os diâmetros dos frutos de cacauzeiros cresceram acentuadamente até os 150 dias, a partir dessa idade o diâmetro diminui.

Analisando as pressuposições dos modelos para a variável diâmetro, verificou-se que para os modelos Logístico, Gompertz e Richards a hipótese de normalidade residual foi aceita pelo teste Shapiro-Wilk, com valor-p de 0,184, 0,153 e 0,322. Pelo teste Breusch-Pagan, a hipótese de variância constante também foi aceita com valor-p iguais a 0,535, 0,809 e 0,256. Para os modelos Logístico e Gompertz, a pressuposição de independência não foi atendida, pelo teste Durbin-Watson, com valor-p de 0,0024 e 0,0002. Para o modelo Richards, esse valor foi de 0,398.

As estimativas dos parâmetros para os três modelos para a variável diâmetro dos frutos, com erros independentes, para o modelo Richards, e erros autorregressivos de primeira ordem, para os modelos Logístico e Gompertz, considerando variância constante, encontram-se na tabela 6.

Tabela 6 Estimativas para os parâmetros dos modelos Logístico, Gompertz e Richards para as médias da variável diâmetro, com erros independentes, para o modelo Richards, AR(1), para os modelos Logístico e Gompertz, considerando variância constante.

Modelos	Parâmetros				
	K	N_0	r	β	ϕ
Logístico	8,2423	0,0285	0,0591	—	0,9063
Gompertz	9,4232	0,0016	0,0275	—	0,9100
Richards	8,4836	0,3628	0,1023	3,7864	—

O parâmetro N_0 , para os modelos Logístico e Gompertz, não foi significativo pelo teste t ao nível de 5% de significância. Além disso, os valores das estimativas do parâmetro N_0 convergiram para valores distantes, quando comparados com os valores observados da tabela 2. Os modelos Richards e Logístico apresentaram, para o parâmetro K , estimativas próximas do diâmetro máximo atingido pelo fruto. Com relação ao ponto de inflexão, para o modelo logístico, obteve-se $\frac{K}{2} \approx 4,4256$ no tempo $\frac{\ln\left(\frac{K}{N_0} - 1\right)}{r} \approx 90,64$ dias. Para o modelo Gompertz, tem-se que $\frac{K}{e} \approx 3,49$ no tempo $\frac{\ln\left(-\ln\left(\frac{N_0}{K}\right)\right)}{r} \approx 78,61$ dias. E o modelo Richards atinge

ponto de inflexão em $K \left(\frac{1}{1+\beta} \right)^{\frac{1}{\beta}} \approx 5,61$ entre 90 e 105 dias.

Os modelos ajustados às médias dos diâmetros encontram-se na figura 2. Há evidências de um bom ajuste para a curva do modelo Richards, enquanto o modelo Gompertz aparenta não ser um modelo adequado para descrever os dados dos diâmetros dos frutos.

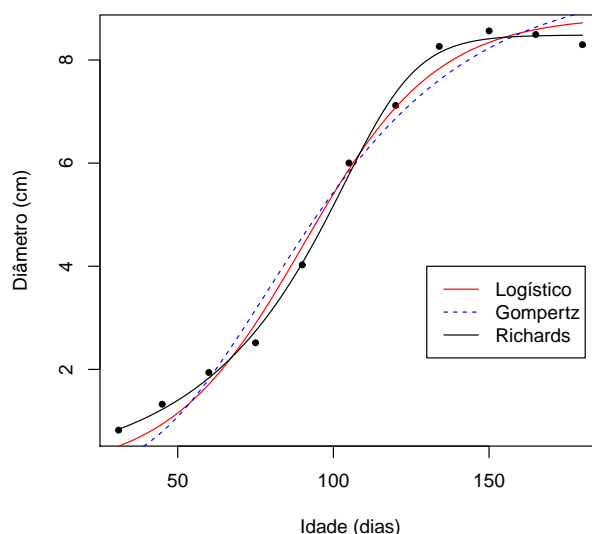


Figura 2 Ajuste dos modelos Logístico, Gompertz e Richards aos dados dos diâmetro dos frutos de cacauzeiros.

As medidas de curvatura e os vieses de Box para a variável diâmetro estão descritos na tabela 7. Mais uma vez, quando comparados com os valores de $2F^{-1/2}$ ao nível de 5% de significância e com 3 e 8 graus de liberdade, calculado para os modelos Logístico e Gompertz foi de 1,098 e para o modelo Richards de 1,053, com 4 e 7 graus de liberdade, o efeito devido a não linearidade intrínseca não foi significativo, para os três modelos, enquanto que a não linearidade devido ao efeito de parâmetros foi significativa, principalmente para o modelo Gompertz, possivelmente uma explicação do ajuste inadequado desse modelo para a variável diâmetro.

Também verifica-se que o modelo Gompertz apresentou um valor bastante elevado para o viés do parâmetro N_0 , para o modelo Richards, os vieses dos parâmetros K e N_0 forneceram os menores valores e uma possível explicação para as boas estimativas apresentadas para esses parâmetros.

Tabela 7 Medidas de curvatura e vieses de Box, em (%), para os modelos Logístico, Gompertz e Richards para a variável diâmetro.

Modelos	Medidas de Curvatura		Viés de Box			
	γ^T	γ^N	K	N_0	r	β
Logístico	3,045	0,150	0,552	12,41	2,162	—
Gompertz	19,78	0,272	1,782	624,165	3,664	—
Richards	3,288	0,281	0,500	1,864	30,883	33,24

Para avaliar a qualidade dos ajustes dos modelos aos dados de diâmetros de cacauzeiros, foram utilizados o critério de informação de Akaike corrigido, o desvio padrão residual e o coeficiente de determinação ajustado. Os resultados estão na tabela 8.

Tabela 8 Estimativas dos avaliadores de qualidade dos ajustes dos modelos Logístico, Gompertz e Richards para as médias da variável diâmetro.

Modelos	AIC_c	DPR	$(R_{aj.}^2)$
Logístico	15,9900	0,6976	0,9463
Gompertz	23,7032	0,5099	0,9735
Richards	6,6600	0,1925	0,9962

De acordo com os critérios adotados para selecionar o modelo mais adequado para descrever os dados de diâmetros dos frutos de cacauzeiros, o modelo Richards apresentou os menores valores para o AIC corrigido e desvio padrão residual.

4.3 Ajuste dos modelos Logístico, Gompertz e Richards para as médias dos volumes

Para analisar o comportamento da variável volume, os modelos foram ajustados aos dados da tabela 2, e verifica-se que até os 105 dias, de acordo com os valores acima, os diâmetros dos frutos de cacauzeiros apresentaram um crescimento mais lento, aumentando gradativamente até os 150 dias e diminuindo acentuadamente a partir dos 165 dias.

Ao realizar os ajustes, o modelo Gompertz não apresentou convergência para a variável diâmetro. As pressuposições de normalidade, independência e homocedasticidade para os modelos Logístico e Richards foram atendidas pelos testes: Shapiro-Wilk com valor-p de 0,134 e 0,485; Breusch-Pagan com valor-p iguais a 0,121 e 0,182 e Durbin-Watson com 0,149 e 0,146.

Os valores das estimativas dos parâmetros para a variável volume dos frutos de cacauzeiros estão descritos na tabela 9, considerando erros independentes e variância constante.

Tabela 9 Estimativas para os parâmetros dos modelos Logístico e Richards para as médias da variável volume, com erros independentes e variância constante.

Modelos	Parâmetros			
	K	N_0	r	β
Logístico	403,89	0,0191	0,0931	–
Richards	402,26	0,1211	0,1100	1,5396

É possível observar, na tabela acima, que o ajuste dos dois modelos convergiram para estimativas aproximadas para o parâmetro K , porém os valores das estimativas do parâmetro K e N_0 estão muito abaixo dos valores observados apresentados na tabela 2, e também verificado na figura 3. Os parâmetros N_0 , r e β para o modelo Richards não foi significativo ao nível de 5% de significância. O parâmetro N_0 também não foram significativos para o modelo Logístico. O ponto de inflexão para o modelo logístico foi alcançado em, $\frac{K}{2} \approx 201,94$ no tempo $\frac{\ln\left(\frac{K}{N_0}-1\right)}{r} \approx 107$ dias. O modelo Richards atinge ponto de inflexão em $K\left(\frac{1}{1+\beta}\right)^{\frac{1}{\beta}} \approx 219,58$ entre 105 e 120 dias.

Os modelos ajustados às médias dos volumes dos frutos encontram-se na figura 3.

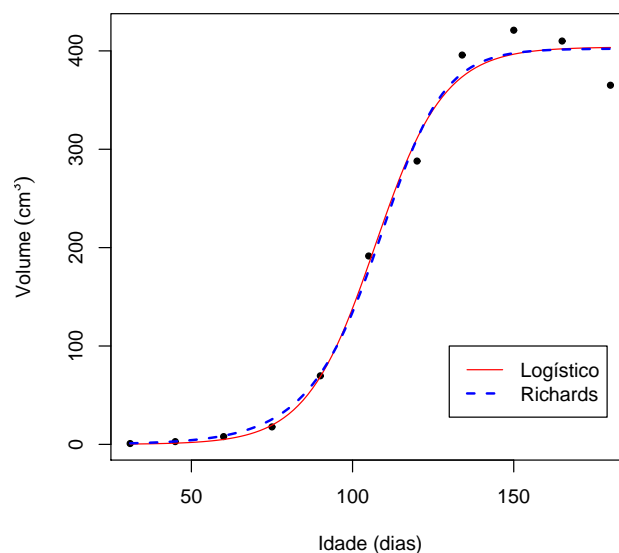


Figura 3 Ajuste dos modelos Logístico e Richards aos dados de volumes dos frutos de cacauzeiros.

As medidas de não linearidade e os vieses de Box para a variável volume estão descritas na tabela 10. Através dos valores de $2F^{-1/2}$ ao nível de 5% de significância com 3 e 8 graus de liberdade, calculado para os modelos Logístico é de 1,098 e para o modelo Richards de 1,053, com 4 e 7 graus de liberdade, verifica-se que o efeito devido a não linearidade intrínseca foi não significativo, para os dois modelos, e a não linearidade devido ao efeito de parâmetros apresenta valores bastante elevados, indicando que essa medida é significativa.

Os parâmetros β e N_0 , do modelo Richards, apresentaram os maiores valores para os vieses de Box, para o modelo Logístico, o parâmetro N_0 também excedeu o limite de 1%, enquanto que os demais mostram-se abaixo, como sugerido por Box (1971).

Tabela 10 Medidas de curvatura e vieses de Box, em (%), para os modelos Logístico e Richards para a variável volume.

Modelos	Medidas de Curvatura		Viés de Box			
	γ^T	γ^N	K	N_0	r	β
Logístico	21,74	0,26	0,0069	3,973	0,0997	—
Richards	50,10	0,51	0,0192	11,549	0,7688	1,159

De acordo com as estimativas dos avaliadores para a qualidade do ajuste, tabela 11, o modelo Logístico mostra-se o mais adequado para descrever os dados de volume dos frutos de cacauzeiros.

Tabela 11 Estimativas dos avaliadores de qualidade dos ajustes dos modelos Logístico e Richards para as médias da variável volume.

Modelos	AIC_c	DPR	$(R_{aj.}^2)$
Logístico	105,17	20,11	0,9879
Richards	110,05	21,15	0,9863

4.4 Algumas considerações sobre os ajustes dos modelos

Na tabela 12, estão apresentados os intervalos de confiança para os parâmetros dos modelos selecionados. Para as variáveis comprimento e diâmetro, modelo Richards, e para a variável volume o modelo Logístico.

Tabela 12 Intervalos de 95% de confiança para os parâmetros do modelo Richards para as variáveis comprimento e diâmetro e para o modelo Logístico para a variável volume dos frutos de cacauzeiros.

	Comprimento		Diâmetro		Volume	
	LI	LS	LI	LS	LI	LS
K	10,225	10,771	8,191	8,776	375,831	431,95
N_0	0,768	1,377	0,202	0,523	-0,043	0,081
r	0,018	0,026	0,042	0,162	0,062	0,123
β	-0,965	16,697	0,940	6,632	-	-

De acordo com a tabela a cima, com exceção dos parâmetros: β para as duas variáveis, e K para a variável volume, os intervalos apresentaram pequena amplitude. Os parâmetros β e N_0 para as variáveis comprimento e volume, o valor zero está incluso, o que os torna não significativos, concordando com os resultados do teste t .

Para a variável volume, na seção 4.3, os modelos subestimaram o valor máximo atingido pelo fruto, provavelmente porque a partir da maturação completa a retração do comprimento e do diâmetro do fruto reflete acentuadamente no volume, sendo atribuída à perda de água na maturação do fruto (BRITO; SILVA, 1983). Assim, é necessário um estudo mais detalhado para descrever o crescimento dos volumes dos frutos de cacauzeiros, pois ao atingir o ponto de estabilização há uma diminuição considerável, com isso os modelos estudados não ajustam

de modo semelhante o desenvolvimento do volume do fruto.

Nas seções 4.1, 4.2 e 4.3, observou-se que a não linearidade devido ao efeito de parâmetros foi significativa para os três modelos e, de acordo com os vieses calculados, o parâmetro N_0 foi o maior responsável na variação das estimativas dos parâmetros nos modelos Logístico e Gompertz. Para o modelo Richards, as variações foram identificadas nos parâmetros r e β . Bates e Watts (1988) afirmam que uma reparametrização pode realizar mudanças na não linearidade devido ao efeito dos parâmetros, porém alertam que reparametrizações para certos tipos de parâmetros e modelos podem reduzi-las para um conjunto de dados e aumentá-las para outro.

5 CONCLUSÃO

O modelo Richards apresenta melhor ajuste aos dados, tornando-se adequado para descrever as variáveis comprimento e diâmetro dos frutos de cacauzeiros em função da idade. O comprimento e diâmetro máximo atingidos pelo fruto são descritos adequadamente pelo modelo, o valor inicial foi subestimado.

Para a variável volume, os modelos em estudo não ajustaram adequadamente o seu desenvolvimento.

Quando adotado erros autorregressivos de primeira ordem, os modelos Logístico e Gompertz para as variáveis comprimento e diâmetro, não apresentaram melhora nos ajustes.

A parametrização dos modelos, embora permita interpretação biológica para três parâmetros, apresentou dificuldades na convergência e causou variações consideráveis na obtenção das estimativas.

Os frutos ao atingirem o ponto de maturação apresentam diminuição nas medidas, causada pela perda de água, e isso é verificado principalmente na variável volume. Diante desse comportamento, realizar estudos utilizando modelos que descrevam essa redução no final do crescimento seria mais adequado, uma vez que os modelos utilizados não satisfazem essa condição. Com isso, uma possibilidade em trabalhos futuros é utilizar modelos conhecidos como modelos platô.

REFERÊNCIAS

- ABDI, H. The method of Least Square. In: _____. **Encyclopedia of measurement and statistics**. Thousand Oaks: UTD, 2007. p. 1-7.
- AKAIKE, H. A. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Minato-Ku, v. 19, n. 6, p. 716-723, Dec. 1974.
- ALMEIDA, A. F.; VALLE, R. R. Análise de crescimento do fruto e das sementes de sete genótipos de Theobroma Cacao L. **Pesquisa Agropecuária Brasileira**, Brasília, v. 30, n. 7, p. 909-916, jul. 1995.
- ALMEIDA, S. L. R.; CHAVES, S. H. G. Análise de crescimento de mudas de cacau CCN-51. **Revista Verde**, Mossoró, v. 6, n. 1, p. 196-200, 2011.
- ALVERSON, W. S. et al. Phylogeny of the coremalvales: evidence from NDHF sequence data. **American Journal of Botany**, Columbus, v. 86, n. 10, p. 1474-1486, Oct. 1999.
- BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its applications**. New York: J. Wiley, 1988. 365 p.
- BENŠIĆ, M. Confidence regions and intervals in nonlinear regression. **Mathematical Communications**, Chichester, v. 2, n. 1, p. 71-76, 1997.
- BERTALANFFY, L. V. Quantitative laws in metabolism and growth. **Biological Research**, Santiago, v. 32, n. 3, p. 217-231, 1957.
- BOX, G. E.; HUNTER, J. S.; HUNTER, W. G. **Statistics for experimenters: desing, innovation, and discovery**. 2nd ed. New Jersey: J. Wiley, 2005. 633 p.
- BOX, M. J. Bias in nonlinear estimation. **Journal of the Royal Statistical Society**, London, v. 33, n. 2, p. 171-201, 1971.
- BRASIL. Ministério da Ciência e Tecnologia. **Conservação, caracterização, avaliação e utilização de germoplasma de cacau (Theobroma Cacao L.) silvestre da Amazônia**. Brasília, 1999. 11 p.

BREUSCH, T. S.; PAGAN, A. R. A simple test for heteroscedasticity and random coefficient variation. **Econometrica**, Chicago, v. 47, n. 5, p. 1287-1294, Sept. 1979.

BRITO, I. C.; SILVA, C. P. Medidas biométricas do fruto do cacau durante seu desenvolvimento. **Sitientibus**, Feira de Santana, v. 2, n. 3, p. 59-66, 1983.

BRUNA, E. D. Curva de crescimento de frutos de pêsego em regiões subtropicais. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 29, n. 3, p. 685-689, 2007.

BURNHAM, K. P.; ANDERSON, D. R. **Model selection and multimodel inference**: a practical information-theoretic approach. New York: Springer-Verlag, 2002. 350 p.

CARVALHO, C. A. L. et al. **Tópicos em ciências agrárias**. Cruz das Almas: Universidade Federal do Recôncavo da Bahia, 2009. v. 1, 296 p.

CENTRO DE PESQUISAS DO CACAU; COMISSÃO EXECUTIVA DO PLANO DA LAVOURA CACAUEIRA. **Características gerais do cacau**. Disponível em: <<http://www.ceplac.gov.br/radar/cacau.html>>. Acesso em: 5 fev. 2013.

DANTAS NETO, A. et al. Caracterização de uma população de cacau para mapeamento de genes de resistência à vassoura-de-bruxa e podridão-parda. **Fitopatologia Brasileira**, Brasília, v. 30, n. 4, p. 380-386, jul./ago. 2005.

DIAS, L. A. S. **Melhoramento genético do cacau**. Viçosa, MG: FUNAPE, 2001. 578 p.

DIAS, L. A. S. et al. Período mínimo de colheita para avaliação de cultivares de cacau em Linhares, ES. **Revista Árvore**, Viçosa, MG, v. 27, n. 4, p. 495-501, 2003.

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

DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression, I. **Biometrika**, London, v. 37, n. 3/4, p. 409-428, 1950.

FERNANDES, T. J. **Curva de crescimento do fruto do cafeeiro em diferentes alinhamentos de plantio utilizando modelos não lineares**. 2012. 80 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2012.

FRANZEN, M.; MULDER, M. B. Ecological, economic and social perspectives on cocoa production worldwide. **Biodiversity and Conservation**, Bengaluru, v. 16, n. 13, p. 3835-3849, June 2007.

FUNDAÇÃO OSWALDO CRUZ. **Vassoura-de-bruxa**. Disponível em: <<http://www.invivo.fiocruz.br/cgi/cgilua.exe/sys/start.htm?infoid=873&sid=9>>. Acesso em: 10 abr. 2014.

GAZOLA, S. et al. Proposta de modelagem não linear do desempenho germinativo de sementes de milho híbrido. **Ciência Rural**, Santa Maria, v. 41, n. 4, p. 551-556, jul./ago. 2001.

GUEDES, M. H. 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, 2004.

HENRIQUES, A. **Thomas Robert Malthus: a teoria malthusiana**. Coimbra: Instituto Politécnico de Coimbra, 2007. 10 p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Levantamento sistemático da produção agrícola**. Rio de Janeiro, 2012. 88 p.

INTERNATIONAL COCOA ORGANIZATION. **Annual report**. London, 2012. Disponível em: <<http://www.icco.org>>. Acesso em: 20 ago. 2013.

JENNRICH, R. I. Asymptotic properties of non-linear least squares estimators. **Institute of Mathematical Statistics**, Beachwood, v. 40, n. 2, p. 633-643, 1969.

LOURENÇO, B.; QUEIROZ, L. B. Crescimento e desenvolvimento puberal na adolescência. **Revista Medicina**, São Paulo, v. 89, n. 2, p. 70-75, 2010.

LUCCHESI, A. A. Utilização prática da análise quantitativa do crescimento vegetal. **Anais da Escola Superior de Agricultura Luiz de Queiroz**, Piracicaba, v. 41, p. 181-202, 1984.

MAIA, E. et al. Método de comparação de modelos de regressão não-linear em bananeiras. **Ciência Rural**, Santa Maria, v. 39, n. 5, p. 1380-1386, 2009.

MALTHUS, T. R. **An essay on the principle of population**. London: Electronic Scholarly, 1798. Disponível em: <<http://www.esp.org>>. Acesso em: 5 out. 2013.

MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameter. **Journal of the Society for Industrial and Applied Mathematics**, Philadelphia, v. 11, n. 2, p. 431-441, 1963.

MARTINS, J. M. et al. **Melhoria da qualidade de cacau**. Ilhéus: CEPLAC/CENEX, 2011. 45 p.

MAZUCHELI, J.; ACHCAR, J. Algumas considerações em regressão não linear. **Acta Scientiarum Technology**, Maringá, v. 24, n. 6, p. 1761-1770, 2002.

MAZZINI, A. R. A. et al. Análise da curva de crescimento de machos Hereford. **Ciência e Agrotecnologia**, Lavras, v. 27, n. 5, p. 1105-1112, set./out. 2003.

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.

MEINHARDT, W. L. et al. Moniliophthora perniciosa, the causal agent of witches' broom disease of cacao: what's new from this old foe? **Molecular Plant Pathology**, London, v. 9, n. 5, p. 577-588, 2008.

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, 2008.

MOTOLSKY, H. J.; RANSNAS, L. A. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. **The FASEB Journal**, New York, v. 1, n. 5, p. 365-374, 1987.

MUNIZ, J. A.; SAVIAN, T. V.; SCALON, J. D. Parameters estimation in the model for in situ degradability of mertens and loften. **Ciência e Agrotecnologia**, Lavras, v. 32, n. 5, p. 1622-1628, set./out. 2008.

OLIVEIRA, M. L.; LUZ, E. D. M. N. **Identificação e manejo das principais doenças do cacaueteiro do Brasil**. Ilhéus: CEPEC, 2005. 19 p.

PEREIRA, J. M.; MUNIZ, J. A.; SILVA, C. A. Nonlinear models to predict nitrogen mineralization in an Oxisol. **Scientia Agricola**, Piracicaba, v. 62, n. 4, p. 395-400, 2005.

PRADO, T. K. L.; 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, maio 2013.

PRAJNESHU. A nonlinear statistical model for aphid population growth. **Journal of the Indian Society of Agricultural Statistics**, New Delhi, v. 51, n. 1, p. 73-80, 1998.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 10 mar. 2014.

RATKOWSKY, D. A. Principles of nonlinear regression modeling. **Journal of Industrial Microbiology**, Berlin, v. 12, n. 3/5, p. 195-199, 1993.

RICHARDS, F. J. A flexible growth function for empirical use. **Journal of Experimental Botany**, Oxford, v. 10, n. 29, p. 290-300, 1959.

ROSAND, P. C. et al. **Exigências nutricionais e fertilização do cacaueteiro**. Itabuna: CEPLAC, 1975. 60 p. (Boletim Técnico, 30).

ROSÁRIO, M. et al. **Cacau história e evolução no Brasil e no mundo: Theobroma cacao, consumo, desenvolvimento agrícola**. Ilhéus: CEPLAC, 1978. 46 p.

SANDLAND, R. L.; MCGILCHRIST, C. A. Stochastic growth curve analysis. **International Biometric Society**, Washington, v. 35, p. 255-271, 1979.

SANTOS, V. R. et al. Crescimento e produtividade agrícola de cana-de-açúcar em diferentes fontes de fósforo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 13, n. 4, p. 389-396, 2009.

SAVAGEAU, M. A. Growth equations: a general equation and a survey of special cases. **Mathematical Biosciences**, New York, v. 48, n. 3, p. 267-278, 1980.

SAVIAN, T. V.; MUNIZ, J. A. A study of in situ degradability: heterogeneity of variances and correlated errors. **Scientia Agricola**, Piracicaba, v. 64, n. 5, p. 548-554, 2007.

SEBER, G. A. F.; WILD, C. J. **Nonlinear regression**. New York: J. Wiley, 1988. 792 p.

SENHOR, R. F. et al. Manejo de doenças pós-colheita. **Revista Verde**, Limoeiro, v. 4, n. 1, p. 1-13, 2009.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality: complete samples. **Biometrika**, London, v. 52, n. 3/4, p. 591-611, 1965.

SMYTH, G. K. Nonlinear regression. In: _____. **Encyclopedia of environmetrics**. Chichester: J. Wiley, 2002. v. 3, p. 1405-1411.

SOUZA, E. M. et al. Modelagem não linear da extração de zinco de um solo tratado com lodo de esgoto. **Acta Scientiarum: Technology**, Maringá, v. 32, n. 2, p. 193-199, 2010.

SOUZA, G. da S. **Introdução aos modelos de regressão linear e não linear**. Brasília: EMBRAPA, 1998. 505 p.

TERRA, M. F.; MUNIZ, J. A.; SAVIAN, T. V. Ajuste dos modelos Logístico e Gompertz aos dados de crescimento de frutos da tamareira-anã (O'brien). **Magistra**, Cruz das Almas, v. 22, n. 1, p. 1-7, 2010.

TSOULARIS, A.; WALLACE, J. Analysis of logistic growth models. **Mathematical Biosciences**, Chicago, v. 179, n. 1, p. 21-55, 2001.

WATTS, D. G.; BATES, D. M. Relative curvature measures of nonlinearity. **Journal of the Royal Statistical Society. Series B, Methodological**, London, v. 42, n. 1, p. 1-25, 1980.

ZEIDE, B. Analysis of growth equations. **Forest Science**, Bethesda, v. 39, n. 3, p. 594-616, 1993.

ZEVIANI, W. M. et al. Modelos não lineares para a liberação de potássio de esterco animal em latossolos. **Ciência Rural**, Santa Maria, v. 42, n. 10, p. 1789-1796, 2012.

ANEXO

```

> ##### ROTINA NO R #####
> # MODELO LOGISTICO - VARIÁVEL COMPRIMENTO
>
> rm(list=ls(all=TRUE))
> cacau1=read.table(
+ "D:/DISSERTAÇÃO_MICHERLANIA/R_DISSERTACAO/comprimento.txt",
+                                     header=T)
> summary(cacau1)
> attach(cacau1)
> ## CARREGANDO BIBLIOTECAS
> require(minpack.lm)
> require(nlme)
> require(nlrwr)
> require(plm)
> require(lattice)
> require(stats)
> require(lmtest)
> require(car)
> require(bstats)
> require(qpcR)
> logistico1 <- nls(comprimento~K/(1-(1-K/NO)*exp(-r*idade)),
+ data=cacau1,start=list(K=11,NO=2.1,r=0.02));summary(logistico1)
> confint2(logistico1) # intervalo de confiança
> R2.aj=Rsq.ad(logistico1);R2.aj # R2_ajustados
> ## OBTENDO AS DERIVADAS
> func.log <- deriv3(~K/(1-(1-K/NO)*exp(-r*idade)),
+ c("K", "NO","r"),function(idade, K,NO,r) NULL) # REDEFININDO FUNCAO
> str(func.log) # MOSTRA E ESTRUTURA INTERNA DE UMA FUNCAO
> args(func.log) # EXIBE OS ARGUMENTOS DE UMA FUNÇÃO
> m <- func.log(idade, K=coef(logistico1)["K"],
+               NO=coef(logistico1)["NO"],r=coef(logistico1)["r"])
> # GRADIENTE E HESSIANA
> grad <- attr(m, "gradient");grad
> hes <- attr(m, "hessian");hes
> nls.I <-lm(comprimento ~ -1+grad,data=cacau1);nls.I

```

```

> # TESTES
> # normalidade
> shapiro.test(residuals(logistico1))
> # heterogeneidade
> bptest(nls.I)
> #processo de autoregressao de ordem 1
> dw.test(nls.I)
> # MEDIDAS DE CURVATURA
> curv.l <- nls(comprimento ~ func.log(idade,K,N0,r),
+             data=cacau1,start=list(K=11,N0=2.1,r=0.02))
> rms.curv(curv.l)
> f <- qf(0.95,10,(10-4))
> valor <- 2*(f)^(-1/2)
> # USANDO A FUNÇÃO GNLS
> logistico2=gnls(comprimento~K/(1-(1-K/N0)*exp(-r*idade)),
+             data=cacau1,start=list(K=10.6,N0=0.8,r=0.01),
+             correlation=corAR1())
> summary(logistico2)
> c(AICc(logistico1),AICc(logistico2)) # AIC corrigido
> # Medida de Vício de Box
> FF <- t(grad)%*%grad
> sigma2 <-summary(logistico1)$sigma
> n=11
> b <- - (sigma2/2)*sapply(1:n,function(x){
+             sum(diag(solve(FF)%*%hes[x,,]))});b
> B <- as.vector(solve(FF)%*%t(grad)%*%b);B
> B.K <- 100*(B/coef(logistico1)[,]);B.K
> #DEFINIDO GRÁFICOS
> par(mfrow=c(1,1))
> plot(comprimento~idade,pch=20,xlab="Idade (dias)",
+       ylab="Comprimento (cm)",data=cacau1)
> with
> curve(10.864150/(1-(1-10.864150/0.425076)*exp(- 0.047269*x)),
+       add=T,col=2,lwd=1)#logistico
> with
> curve(11.204425*(exp(log(0.04581/11.204425)*exp(-0.031*x))),
+       add=T,col="blue",lty=2,lwd=1)#gompertz

```

```
> with
> curve(10.498474/((1+((10.498474/1.072978)^(7.866452)
+       - 1)*exp(-0.022510*7.866452*x))^(1/7.866452)),
+       add=T,col=1,lty=1,lwd=1)#richards
> legend(130,4,c("Logístico","Gompertz","Richards"),
+       col=c(2,"blue",1),lty=c(1,2,1),lwd=c(1,1,1))
>
```