



THALITA KELEN LEAL DO PRADO

**MODELOS NÃO-LINEARES NO
CRESCIMENTO DE FRUTOS DE COQUEIRO**

LAVRAS - MG

2011

THALITA KELEN LEAL DO PRADO

**MODELOS NÃO-LINEARES NO CRESCIMENTO DE FRUTOS DE
COQUEIRO**

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.

Orientador

Dr. Joel Augusto Muniz

Co-Orientadores

Dra. Taciana Villela Savian

Dra. Thelma Sáfydi

LAVRAS-MG

2011

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Prado, Thalita Kelen Leal do.

Modelos não-lineares no crescimento de frutos de coqueiro /
Thalita Kelen Leal do Prado. – Lavras : UFLA, 2011.
92 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2011.
Orientador: Joel Augusto Muniz.
Bibliografia.

1. Modelos de crescimento. 2. Método de Monte Carlo via
cadeias de Markov. 3. Mínimos quadrados. I. Universidade Federal
de Lavras. II. Título.

CDD – 519.536

THALITA KELEN LEAL DO PRADO

**MODELOS NÃO-LINEARES NO CRESCIMENTO DE FRUTOS DE
COQUEIRO**

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 14 de fevereiro de 2011.

Dra. Taciana Villela Savian ESALQ

Dra. Thelma Sáfyadi UFLA

Dr. Gerson Barreto Mourão ESALQ

Dr. Joel Augusto Muniz
Orientador

LAVRAS-MG

2011

*Aos meus pais, Jair e Maria, que me apoiaram, aconselharam e incentivaram
para que eu chegasse até aqui e nunca desistisse...*

Aos meus irmãos, Junior e Suziléa.

À minha cunhada Ariane.

DEDICO.

AGRADECIMENTOS

A Deus, em primeiro lugar, por permitir mais esta conquista em minha vida.

Ao professor Dr. Joel Augusto Muniz pela oportunidade, incentivo e orientação no decorrer deste curso.

À professora Dra. Taciana Villela Savian pela disponibilidade, orientação, paciência e grande auxílio prestado na elaboração desta dissertação.

À professora Dra. Thélma Sáfy pela colaboração no projeto de pesquisa, do qual se desenvolveu as análises aqui apresentadas e a leitura da dissertação.

Aos professores do Departamento de Ciências Exatas (DEX), pelos ensinamentos transmitidos que colaboraram para o meu crescimento.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX) pela oportunidade de realizar o mestrado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudos.

À Josi, secretária do Departamento de Ciências Exatas (DEX), pela atenção e simpatia sempre prestada.

Aos amigos Alexandre, Diogénes, Suzana, Lucas, Iábita, Thaís, Juliano, Andressa, Luzia, Vanessa, Carol, Eugênia, Maria Cecília, Dedê, enfim todos que contribuíram durante a minha estada em Lavras, pelos momentos de descontração e aprendizado.

Em especial, às amigas, Adriana, Suzana, Vania, Marina, Débora, pelo auxílio e amizade sincera, compartilhando e apoiando sempre nos momentos tristes e felizes. Sempre serei grata.

À minha família, principalmente aos meus pais, pelo amor, paciência, dedicação, atenção, cuidado, orações, essenciais para que eu finalizasse mais

uma etapa da minha formação profissional. Serei eternamente grata por tudo e em muito me alegro por viverem comigo mais uma etapa da minha vida.

A Pedra

O distraído nela tropeçou...

O bruto a usou como projétil.

O empreendedor, usando-a, construiu.

O camponês, cansado da lida, dela fez assento.

Para meninos, foi brinquedo.

Drummond a poetizou.

Já, Davi, matou Golias, e Michelangelo extraiu-lhe a mais bela escultura...

E em todos esses casos, a diferença não esteve na pedra, mas no homem!

Não existe "pedra" no seu caminho que você não possa aproveitá-la para o seu próprio crescimento.

Independente do tamanho das pedras, no decorrer de sua vida, não existirá uma, que você não possa aproveitá-la para seu crescimento espiritual. Quanto a sua pedra atual, tenho certeza que Deus irá te dar sabedoria, para mais tarde você olhar para ela, e ter orgulho da maravilhosa experiência que causou em sua vida, no seu crescimento espiritual.

Autor Desconhecido.

Entrega o teu caminho ao Senhor, confia nele, e o mais Ele fará.

Salmos 37.5-Bíblia Sagrada.

RESUMO

A avaliação dos processos de desenvolvimento de frutos permite estabelecer as bases para definir o ponto mais adequado de colheita e estratégias para sua conservação, visando o aumento da vida útil pós-colheita. A descrição do crescimento de plantas e de frutos, por meio do acúmulo de biomassa ao longo do tempo, é fundamental para o manejo adequado e detecção de fatores que afetam o desenvolvimento das plantas, tais como: ocorrências de deficiências hídricas e nutricionais, competição de plantas daninhas, solos compactados ou mal drenados entre outros, podendo auxiliar no aperfeiçoamento de metodologias de pesquisa sobre a cultura, como, por exemplo, na identificação de parâmetros mais adequados para a avaliação do crescimento da planta nas diversas fases do ciclo. A estimação dos parâmetros dos modelos não lineares, Logístico e Gompertz, foi feita pelo método de Mínimos Quadrados e por meio do método iterativo de Monte Carlo via cadeias de Markov, considerando as estruturas de erros independentes e autorregressivos AR(1). Posteriormente, foram utilizados os critérios de comparação de modelos: coeficiente de determinação ajustado, desvio padrão residual e critério de informação de Akaike. Foi realizado um estudo por meio de dados simulados e dados reais do diâmetro externo longitudinal e transversal dos frutos de coqueiro da variedade anã verde. Pelo processo de simulação, a precisão do método mostrou-se eficiente, pois apresentou estimativas próximas ao valor paramétrico. As estimativas obtidas com base nos modelos ajustados além da interpretação biológica dos parâmetros estão em conformidade com a descrição das características das variedades de coqueiros na literatura científica. O modelo Logístico proporcionou melhor qualidade no ajuste em ambos os métodos de estimação de parâmetros aos dados experimentais, considerando as diferentes estruturas de erros.

Palavras-chave: Modelos de crescimento. Método de Monte Carlo via Cadeia de Markov. Mínimos Quadrados. Inferência clássica.

ABSTRACT

The evaluation of the processes of fruit development allows us to establish the basis for defining the most appropriate point of collection and strategies for their conservation, aiming to increase shelf-life. The description of the growth of plants and fruits, through the accumulation of biomass over time is essential for the proper handling and detection of factors that affect plant growth, such as instances of water shortages and nutrition, competition from weeds, compacted soils or poorly drained among others, could help in the development of methods for research on culture, for example, the identification of more appropriate parameters for the assessment of plant growth at different stages of the cycle. The parameter estimation of nonlinear models, Logistic and Gompertz, was made by the Least Squares method and through an iterative Monte Carlo Markov chain, considering the structures of independent errors and autoregressive AR (1). Subsequently, criteria were used for comparison of models: adjusted coefficient of determination, residual standard deviation and the Akaike information criterion. A study was conducted using simulated data and real data from the longitudinal and transverse outer diameter of the fruits of green dwarf coconut trees. For the simulation process, the accuracy of the method was efficient, as it showed estimates close to the parametric value. The estimates obtained based on the adjusted models beyond the biological interpretation of the parameters are in accordance with the description of the characteristics of the varieties of palm trees in the scientific literature. The Logistic model provided the best quality in the setting of both the methods of parameter estimation to experimental data, considering the different error structures.

Keywords: Growth models. Method of Monte Carlo Markov Chain. Least Squares. Classical inference.

SUMÁRIO

1	INTRODUÇÃO.....	11
2	REFERENCIAL TEÓRICO.....	14
2.1	Modelos de crescimento.....	14
2.2	Modelos não-lineares.....	18
2.3	Método de estimação por mínimos quadrados.....	21
2.3.1	Método dos mínimos quadrados generalizados.....	22
2.4	Métodos de simulação.....	23
2.4.1	Método de Monte Carlo via cadeias de Markov.....	24
2.4.1.1	Algoritmo de Metropolis-Hastings.....	26
2.4.1.2	Amostrador de Gibbs.....	26
2.4.2	Diagnóstico de convergência dos Métodos de Monte Carlo via cadeias de Markov.....	28
2.4.3	Erro de Monte Carlo.....	33
3	MATERIAIS E MÉTODOS.....	35
3.1	Procedimentos de ajuste por meio do método de mínimos quadrados.....	37
3.2	Procedimentos de ajuste por meio do método de Monte Carlo via cadeias de Markov.....	40
4	RESULTADOS E DISCUSSÃO.....	41
4.1	Ajuste pelo método de mínimos quadrados.....	41
4.2	Ajuste por meio do método de Monte Carlo via Cadeias de Markov.....	45
4.2.1	Obtenção das distribuições condicionais completas para os parâmetros do modelo Logístico.....	47
4.2.2	Obtenção das distribuições condicionais completas para os parâmetros do modelo Gompertz.....	52
4.3	Análise dos dados simulados.....	56
4.3.1	Modelo Logístico.....	56
4.3.2	Modelo Gompertz.....	66
4.4	Análise dos dados reais.....	78
4.4.1	Modelo Logístico.....	78
4.4.2	Modelo Gompertz.....	80
4.4.3	Comparação dos modelos.....	84
5	CONCLUSÃO.....	85
	REFERÊNCIAS.....	86

1 INTRODUÇÃO

Nos trabalhos com frutíferas um aspecto importante a ser considerado é o “ponto de colheita” que na maioria das vezes, é determinado de maneira subjetiva, considerando-se a coloração externa e o tamanho do fruto. A utilização de índices de maturação permite que se conheça a fase ideal de colheita para cada situação mercadológica. A análise do crescimento e desenvolvimento dos frutos é fundamental para descrever as mudanças ocorridas nas características físicas e componentes químicos, possibilitando conhecer a dinâmica envolvida, auxiliando nas estratégias adotadas na gestão do processo de produção e colheita.

A avaliação dos processos de desenvolvimento de um fruto permite estabelecer as bases para definir o ponto mais adequado de colheita e estratégias para sua conservação, visando o aumento da vida útil pós-colheita. O conhecimento do crescimento de plantas e frutos, medido por meio do acúmulo de biomassa ao longo do tempo, são fundamentais para o manejo adequado e a detecção de fatores que influenciam o desenvolvimento das plantas, tais como: ocorrências de deficiências hídricas e nutricionais, competição de plantas daninhas, solos compactados ou mal drenados, auxiliando no aprimoramento de metodologias de pesquisa sobre a cultura, como, por exemplo, na identificação de parâmetros mais adequados para a avaliação do crescimento da planta nas diversas fases do ciclo.

Os frutos dos coqueiros anã e híbrido destinados ao consumo “in natura” de água de coco geralmente são colhidos entre o sexto e o sétimo mês, após a abertura natural da inflorescência. Nessa idade ocorrem os maiores valores para peso de fruto e produção de água de coco, teores de frutose, glicose, grau brix e sais minerais, principalmente potássio, os quais, conferem melhor sabor à água de coco. A água proveniente de frutos com idade em torno de cinco meses, é

menos doce (menores teores de glicose e frutose e menor grau brix), enquanto na dos frutos com oito meses de idade, já ocorrem quedas nos teores de glicose e frutose e no grau brix, e aumento no teor de sacarose e provavelmente no de gordura, ocasionando um sabor rançoso a água de coco.

Os frutos de coqueiro da variedade Anã são compostos das cultivares Verde, Amarela, Vermelha da Malásia e Vermelha de Camarões. A variedade Anã Verde tem sido cultivada com o objetivo principal de sua comercialização para o consumo da água-de-coco, seja sob a forma “in natura” diretamente dos frutos ou processadas nas agroindústrias.

Os estudos de curvas de crescimento são importantes uma vez que podem ser realizados em situações em que a informação contida numa seqüência de pontos “tamanho-tempo”, do crescimento de uma espécie seja reduzida e sintetizada num conjunto de parâmetros que tenham interpretação prática e preditiva.

Vários tipos de modelos estatísticos podem ser usados para estudar os processos de crescimento envolvidos no sistema de produção vegetal. Na maioria dos casos, as pesquisas que avaliam crescimento e desenvolvimento de frutos, utilizam modelos de regressão polinomial, e em algumas situações não tem sido adequado para descrever os fenômenos. Os modelos não-lineares fornecem informações e estimativas úteis, particularmente se forem baseados em parâmetros que permitem interpretação biológica. Os parâmetros dos modelos não-lineares são utilizados para se obter, por exemplo, taxa de crescimento, peso à maturidade e grau de amadurecimento.

Entre os modelos não-lineares que podem ser usados para expressar o crescimento de plantas e seus componentes, estão os modelos de Gompertz, Logístico, Brody e Von Bertalanffy. A estimação dos parâmetros pode ser feita utilizando-se diferentes métodos, a saber: métodos de mínimos quadrados,

método da máxima verossimilhança e métodos dos momentos, considerando-se diferentes estruturas de erros.

Pretende-se, com esse trabalho, estudar o ajuste dos modelos de Gompertz e Logístico, por meio do método de mínimos quadrados e do método de Monte Carlo via Cadeias de Markov, utilizando dados de crescimento de fruto de coqueiro da variedade Anã Verde, considerando-se as estruturas de erros independentes e erros autorregressivos de primeira ordem (AR1) e avaliar a metodologia utilizada no procedimento de ajuste por meio de um estudo de simulação.

2 REFERENCIAL TEÓRICO

2.1 Modelos de crescimento

Cardoso et al. (2006), destacam que uma forma de maximizar a eficiência de plantas em captação dos recursos limitantes (nutrientes, luz, água entre outros) ao crescimento, desenvolvimento e produção, é manejar adequadamente a cultura, com a utilização de espaçamento e/ou arranjo espacial eficientes, uma população de plantas adequada, cultivar adaptada ao local específico, entre outros. A análise de crescimento neste caso entra como uma ferramenta importante que avalia a eficiência desses manejos, sendo bastante utilizada por pesquisadores das Ciências Agrárias.

O estágio de desenvolvimento dos frutos no momento da colheita tem influência na qualidade do fruto maduro. Quando os frutos são colhidos verdes ou fisiologicamente imaturos, não amadurecem, enrugam e apresentam exsudação da seiva, ou quando o amadurecimento ocorre, a qualidade dos frutos é prejudicada. Por outro lado, os frutos colhidos muito maduros deterioram-se rapidamente, não podendo ser armazenados e/ou comercializados em locais distantes (KAYS, 1997). A avaliação do padrão de desenvolvimento de um fruto a partir do florescimento auxilia no estabelecimento de índices de maturidade (COOMBE, 1976).

Segundo Cavalini et al. (2006), o ponto de colheita é um dos principais fatores de perda pós-colheita e é determinado, na grande maioria das vezes, de maneira subjetiva através da coloração externa e tamanho do fruto, não havendo uma padronização consensual entre os produtores. A utilização de índices de maturação permite que se conheça o ponto de colheita ideal para cada situação mercadológica.

Vários critérios têm sido utilizados na determinação da maturidade de frutos, baseados no aspecto aparente (tamanho, diâmetro, cor, etc.) e na composição química (sólidos solúveis, acidez titulável, etc.) do produto na época da colheita. Porém, estes índices podem variar consideravelmente, dependendo do local de cultivo, cultivares e condições climáticas do ano de crescimento (BIALE; YOUNG, 1964). Um dos índices mais utilizados na determinação do ponto de colheita é o número de dias desde a floração até o desenvolvimento pleno do fruto (WARRINGTON et al., 1999).

No estudo de crescimento de plantas, Urchei, Rodrigues e Stone (2000) citam que para avaliar os efeitos de sistemas de manejo sobre as plantas, a análise de crescimento é fundamental, pois descreve as mudanças na produção vegetal em função do tempo, o que não é possível com o simples registro do rendimento. Segundo Kvet et al. (1971), o estudo do crescimento de comunidades vegetais é um dos primeiros passos na análise de produção primária, caracterizando-se como elo de ligação entre o simples registro do rendimento das culturas e a análise destas por meio de métodos fisiológicos, podendo ser utilizada para conhecer a adaptação ecológica das plantas a novos ambientes, a competição interespecífica, os efeitos de sistemas de manejo e a capacidade produtiva de diferentes genótipos.

Pereira e Machado (1987) afirmam que a avaliação do crescimento representa a referência inicial na análise de produção das espécies vegetais, requerendo informações que podem ser obtidas sem a necessidade de equipamentos sofisticados. Tais informações são a quantidade de material contido na planta toda e em suas partes (folhas, colmos, raízes e frutos), e o tamanho do aparelho fotossintetizante (área foliar), obtidas a intervalos de tempo regulares durante o desenvolvimento fenológico da planta. Os trabalhos encontrados na área vegetal destacam que análise de crescimento é fundamental, pois descreve as mudanças na produção vegetal em função do tempo. Porém,

muitos desses trabalhos fazem uso de modelos de regressão linear ou polinomial o que não é adequado neste tipo de estudo, pois os parâmetros não possuem interpretação prática. Os modelos não-lineares cujos parâmetros apresentam interpretação biológica e que descrevem as curvas de crescimento são amplamente utilizados em experimentos com animais.

Alvarez e Castro (1999) analisaram o comportamento do crescimento de cana crua e cana queimada nos primeiro e segundo anos de rebrota por meio de curvas adaptadas e verificaram que a altura das plantas apresentou curvas de crescimento de formato sigmoidal. Alvarez e Castro (1999) ajustaram, dentre outros modelos, o modelo de crescimento de Gompertz aos dados de matéria seca de folha e colmo, índice de área foliar e acúmulos de graus-dia para cana-de-açúcar.

Pereira et al. (2000) estudaram o comportamento da bananeira (*Musa* spp.) ‘Prata Anã’ no primeiro ciclo de produção sob diferentes espaçamentos e verificaram um padrão sigmoidal para o crescimento das plantas ajustando o modelo logístico para a característica altura das plantas. Corsato (2004), estudando a fenologia e carboidratos de reserva do caquizeiro (*Diospyros kaki* L.) ‘Rama Forte’ em clima tropical no Brasil encontraram para as curvas de crescimento do fruto um padrão sigmoidal duplo. Martins Filho et al. (2008) ajustaram o modelo Logístico para os dados de cultivares de crescimento de duas cultivares de feijoeiro.

Vários tipos de modelos estatísticos podem ser usados, de acordo com suas características, para facilitar a interpretação dos processos envolvidos no sistema de produção vegetal. Esses modelos podem fornecer também informações e estimativas úteis, particularmente se forem baseados em parâmetros que permitem interpretação biológica. Os modelos empíricos usados freqüentemente para estimar o crescimento de frutos incluem a função Logística e de Gompertz.

Terra, Muniz e Savian (2010) verificaram a qualidade do ajuste dos modelos Logístico e Gompertz, aos dados de crescimento, em comprimento e diâmetro dos frutos de tamareira-anã onde o modelo Logístico se ajustou melhor aos dados.

De modo geral, na literatura, diversos trabalhos abordam a importância de se estudar o modelo de crescimento de animais. Brown, Brown e Butts (1972), utilizaram a função Brody para obter estimativas do peso assintótico e da taxa de maturidade em bovinos da raça Hereford e Angus. Denise e Brinks (1985), analisando as mesmas raças, compararam as funções de Brody e Richards em relação à qualidade do ajuste dos parâmetros da curva de crescimento. Bergamasco et al. (2001), compararam os modelos Logístico, Brody e Gompertz, na descrição do crescimento de fêmeas da raça Holandesa, do nascimento aos 2 anos de idade. Guedes et al. (2005) ajustou o modelo Logístico para o estudo de crescimento de cordeiros da raça Santa Inês. Mendes et al. (2009) estudaram o crescimento de fêmeas Hereford por meio da função não linear de Gompertz. Silva et al. (2010) utilizou o modelo de crescimento Von Bertalanffy na avaliação genética de curvas de crescimento de bovinos de corte.

Segundo Vieira e Mischon (1976), a função de Gompertz, embora várias vezes utilizada, não é tão conhecida como a Logística. Primeiramente, a função logística foi indicada para o estudo do crescimento de populações humanas por Verhulst em 1838, conforme Hoffmann e Vieira (1998), o qual a denominou de “curva logística”. Mais tarde, outros autores utilizaram a mesma curva para descrever o crescimento da população americana, de 1870 a 1910. A função de Gompertz foi estudada por Gompertz, Winsor e Laird para descrever a taxa de mortalidade numa população.

2.2 Modelos não lineares

Em regressão, os modelos são classificados basicamente como: modelos lineares (aqueles que são lineares em relação aos parâmetros); modelos linearizáveis (aqueles que podem ser transformados em lineares através de alguma transformação) e modelos não-lineares (modelos em que pelo menos uma das derivadas parciais depende de algum parâmetro do modelo). Alguns modelos de regressão não-linear são utilizados para representar curvas de crescimento e apresentam algumas peculiaridades quanto ao procedimento de ajuste.

Sob o ponto de vista da inferência estatística clássica, para os modelos de regressão linear, os estimadores de mínimos quadrados apresentam propriedades ótimas, isto é, são não-viesados, normalmente distribuídos, e ainda possuem variância mínima entre qualquer outra classe de estimadores. Essas propriedades são consideradas como as melhores que uma classe de estimadores pode apresentar (SEARLE, 1971). Já no caso de modelos não lineares, as propriedades somente são válidas assintoticamente, isto é, quando o tamanho da amostra é suficientemente grande. Em geral, em pequenas amostras essas propriedades não são satisfeitas.

Para os modelos de regressão é comum assumir que os erros sejam independentes, o que não é apropriado quando se trabalha com séries cronológicas de dados, em que o erro da observação relativa a um período está correlacionado com o erro da observação anterior (HOFFMANN; VIEIRA, 1998). Segundo Morettin e Toloí (2004), a característica geral da dependência dos resíduos é a de existir uma variação sistemática dos valores em observações sucessivas. Quando isso ocorre, diz-se que os resíduos são autocorrelacionados. Hoffman e Vieira (1998) consideram que, além da autocorrelação residual pode ocorrer heterogeneidade de variâncias. Neste caso, o método dos quadrados

mínimos generalizado é mais eficiente do que o método dos quadrados mínimos ponderados e ordinários.

Segundo Neter, Wasserman e Kutner (1985), o uso do método dos quadrados mínimos ordinário, nos casos de regressão com erros autocorrelacionados positivamente têm importantes conseqüências, como por exemplo: os estimadores dos coeficientes não são tendenciosos, apesar de levarem à subestimação da variância, podendo ser completamente ineficientes; o quadrado médio do resíduo pode subestimar a variância dos erros; em conseqüência, o desvio padrão calculado de acordo com o método dos mínimos quadrados ordinários pode subestimar o verdadeiro desvio padrão do coeficiente de regressão estimado (em modelos lineares ou não lineares), fazendo com que os intervalos de confiança e testes usando as distribuições t e F sejam inválidos.

Num estudo sobre a predição do nitrogênio mineralizado em latossolo pelo uso de modelos não-lineares, Pereira, Muniz e Silva (2005) incorporaram a autocorrelação residual e encontraram estimativas mais precisas para os parâmetros envolvidos nos modelos, em relação aos modelos assumindo independência dos erros.

Nos estudos de curva de crescimento que envolve medidas repetidas no tempo, tomadas num mesmo indivíduo, é comum ocorrer heterogeneidade de variâncias, e uma alternativa para análise dos dados com essa característica é o da ponderação. Guedes et al. (2004) utilizaram a ponderação pelo inverso dos pesos dos cordeiros das raças Santa Inês e Bergamácia. Mazzini et al. (2005) utilizaram a ponderação pelo inverso da variância dos pesos, juntamente com a estrutura de erros autorregressivos, verificando que houve melhor qualidade de ajuste para as funções de crescimento nas raças Hereford e Nelore, respectivamente. Savian e Muniz (2007) utilizaram a ponderação pelo inverso da variância e também a estrutura de erros autorregressivos para avaliar o resíduo potencialmente degradável da fibra em detergente neutro (FDN).

Em estudo de modelos não-lineares que descrevem o crescimento vegetal é bastante razoável incorporar a autocorrelação, tendo em vista que as medidas de crescimento são tomadas em uma mesma unidade experimental, planta ou fruto, ao longo do tempo, estando, portanto, provavelmente correlacionadas.

Na área de produção vegetal, em diversas situações experimentais, os fenômenos são descritos por modelos não-lineares como, por exemplo: o ajuste de curvas para dinâmica e disponibilização de nutrientes no sistema solo-planta; modelos de crescimento de espécies vegetais, de uma maneira geral, bem como de sementes e frutos; estudos de velocidade de reações químicas – cinética enzimática (Modelo de Michaelis-Menten); processos que descrevem fenômenos envolvidos na fotossíntese.

Toralles et al. (2004), utilizaram o modelo de Michaelis-Menten para estudar o escurecimento enzimático “*in vitro*” em pêssegos e encontraram que o coeficiente de especificidade, função dos parâmetros do modelo, é um bom indicador cinético do estudo “*in vitro*” e, portanto, possível de ser utilizadas como indicador de possível escurecimento em polpa. Machado (2006) estudou de maneira aprofundada o processo de estimação de parâmetros em mecanismos cinéticos de Michaelis-Menten.

Na área vegetal alguns autores têm trabalhado com modelos não-lineares em detrimento aos lineares. Corrêa et al. (2007), na descrição do processo de secagem do feijão (*Phaseolus vulgaris* L.) utilizou doze modelos não lineares dentre os quais o exponencial de dois termos e o Thompson; Benassi et al. (2007) ao avaliarem a biometria de frutos de coqueiro *Cocos nucifera* L. variedade anã-verde, em diferentes estádios de desenvolvimento com o modelo Logístico e Logístico combinado com uma quadrática; Carlesso et al. (2007) compararam cinco modelos não-lineares exponenciais e variações deste, com o objetivo de determinar o modelo que melhor descreve a secagem por convecção

em camada fina de sementes de maracujá-amarelo; Jesus, Martins e Barbosa (2008) no desenvolvimento de frutos Longan (*Dimocarpus longan* lour) pelo modelo Logístico com uma exponencial quadrática

2.3 Método de estimação por mínimos quadrados

Gallant (1987) afirma que o método dos quadrados mínimos é utilizado na estimação dos parâmetros em modelos não-lineares, da mesma maneira que em modelos lineares. Seja uma equação de regressão não-linear dada por:

$$Y_t = f(X_t, \theta) + e_t$$

em que: $t = 1, 2, 3, \dots, n$.

Assumindo-se que $E(e_t) = 0$, $\text{Var}(e_t) = \sigma^2$ e $e_t \sim N(0, \sigma^2)$, esta função pode ser reescrita na forma matricial

$$Y = f(\theta) + e$$

$$\text{em que: } Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ Y_n \end{bmatrix}; f(\theta) = \begin{bmatrix} f(X_1, \theta) \\ f(X_2, \theta) \\ \cdot \\ f(X_n, \theta) \end{bmatrix}; e = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ e_n \end{bmatrix}.$$

A estimativa para θ é dada pelo vetor $\hat{\theta}$ que minimiza a soma de quadrados do resíduo,

$$S(\theta) = \sum_{t=1}^n [Y_t - f(X_t; \theta)]^2$$

Escrevendo-se $S(\theta)$ na forma matricial, tem-se:

$$S(\theta) = [Y - f(\theta)]' [Y - f(\theta)]$$

Derivando $S(\theta)$ em relação à θ , igualando-se a zero obtém-se o sistema de equações normais (SEN) não-linear, onde não é possível obter uma solução explícita para os parâmetros. Assim, as estimativas dos parâmetros são obtidas por meio da utilização de métodos iterativos.

2.3.1 Método dos mínimos quadrados generalizados

Conforme apresentado por Hoffman e Vieira (1998), o método dos mínimos quadrados generalizados é mais eficiente do que o método dos mínimos quadrados ordinários, na presença de heterogeneidade de variâncias e autocorrelação residual.

Para o modelo linear matricial:

$$Y = X\beta + u$$

supondo-se que $u \sim N(0; W\sigma^2)$, em que W é uma matriz simétrica, positiva definida, que representa as variâncias e covariâncias dos erros. Admitindo-se que os erros são autocorrelacionados na forma de um processo autorregressivo estacionário de primeira ordem AR (1),

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

em que t é o tempo medido em horas, dias, semanas, etc e tem-se que $E(\xi_t) = 0$, $E(\xi_t^2) = \sigma_\varepsilon^2$, $E(\xi_t \xi_{t-h}) = 0$ se $h \neq 0$.

O modelo u_t será estacionário se

$$-1 < \phi_1 < +1$$

para $t = 1, 2, \dots, n$.

$$\text{Nessas condições, } \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \text{ e } \text{Cov}_u = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \cdot \phi_1^h = \sigma_u^2 \phi_1^h.$$

Segundo Draper e Smith (1998) e Hoffman e Vieira (1998), o vetor de estimativas dos parâmetros é dado por:

$$\hat{\beta} = (X'W^{-1}X)^{-1}X'W^{-1}Y, \text{ em que}$$

$$W = \frac{\sigma_{\varepsilon}^2}{1-\phi_1^2} \begin{bmatrix} 1 & \phi_1 & \phi_1^2 & \dots & \phi_1^{n-1} \\ \phi_1 & 1 & \phi_1 & \dots & \phi_1^{n-2} \\ \phi_1^2 & \phi_1 & 1 & \dots & \phi_1^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1^{n-1} & \phi_1^{n-2} & \phi_1^{n-3} & \dots & 1 \end{bmatrix}$$

Outros modelos possíveis para o erro são o de médias móveis (MA) e o modelo misto, autorregressivo e médias móveis (ARMA).

2.4 Métodos de simulação

Às vezes é impossível obter expressões explícitas para os estimadores de máxima verossimilhança para o caso de modelos de regressão não-linear. A inferência exata somente será possível se as funções puderem ser calculadas analiticamente, caso contrário os métodos aproximados deverão ser utilizados. Os métodos aproximados são baseados em simulação estocástica que consistem no tratamento de problemas reais a partir de reproduções em ambientes controlados pelo pesquisador. Alguns problemas apresentam componentes aleatórios, os quais não podem ser descritos de forma exata e sim baseados em informações probabilísticas. Para estes problemas, o processo de simulação é estocástico, ou seja, baseado em distribuições de probabilidades.

A inferência baseada em técnicas de simulação utiliza amostras da distribuição de interesse, no caso da inferência clássica, a função de verossimilhança, para extrair informação a seu respeito. Como uma amostra é um substituto parcial da informação contida em uma densidade, os métodos

baseados em simulação são inerentemente aproximados e devem ser utilizados quando for constatada a impossibilidade de extração analítica de informação da distribuição (PAULINO; TURKMAN; MURTEIRA, 2003). Dada uma amostra (y_1, y_2, \dots, y_n) a função de densidade de probabilidade conjunta vista como uma função do vetor de parâmetros desconhecido é denominada de função de verossimilhança, dada por:

$$L(\theta|y) = \prod_{t=1}^n f(y_t, \theta)$$

Existem métodos de simulação em que as amostras são obtidas da distribuição de interesse em um único passo, chamados de métodos não iterativos, são eles os métodos de Monte Carlo (ordinário e via função de importância) e os métodos de Reamostragem (método da rejeição e método da reamostragem ponderada) (EHLERS, 2007). Os métodos que fazem uso de técnicas de simulação iterativa, baseadas em cadeias de Markov são os métodos de Monte Carlo via cadeias de Markov (métodos MCMC) sendo que os mais utilizados são o amostrador de Gibbs e o algoritmo de Metropolis-Hastings. Os métodos de simulação de Monte Carlo, apesar de intensamente utilizados em inferência bayesiana, podem ser utilizados na inferência clássica sem restrições como um meio de amostrar a função de verossimilhança obtendo como resultados estimativas por ponto e por intervalo dos parâmetros.

2.4.1 Método de Monte Carlo via cadeias de Markov

A idéia básica por detrás desses métodos é a de transformar o problema estático, em consideração, num problema de natureza dinâmica, construindo para tal efeito um processo estocástico temporal, artificial, que seja fácil de simular, e que convirja para a distribuição original (EHLERS, 2007).

Metropolis et al. (1953) desenvolveram um algoritmo para esse efeito, o qual foi mais tarde generalizado por Hastings (1970), sendo atualmente conhecido na literatura por algoritmo de Metropolis-Hastings. A associação destes algoritmos para a simulação de distribuições com o método de Monte Carlo para aproximação de integrais conduz aos chamados Métodos de Monte Carlo via Cadeias de Markov (MCMC).

Um caso particular destes métodos é o método de amostragem Gibbs, o qual foi introduzido por Geman e Geman (1984) para simular distribuições multivariadas, de natureza bastante complexas, que surgem em problemas de reconstrução de imagens. Gelfand e Smith (1990) mostraram como o algoritmo pode ser usado para simular distribuições *a posteriori* e, conseqüentemente, a sua importância na resolução de problemas em inferência Bayesiana.

A inferência sobre os parâmetros consiste em determinar *a posteriori* conjunta do modelo, com o objetivo de encontrar uma distribuição para um parâmetro específico, chamada distribuição marginal de θ_i . Para isso, integra-se *a posteriori* conjunta em relação aos outros parâmetros do modelo. Porém, muitas vezes a forma analítica da marginal é complexa e multidimensional ou até mesmo impossível de ser calculada analiticamente. Nesse caso, utilizam-se métodos numéricos iterativos como método de Monte Carlo via Cadeias de Markov, (MCMC) para gerar valores de uma distribuição condicional *a posteriori* para cada parâmetro, pois tais valores aproximam das distribuições marginais (PAULINO; TURKMAN; MURTEIRA, 2003). O estudo descrito acima é feito utilizando-se a função de verossimilhança, ao invés da *posteriori*, no caso da inferência clássica.

2.4.1.1 Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings usa a idéia de que um valor é gerado de uma distribuição auxiliar e aceito com uma dada probabilidade. Este mecanismo de correção garante a convergência da cadeia para a distribuição de equilíbrio (EHLERS, 2007).

Suponha que a cadeia esteja no estado θ e um valor θ' é gerado de uma distribuição proposta $q(\cdot|\theta)$. Note que a distribuição proposta pode depender do estado atual da cadeia, θ . O novo valor θ' é aceito com probabilidade

$$\alpha(\theta, \theta') = \min \left(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right),$$

em que: π é a distribuição de interesse, no caso da inferência clássica é a função de verossimilhança.

2.4.1.2 Amostrador de Gibbs

O algoritmo Gibbs, ou método de amostragem Gibbs, é baseado no fato de que se a função de verossimilhança representar o núcleo de uma distribuição conhecida, e os estimadores dos parâmetros é de difícil obtenção, então a função pode ser amostrada diretamente a partir das distribuições condicionais completas para os parâmetros do modelo, $p(\theta_i|x, \theta(-i))$, $i = 1, \dots, k$, em que $\theta = (\theta_1, \dots, \theta_k)$ e $\theta(-i) = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$, ou seja, é o vetor θ sem a i -ésima componente (PAULINO; TURKMAN; MURTEIRA, 2003).

O mesmo autor ainda descreve o algoritmo Gibbs da seguinte maneira. Seja $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$ um valor arbitrário inicial para o vetor θ , então:

1) obtém-se $\theta_1^{(1)}$ de $p(\theta_1|x, \theta_2^{(0)}, \dots, \theta_k^{(0)})$,

...

obtém-se $\theta_k^{(1)}$ de $p(\theta_k|x, \theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \dots, \theta_{k-1}^{(1)})$.

Completa-se assim uma iteração do esquema e uma transição de θ^0 para $\theta^1 = (\theta_1^{(1)}, \dots, \theta_k^{(1)})$.

2) O esquema anterior é repetido com θ^1 , obtido anteriormente, como vetor inicial para obter um novo vetor θ^2 e assim haver uma transição de θ^1 para $\theta^2 = (\theta_1^{(2)}, \dots, \theta_k^{(2)})$,

3) Repete-se t vezes este ciclo de geração de observações aleatórias de cada uma das distribuições condicionais, produzindo assim $\theta^0, \dots, \theta^t$.

A sucessão $\theta^0, \dots, \theta^t, \dots$ é uma realização de uma cadeia de Markov com espaço de estados, Θ , o espaço paramétrico. Resultados teóricos permitem concluir que, quando $t \rightarrow \infty$, $(\theta_1^{(t)}, \dots, \theta_k^{(t)})$ tende em distribuição para um vetor aleatório cuja função densidade de probabilidade conjunta é $p(\theta|x)$. Em particular, $\theta_i^{(t)}$ tende em distribuição para uma quantidade aleatória cuja densidade é $p(\theta_i|x)$, que é a densidade marginal do parâmetro e,

$\frac{1}{t} \sum_{j=1}^t \theta_i^j \rightarrow E(\theta_i)$, quase certamente, em que: $E(\theta_i)$ representa o valor esperado de θ_i .

De acordo com a teoria exposta, o método de amostragem Gibbs permite, por meio de um processo iterativo, gerar realizações de uma cadeia de Markov de tal modo que, à medida que o número de iterações aumenta, a cadeia se aproxima da sua condição de equilíbrio. No entanto, sucessivas realizações de

uma mesma cadeia ao longo do tempo, não constituem uma amostra aleatória da distribuição que se pretende gerar. Com efeito, os vetores θ^j que vão sendo gerados são correlacionados. A questão que se põe é como utilizar as realizações da cadeia, após o equilíbrio, para fazer inferências sobre θ , ou qualquer outra função de θ . Há essencialmente dois métodos em disputa. A teoria sugere que as inferências devem ser feitas com base nos resultados obtidos de uma única realização da cadeia durante um longo período. Por outro lado, gerações múltiplas de cadeias independentes também podem ser utilizadas para obter uma amostra independente e identicamente distribuída, i.i.d.

2.4.2 Diagnóstico de convergência dos Métodos de Monte Carlo via cadeias de Markov

Como descrito anteriormente, a idéia dos métodos de Monte Carlo via cadeias de Markov é a de transformar o problema estático, em consideração, num problema de natureza dinâmica, construindo para isto, um processo estocástico temporal, artificial, que seja fácil de simular, e que convirja para a distribuição original (PAULINO; TURKMAN; MURTEIRA, 2003). A questão é como monitorar a convergência das cadeias.

Alguns métodos, ou estratégias, são utilizados para diagnosticar esta convergência sendo eles classificados como métodos informais e formais. Como método de monitoração informal tem-se o método proposto por Gelfand e Smith (1990), baseado na visualização gráfica da cadeia gerada, e como métodos de monitoração formal têm-se, dentre os mais utilizados, aqueles propostos por Gelman e Rubin (1992), Geweke (1992), Heidelberger e Welch (1993) e Raftery e Lewis (1992).

Gelman e Rubin (1992) propuseram um método que se baseia no uso de componentes de variância de seqüências múltiplas da cadeia, simuladas a partir

de pontos dispersos. A popularidade deste método é devido, em grande parte, à simplicidade de sua implementação e ao fato de sua metodologia ter sido extensivamente avaliada. Este método preconiza que sejam simuladas $m \geq 2$ seqüências, cada uma de comprimento $2n$, e com pontos iniciais simulados de uma distribuição sobredispersa com relação à distribuição de equilíbrio. As primeiras n iterações, de cada cadeia, são descartadas para evitar o período de aquecimento (“burn-in”), ou seja, para evitar a influência dos valores iniciais.

Considere ainda uma quantidade escalar de interesse que se pretende estimar, U , geralmente é uma função do parâmetro θ . Designando u_{ij} o valor que U assume na j -ésima iteração da i -ésima cadeia, calcula-se a média e variância, respectivamente, de U para cada uma das m cadeias consideradas como se segue:

$$\bar{u}_i = \frac{1}{n} \sum_{j=1}^n u_{ij}, \quad s_i^2 = \frac{1}{n-1} \sum_{j=1}^n (u_{ij} - \bar{u}_i)^2$$

Calculam-se as seguintes componentes da variância:

$$W = \frac{1}{m} \sum_{i=1}^m s_i^2,$$

a média das m variâncias, cada uma baseada em $n-1$ graus de liberdade, e

$$\frac{B}{n} = \frac{1}{m-1} \sum_{i=1}^m (\bar{u}_i - \bar{u}_{..})^2,$$

a variância entre as médias das m seqüências, cada uma baseada em n valores de U .

Estima-se a média alvo de U como uma média amostral de todos os mn valores simulados de U .

$$\hat{\mu} = \bar{U}_{..} = \frac{1}{m} \sum_{i=1}^m \bar{u}_i.$$

Estima-se a variância alvo de U como uma média ponderada de W e B , como se segue:

$$\hat{\sigma}^2 = \frac{n-1}{n}W + \frac{1}{n}B$$

Cria-se uma distribuição t-Student para U com parâmetro de localização $\hat{\mu}$ de dispersão $\sqrt{\hat{V}}$ e números de graus de liberdade ν , em que:

$$\sqrt{\hat{V}} = \sqrt{\hat{\sigma}^2 + \frac{B}{mn}},$$

$$\nu = \frac{2\hat{V}^2}{\text{var}(\hat{V})},$$

sendo:

$$\begin{aligned} \text{var}(\hat{V}) = & \left(\frac{n-1}{n}\right)^2 \frac{1}{m} \text{var}(s_i^2) + \left(\frac{m+1}{mn}\right)^2 \frac{2}{m-1} B^2 + \\ & + 2 \frac{(m-1)(n-1)}{mn^2} \frac{n}{m} [\text{cov}(s_i^2, \bar{u}_i) - 2\bar{u}_{..} \text{cov}(s_i^2, \bar{u}_i)] \end{aligned}$$

as variâncias e covariâncias são estimadas a partir dos m valores amostrais de s_i^2 , \bar{u}_i e \bar{u}_i^2 .

A comparação entre as variâncias é expressa por:

$$\hat{R} = \frac{\hat{V}}{W} \frac{\nu}{\nu-2},$$

o qual chama-se fator de redução de escala e pode ser interpretado como um fator de diagnóstico de convergência. Quando \hat{R} afasta-se de 1, sugere-se que qualquer das estimativas das variâncias pode ser reduzida com a utilização de mais iterações, desde que as seqüências simuladas ainda não tenham feito um passeio por toda a função de interesse. Se o fator de redução de escala, \hat{R} , aproxima-se de 1, pode-se concluir que cada m conjunto de n iterações se aproxima da distribuição de interesse.

O critério proposto por Raftery e Lewis (1992) é um método que estima quantas iterações são necessárias para que o amostrador de Gibbs apresente convergência à distribuição estacionária, propondo quantidades necessárias para isso. O método fornece as estimativas do *burn-in*, o número total de iterações que devem ser computadas e a distância mínima de uma iteração à outra (*thin*), para se obter a subamostra aproximadamente independente. A regra de decisão do critério de Raftery e Lewis (1992) é baseada no fator de dependência, que é responsável pelo acréscimo multiplicativo ao número de iterações necessárias para se alcançar a convergência. De acordo com os autores, se o fator de dependência for maior que cinco, pode-se concluir que a cadeia não atingiu convergência.

O critério proposto por Geweke (1992) consiste em indicar a convergência da média da quantidade amostrada, $g(\theta)$, em uma única cadeia, isto é, os valores $g(\theta^j)$ são calculados a cada iteração da simulação, formando uma série temporal. Em um primeiro passo, divide-se a cadeia de tamanho N em duas seqüências. A primeira seqüência possui N_a primeiras iterações e a segunda, N_b restantes iterações. O autor sugere que as médias devem ser estimadas após o descarte de algumas iterações iniciais e que sejam considerados $N_a = 0,1N$ e $N_b = 0,5N$. Em um segundo passo, estima-se as

médias \hat{g}_a e \hat{g}_b e as variâncias assintóticas $\hat{S}_{g_a}^2(0)$ e $\hat{S}_{g_b}^2(0)$, respectivamente. As variâncias são determinadas pela estimação da densidade espectral. Se a diferença padronizada entre as médias for muito grande, existe indicação de ausência de convergência. A regra de decisão do critério de Geweke é baseada no valor p . Se esse valor for menor que um nível de significância adotado pelo pesquisador, conclui-se que a cadeia não atingiu convergência.

O critério de Heidelberger e Welch (1993) utiliza testes estatísticos para avaliar a hipótese nula de estacionariedade da amostra gerada. Se a hipótese nula for rejeitada para um dado valor de descarte das observações iniciais, o teste é repetido depois de descartados os 10% valores iniciais da amostra. Se a hipótese é novamente rejeitada, mais 10% dos valores iniciais são descartados, e assim sucessivamente, até serem descartados os 50% valores iniciais. Se a hipótese for novamente rejeitada, isso indica que é necessário um número maior de iterações. Caso contrário, o número de iterações descartadas é indicado como o tamanho do *burn-in*.

Uma avaliação detalhada dos critérios de convergência para o método de Monte Carlo via Cadeias de Markov pode ser encontrado em Nogueira, Sáfyadi e Ferreira (2004). Os autores concluíram que o seguinte procedimento deveria ser utilizado para que se obtivesse uma avaliação mais precisa da convergência: aplicar o critério de Raftery e Lewis (1992) em uma amostra piloto e determinar o tamanho ideal da seqüência; monitorar a convergência das seqüências nas proximidades do tamanho ideal, indicado pelo critério de Raftery e Lewis (1992), por meio dos critérios de Gelman e Rubin (1984) e Geweke (1992) e determinar o tamanho de *burn-in*, pelo critério de Heidelberger e Welch (1993).

2.4.3 Erro Monte Carlo

O método de Monte Carlo é utilizado para o cálculo de integrais definidas,

utilizando-se aproximações. Paulino, Turkman e Murteira (2003), sugerem considerar o problema de aproximar uma integral da forma:

$$\int g(\theta)h(\theta|x)d\theta = E[g(\theta)|x],$$

em que θ e x podem ser vetores, cuja existência se admite. Valores *a posteriori* de interesse podem ser expressos pela equação acima, para algum tipo de função $g(\theta)$ integrável. É o caso de médias *a posteriori*, θ_i , de componentes de um vetor de parâmetros θ . Se for possível simular uma amostra aleatória $(\theta_1, \dots, \theta_m)$ da densidade *a posteriori* $h(\theta|x)$, o Método Monte Carlo aproxima a integral acima pela média empírica, dada por:

$$\hat{E}[g(\theta)|x] = \frac{1}{m} \sum_{i=1}^m g(\theta_i),$$

a qual pela Lei Forte dos Grandes Números, converge quase certamente para $\hat{E}[g(\theta)|x]$. No caso da inferência clássica, a função $h(\theta|x)$ é a função de verossimilhança. A precisão desta aproximação pode ser medida pelo erro padrão de Monte Carlo, dado por:

$$\frac{1}{\sqrt{m(m-1)}} \left\{ \sum_{i=1}^m \left[g(\theta_i) - \frac{1}{m} \sum_{i=1}^m g(\theta_i) \right]^2 \right\}^{1/2},$$

isto é, a aproximação pode ser tão acurada quanto se deseja, bastando aumentar o valor de m (número de iterações).

Este método consiste em gerar vários números aleatórios uniformemente distribuídos no intervalo de interesse, avaliando-se o valor da função a ser

integrada em cada um destes pontos e, finalmente calculando-se a média desses valores obtidos para se obter a aproximação da integral definida. Como o método de Monte Carlo é uma aproximação, não se pode afirmar que o valor encontrado por ele seja o valor exato da integral. Entretanto, quanto mais pontos forem gerados, maior será a confiabilidade na aproximação obtida.

O Erro Monte Carlo utilizado com frequência na inferência bayesiana, pode ser utilizado na inferência clássica, utilizando-se a função de verossimilhança.

3 MATERIAL E MÉTODOS

Os dados utilizados são de um experimento (BENASSI, 2006) em que foi mensurado, dentre outras características, os diâmetros externos, longitudinal (DEL) e transversal (DET), de frutos de coqueiro da variedade anão verde. As avaliações biométricas dos frutos foram realizadas em 26 épocas de colheitas, com intervalo quinzenal, durante o seu crescimento. A primeira avaliação foi realizada no dia da abertura da inflorescência e a última coleta de frutos aos 375 dias após a abertura da inflorescência (d.a.a.i.). As inflorescências, uma por planta, foram marcadas no momento de sua abertura, com uma fita plástica de 1,0 metro de comprimento. Para atingir o número necessário, efetuaram-se três visitas à área, marcando-se um total de 108 inflorescências recém abertas. Desse total, 8 inflorescências, que abriram, foram identificadas e não tiveram seus frutos colhidos ao longo do trabalho. Estes frutos foram numerados e, quinzenalmente, mensurados quanto aos diâmetros externos, longitudinal e transversal. Estes frutos serviram como controle para verificar se ocorreu diferença significativa entre o crescimento dos frutos que permaneceram ligados à planta e aqueles que quinzenalmente foram colhidos e avaliados. Nas 100 inflorescências restantes, previamente marcadas, quinzenalmente, coletou-se 25 frutos, sendo em média 3 por cacho, retirando-se frutos das regiões apical, mediana e basal, objetivando-se maior uniformidade amostral.

Aos dados, foram ajustados os modelos de crescimento não-linear Logístico e Gompertz considerando erros independentes e estrutura de erros auto-regressivos de primeira ordem. Estes modelos são descritos pelas seguintes expressões:

1. Modelo Logístico com erros independentes e autorregressivos de primeira ordem, respectivamente:

$$W_i = \frac{A}{\{1+B[\exp(-Kt_i)]\}} + u_i$$

$$W_i = \frac{A}{\{1+B[\exp(-Kt_i)]\}} + \phi u_{i-1} + \xi_i$$

2. Modelo de Gompertz com erros independentes e autorregressivos de primeira ordem, respectivamente:

$$W_i = A \exp\{-B[\exp(-Kt_i)]\} + u_i$$

$$W_i = A \exp\{-B[\exp(-Kt_i)]\} + \phi u_{i-1} + \xi_i$$

em que:

W expressa o valor da variável dependente (diâmetro externo longitudinal (DEL) e o diâmetro externo transversal (DET));

A corresponde à assíntota do modelo, indicando o valor de estabilização da variável dependente em relação ao tempo;

B corresponde ao parâmetro de locação, sem interpretação biológica;

k determina a taxa de expansão do crescimento;

\exp refere-se à base do logaritmo neperiano;

e_i é o erro experimental com média zero e variância σ^2 ;

t_i refere-se ao tempo de análise da variável dependente (dias), $i = 1, 15, 30, \dots, 375$;

ϕ corresponde ao parâmetro auto-regressivo de primeira ordem;

ξ_i é o ruído branco.

3.1 Procedimentos de ajuste por meio do método de mínimos quadrados

Na verificação da presença de autocorrelação residual foi utilizado o teste de Durbin-Watson (DW) dado por:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

em que: e_t é o resíduo do ajuste no tempo t , e_{t-1} é o resíduo do ajuste no tempo imediatamente anterior a t . Foram obtidas estimativas dos parâmetros do modelo em estudo considerando a estrutura de autocorrelação residual.

Conforme visto anteriormente, o sistema de equações normais (SEN) não linear não apresenta solução explícita sendo necessário o uso de métodos iterativos. Como padrão será utilizado o método de Gauss-Newton, na forma original, que consiste no desenvolvimento em série de Taylor até o termo de primeira ordem da função, $f(X_i, \theta)$ em torno do ponto θ^0 , cuja idéia geral é feita considerando o modelo não-linear:

$$Y_i = f(X_i; \theta) + e_i,$$

A expansão em série de Taylor é dada por:

$$f(\theta) = f(\theta^0) + F(\theta^0)(\theta - \theta^0).$$

Assim, o SEN (não-linear),

$$X' f(\hat{\theta}) = X' Y$$

pode ser escrito como: $X' [f(\theta^0) + F(\theta^0)(\theta - \theta^0)] = X' Y$,

em que: $F(\theta^0)$ é a matriz de derivadas parciais X do modelo. Logo, substituindo-se no SEN, obtém-se:

$$X' [f(\theta^0) + X(\theta - \theta^0)] = X' Y$$

Fazendo-se o produto matricial e reagrupando termos semelhantes, encontra-se:

$$\theta - \theta^0 = (X'X)^{-1}X'e.$$

Portanto, a fórmula iterativa conhecida como Método de Gauss-Newton é dada por:

$$\theta^1 = \theta^0 + (X'X)^{-1}X'e$$

Com base em um valor inicial para θ^0 , este processo é repetido colocando θ^1 no lugar de θ^0 (vetor de estimativas iniciais) até que algum critério de convergência seja aceito, isto é, $\theta^n \rightarrow \hat{\theta}$ em que n é o número de vezes que o processo foi repetido.

A seleção do modelo adequado para explicar o crescimento do fruto terá como base a precisão dos ajustes e obedecerá aos seguintes critérios:

- a. Coeficiente de determinação ajustado (R^2_{aj}), dado por:

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

em que: R^2 é o coeficiente de determinação não ajustado, n é o número de observações e p é o número de parâmetros do modelo.

- b. Desvio padrão residual (DPR), dado por:

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

em que: QME é o quadrado médio residual, n é o número de observações e p é o número de parâmetros do modelo;

- c. Critério de Informação de Akaike (AIC), dado por:

$$AIC = \ln(QME) + 2(p+1)/n$$

em que: n é o número de observações, QME é o quadrado médio residual e p é o número de parâmetros do modelo;

Intervalos de 95% de confiança foram construídos para os parâmetros dos modelos em estudo e de acordo com Draper e Smith (1998), as estimativas da matriz assintótica de variâncias e covariâncias estrutura-se da seguinte forma:

$$\hat{V}(\hat{\theta}) = (X'X)^{-1} \hat{\sigma}^2 = (X'X)^{-1} QME$$

em que: X é a matriz de derivadas parciais do modelo e QME é o quadrado médio residual. Logo, o erro padrão da estimativa de determinado parâmetro do modelo, θ_j , é dado por:

$$s(\hat{\theta}_j) = \sqrt{\hat{V}(\hat{\theta}_j)}$$

em que: $\hat{V}(\hat{\theta}_j)$ é um elemento da diagonal principal da matriz assintótica de variâncias e covariâncias.

Desta forma, define-se o intervalo de confiança para o parâmetro, θ_j , como:

$$IC(\theta_j) = \hat{\theta}_j \pm t_{(gl.erro, \alpha/2)} \sqrt{\hat{V}(\hat{\theta}_j)}$$

em que: $t_{(gl.erro, \alpha/2)}$ é o percentil superior $\alpha/2$ da distribuição t de Student com graus de liberdade residual.

Os procedimentos de ajuste acima mencionados foram implementados por meio dos programas computacionais R (R DEVELOPMENT CORE TEAM, 2008) e SAS® (STATISTICAL ANALYSIS SYSTEM INSTITUTE - SAS INSTITUTE, 1999).

3.2 Procedimentos de ajuste por meio do método de Monte Carlo via Cadeias de Markov

De modo semelhante ao realizado na estimação de parâmetros pelo método de mínimos quadrados, no ajuste dos modelos e estimação de parâmetros pelo método de amostragem da função de verossimilhança foi verificada a presença de autocorrelação residual, por meio do teste de Durbin-Watson. Os resultados dos testes para autocorrelação residual orientaram a maneira de se construir a função de verossimilhança, e que foi posteriormente amostrada por meio dos métodos de Monte Carlo via cadeias de Markov.

A convergência das cadeias de cada um dos parâmetros do modelo foi monitorada pela visualização gráfica da cadeia gerada pelo MCMC e pelos critérios Geweke (1992), Heidelberger e Welch (1993) e Raftery e Lewis (1992).

Foram obtidas estimativas pontuais (média, moda) e intervalares por meio do intervalo de credibilidade de máxima densidade de probabilidade (*highest probability density*-HPD) para os parâmetros dos modelos. Ehlers (2007) define C como um intervalo de credibilidade de $100(1-\alpha)\%$ para um parâmetro qualquer θ se $P(\theta \in C) \geq 1-\alpha$. Assim, quanto menor for o tamanho do intervalo, mais concentrada é a distribuição do parâmetro, ou seja, o tamanho do intervalo informa sobre a dispersão de θ .

De acordo com a definição anterior, é possível construir uma infinidade de intervalos, mas são de principal interesse aquele com menor comprimento possível. Os intervalos de comprimento mínimo são obtidos tomando-se valores de θ com maior densidade, denominados intervalos de credibilidade de máxima densidade de probabilidade ou HPD.

Existem diversos critérios, descritos na literatura, para a seleção de modelos. Alguns deles são: Critério de Informação de Akaike (AIC), Critério de Informação Deviance (DIC).

4 RESULTADOS E DISCUSSÃO

4.1 Ajuste pelo método de mínimos quadrados

No ajuste dos modelos e estimação dos parâmetros, utilizou-se o PROC MODEL do software Statistical Analysis System e a seleção do modelo mais adequado teve como base a precisão dos ajustes obedecendo aos seguintes critérios: coeficiente de determinação (R^2_{aj}) ajustado, desvio padrão residual (DPR), critério de informação de Akaike (AIC), teste de Durbin-Watson (DW) e número de iterações.

A estatística do teste de Durbin-Watson, que testa a hipótese nula de que os resíduos não são correlacionados contra a hipótese alternativa de que os resíduos seguem um processo autorregressivo de primeira ordem (AR1), varia em um intervalo de valores entre de 0 a 4 sendo que valores em torno de 0 indicam a presença autocorrelação positiva. Para o ajuste dos modelos, Logístico e Gompertz considerando erros independentes, o valor da estatística DW encontrava-se em torno do valor zero (0,38 e 0,50 para os modelos Logístico e Gompertz respectivamente, considerando o diâmetro externo longitudinal (DEL); 0,23 e 0,29 para o Logístico e Gompertz, considerando o diâmetro externo transversal (DET)) indicando a necessidade de considerar um processo autorregressivo (AR) para a série de resíduos. A escolha da ordem do processo AR a ser ajustado foi feita a partir da significância das estimativas dos parâmetros autorregressivos e segundo esse critério (Tabela 1) para ambos os modelos, considerou-se um processo autorregressivo de primeira ordem.

Tabela 1 Avaliadores da qualidade de ajuste dos modelos Logístico e Gompertz aos dados de diâmetro, longitudinal (DEL) e transversal (DET), de frutos de coqueiro anão verde considerando a estrutura de erros autorregressivos (AR)

Avaliadores	DEL		DET	
	Logístico	Gompertz	Logístico	Gompertz
Parâmetro A	21,43	24,33	15,81	18,37
Parâmetro B	11,07	2,70	8,33	2,25
Parâmetro k	0,0146	0,0075	0,0131	0,0062
ϕ (<i>valor p</i>)	0,63 ($p \leq 0,0016$)	0,85 ($p \leq 0,0001$)	0,88 ($p \leq 0,0001$)	0,92 ($p \leq 0,0001$)
R_{aj}^2	0,9968	0,9954	0,9916	0,9914
DPR	0,0797	0,0956	0,0947	0,0953
DW	1,771	1,537	1,396	2,203
AIC	-1,5929	-1,2198	-1,2389	-1,2248

Os resultados da Tabela 1 indicam que, tanto para o diâmetro longitudinal como para o diâmetro transversal dos frutos, o modelo Logístico é considerado o mais adequado para descrever o crescimento de frutos de coqueiro da variedade anã verde, pois apresentaram o menor valor para *AIC* e *DPR* e melhores valores para o R_{aj}^2 . O teste de Durbin-Watson, após incorporar a estrutura de erro AR(1) nos modelos, resultou em independência para os resíduos e somente para o caso DET do modelo Logístico o teste foi inconclusivo (Tabela 1).

Com relação às estimativas dos parâmetros A, B e k do modelo Logístico, verificou-se que os frutos à maturação apresentam diâmetro longitudinal em torno de 21,5 cm e diâmetro transversal em torno de 15,8 cm.

Os intervalos de confiança para os parâmetros do modelo Logístico são apresentados para o DEL e DET, na Tabela 2:

Tabela 2 Intervalos de 95% de confiança com para os parâmetros do modelo Logístico para o diâmetro externo longitudinal (DEL) e diâmetro externo transversal (DET)

Parâmetros	DEL		DET	
	LI	LS	LI	LS
A	20,21	22,65	12,88	18,73
B	8,32	13,82	4,23	12,43
k	0,01	0,02	0,008	0,02
ϕ	0,28	0,98	0,61	1,00

O ajuste do modelo Logístico aos dados de diâmetro, longitudinal e transversal, dos frutos de coqueiro da variedade anã pode ser visualizado no Gráfico 1.

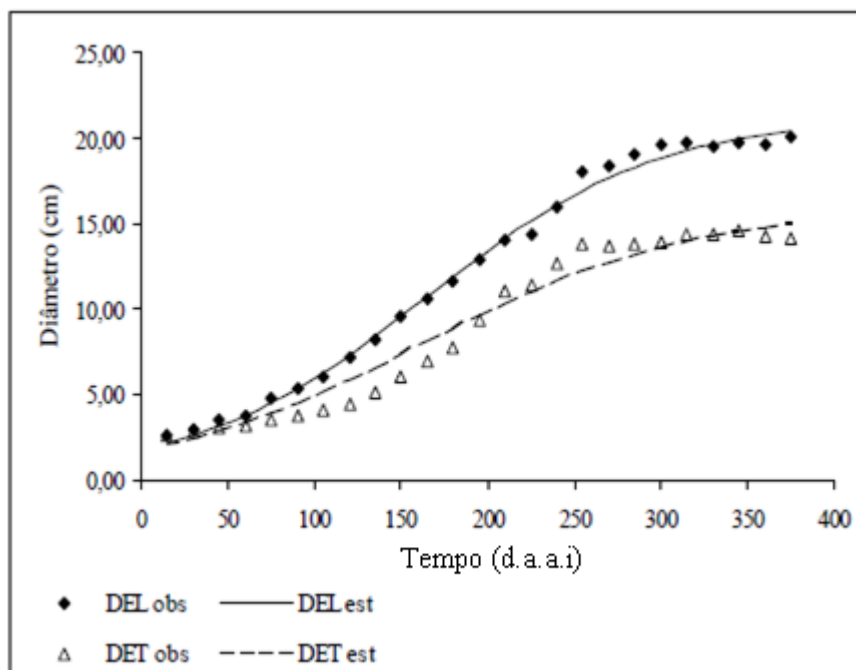


Gráfico 1 Diâmetros médios, longitudinal (DEL) e transversal (DET), observados e estimados por meio do modelo Logístico com estrutura de erros autorregressivos de primeira ordem

Inicialmente os cocos crescem mais em comprimento do que em largura, ficando relativamente alongados e compridos, ou seja, diâmetro transversal menor que o diâmetro longitudinal (GOMES, 1984). O mesmo autor também afirma que o coco adquire tamanho máximo quando tem cerca de seis meses de idade, mantendo-se constante durante um ou dois meses, com posterior decréscimo. No presente estudo verifica-se que os diâmetros dos frutos começam a estabilizar mais tardiamente, ou seja, para o diâmetro transversal a estabilização ocorre por volta dos 250 dias (8 meses) e para o diâmetro longitudinal em torno de 300 dias (10 meses).

Magalhães (1999), estudando o diâmetro de frutos de coqueiro da variedade anã verde colhidos aos oito meses de idade relata valores de 18,2cm de diâmetro longitudinal e 15,7cm de diâmetro transversal. Corrêa et al. (2002), avaliando frutos de coqueiro da mesma variedade, obtiveram valores médios de 16,6 cm e 13,1 cm para o comprimento e largura dos frutos, respectivamente. Valores menores aos encontrados neste estudo.

As estimativas obtidas com base no modelo ajustado além da interpretação biológica dos parâmetros estão em conformidade com a descrição das características das variedades de coqueiros, na literatura científica, onde estão relatados que, frutos do coqueiro anã são menores que os frutos dos coqueiros híbridos ou gigantes (CAMBOIM NETO, 2002).

4.2 Ajuste por meio do método de Monte Carlo via cadeias de Markov

Considerando a estrutura de erro independente e autorregressivo (AR1) obtém-se a função de verossimilhança para os dados fazendo $\tau = 1/\sigma^2$. Para o modelo Logístico, as funções de verossimilhança com erro independente e autorregressivo AR(1), são dadas, respectivamente por:

$$L(A, B, k, \tau | Y) = \tau^{n/2} \exp \left\{ \frac{-\tau}{2} \sum_{i=1}^n \left(Y_i - \frac{A}{1 + B e^{-kt_i}} \right)^2 \right\}$$

$$L(A, B, k, \phi, \tau | Y) = \tau^{n/2} \exp \left\{ \frac{-\tau}{2} \sum_{i=2}^n \left(Y_i - \frac{A}{1 + B e^{-kt_i}} - \phi u_{i-1} \right)^2 \right\}$$

Da mesma forma, para o modelo Gompertz, as funções de verossimilhança com erro independente e autorregressivo AR(1), são dadas, respectivamente por:

$$L(A, B, k, \tau | Y) = \tau^{n/2} \exp \left\{ \frac{-\tau}{2} \sum_{i=1}^n \left(Y_i - A e^{-B e^{-k t_i}} \right)^2 \right\}$$

$$L(A, B, k, \tau | Y) = \tau^{n/2} \exp \left\{ \frac{-\tau}{2} \sum_{i=2}^n \left(Y_i - A e^{-B e^{-k t_i}} \right)^2 - \phi u_{i-1} \right\}$$

Assim as distribuições conjuntas dos parâmetros, em cada estrutura de erro, podem ser denotadas por:

$$L(A, B, k, \tau | Y) = P(A, B, k, \tau | Y)$$

$$L(A, B, k, \phi, \tau | Y) = P(A, B, k, \tau, \phi | Y)$$

Portando, tem-se as distribuições conjuntas dos parâmetros e a partir destas foram obtidas as distribuições condicionais completas para cada parâmetro, nas quais foi aplicada a amostragem Gibbs e o algoritmo Metrópolis-Hastings, para obter aproximações das distribuições marginais dos parâmetros. A distribuição auxiliar utilizada no algoritmo Metropolis-Hastings foi uma distribuição normal cuja média e variância foi baseada nos resultados do método de mínimos quadrados apresentados na tabela 1. Para o estudo considerando a estrutura de erro autorregressivo AR(1), a distribuição condicional encontrada para o parâmetro ϕ , que foi de uma distribuição normal, foi truncada no intervalo de -1 a 1, que é o intervalo de interesse para o parâmetro autorregressivo AR(1).

Os erros u_t foram obtidos a partir do ajuste dos modelos assumindo erros independentes. O erro independente gerado pelo ajuste foi tomado como a sequência u_t , perdendo a primeira observação visto que o AR é de primeira ordem.

4.2.1 Obtenção das distribuições condicionais completas para os parâmetros do modelo Logístico

A partir da distribuição conjunta dos parâmetros do modelo Logístico com estrutura de erro independente, a distribuição condicional para o parâmetro A , considerando que os termos que não dependem de A são termos constantes para este parâmetro, pode ser escrita como se segue:

$$P(A | B, k, \tau, Y) \propto \exp \left\{ \frac{-\tau}{2} \sum_{i=1}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} \right)^2 \right\}$$

Desenvolvendo-se o quadrado acima, tem-se:

$$\propto \exp \left\{ \frac{-\tau}{2} \sum_{i=1}^n \left[Y_i^2 - 2A \frac{Y_i}{1 + Be^{-kt_i}} + A^2 \frac{1}{(1 + Be^{-kt_i})^2} \right] \right\}$$

Rearranjando os termos e retirando os termos que não dependem do parâmetro A , tem-se:

$$\propto \exp \left\{ \frac{-\tau \sum_{i=1}^n \frac{1}{(1 + Be^{-kt_i})^2}}{2} \left(A - \frac{\sum_{i=1}^n \frac{Y_i}{1 + Be^{-kt_i}}}{\sum_{i=1}^n \frac{1}{(1 + Be^{-kt_i})^2}} \right)^2 \right\}$$

Assim, pelo resultado acima, encontra-se o núcleo de uma distribuição normal para o parâmetro A com média e variância dadas por:

$$N \left(\frac{\sum_{i=1}^n \frac{Y_i}{1 + Be^{-kt_i}}}{\sum_{i=1}^n \frac{1}{(1 + Be^{-kt_i})^2}}, \frac{\sigma^2}{\sum_{i=1}^n \frac{1}{(1 + Be^{-kt_i})^2}} \right)$$

Analogamente, a distribuição condicional completa para o parâmetro τ pode ser obtida, como se segue:

$$P(\tau | A, B, k, Y) \propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} \right)^2 \right\}$$

Assim, a expressão acima representa o núcleo de uma distribuição gama com seguintes parâmetros:

$$Gama \left(1 + \frac{n}{2}; (1/2) \sum_{i=1}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} \right)^2 \right)$$

A distribuição condicional completa para os parâmetros B e k podem ser escritas de forma proporcional, como se segue:

$$P(B | A, k, \tau, Y) \propto \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} \right)^2 \right\}$$

$$P(k | A, B, \tau, Y) \propto \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} \right)^2 \right\}$$

As expressões acima não apresentam uma distribuição de forma conhecida então foi necessária a utilização do algoritmo Metropolis-Hastings para os parâmetros B e k.

A partir da distribuição conjunta dos parâmetros do modelo Logístico com estrutura de erro autorregressivo (AR1), a distribuição condicional completa para o parâmetro A do modelo pode ser escrita como se segue:

$$P(A|B, k, \tau, Y, \phi) \propto \exp \left\{ \frac{-\tau}{2} \sum_{i=2}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} - \phi u_{i-1} \right)^2 \right\}$$

Fazendo $Y_i^* = Y_i - \phi u_{i-1}$ e desenvolvendo o quadrado da expressão acima, obtém-se:

$$\propto \exp \left\{ \frac{-\tau}{2} \sum_{i=2}^n \left(Y_i^{*2} - 2Y_i^* \frac{A}{1 + Be^{-kt_i}} + \left(\frac{A}{1 + Be^{-kt_i}} \right)^2 \right) \right\}$$

Assim, retirando-se os termos que não dependem do parâmetro A, ou seja, termos constantes, tem-se:

$$\propto \exp \left\{ \frac{-\tau}{2} \left(-2A \sum_{i=2}^n \frac{Y_i^*}{1 + Be^{-kt_i}} + A^2 \sum_{i=2}^n \frac{1}{(1 + Be^{-kt_i})^2} \right) \right\}$$

Rearranjando os termos da expressão anterior, segue:

$$\propto \exp \left\{ \frac{-\tau \sum_{i=2}^n \frac{1}{1 + Be^{-kt_i}}}{2} \left(A - \frac{\sum_{i=2}^n \frac{Y_i^*}{1 + Be^{-kt_i}}}{\sum_{i=2}^n \frac{1}{(1 + Be^{-kt_i})^2}} \right)^2 \right\}$$

Substituindo a expressão $Y_i - \phi u_{i-1}$ em Y_i^* , a distribuição condicional resulta em:

$$\propto \exp \left\{ \frac{-\tau \sum_{i=2}^n \frac{1}{1 + Be^{-kt_i}}}{2} \left(A - \frac{\sum_{i=2}^n \frac{Y_i - \phi u_{i-1}}{1 + Be^{-kt_i}}}{\sum_{i=2}^n \frac{1}{(1 + Be^{-kt_i})^2}} \right)^2 \right\}$$

Assim encontramos o núcleo de uma distribuição normal para o parâmetro A , com média e variância dadas abaixo:

$$N \left(\frac{\sum_{i=2}^n \frac{Y_i - \phi u_{i-1}}{1 + Be^{-kt_i}}}{\sum_{i=2}^n \frac{1}{(1 + Be^{-kt_i})^2}}, \frac{\sigma^2}{\sum_{i=2}^n \frac{1}{1 + Be^{-kt_i}}} \right)$$

Analogamente, a distribuição condicional completa para o parâmetro ϕ pode ser escrita da forma proporcional, como se segue:

$$P(\phi | A, B, k, \tau, Y) \propto \exp \left\{ \frac{-\tau}{2} \sum_{i=2}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} - \phi u_{i-1} \right)^2 \right\}$$

Fazendo $Y_i^* = Y_i - \frac{A}{1 + Be^{-kt_i}}$ e desenvolvendo o quadrado da expressão acima, tem-se:

$$\propto \exp \left\{ \frac{-\tau}{2} \sum_{i=2}^n \left(Y_i^{*2} - 2Y_i^* \phi u_{i-1} + \phi^2 u_{i-1}^2 \right) \right\}$$

Aplicando o somatório em cada um dos termos, tem-se:

$$\propto \exp \left\{ \frac{-\tau}{2} \left(\sum_{i=2}^n Y_i^{*2} - 2\phi \sum_{i=2}^n Y_i^* u_{i-1} + \phi^2 \sum_{i=2}^n u_{i-1}^2 \right) \right\}$$

Retirando os termos constantes, ou seja, os termos que não dependem do parâmetro ϕ e rearranjando a expressão, obtém-se:

$$\propto \exp \left\{ \frac{-\tau \sum_{i=2}^n u_{i-1}^2}{2} \left(\phi - \frac{\sum_{i=2}^n Y_i^* u_{i-1}}{\sum_{i=2}^n u_{i-1}^2} \right)^2 \right\}$$

Assim substituindo $Y_i - \frac{A}{1 + Be^{-kt_i}}$ por Y_i^* , fixado inicialmente, segue

que:

$$\propto \exp \left\{ \frac{-\tau \sum_{i=2}^n u_{i-1}^2}{2} \left(\phi - \frac{\sum_{i=2}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} \right) u_{i-1}}{\sum_{i=2}^n u_{i-1}^2} \right)^2 \right\}$$

Portanto, encontramos para o parâmetro ϕ o núcleo de uma distribuição normal com média e variância dadas abaixo:

$$N \left(\frac{\sum_{i=2}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} \right) u_{i-1}}{\sum_{i=2}^n u_{i-1}^2}, \frac{\sigma^2}{\sum_{i=2}^n u_{i-1}^2} \right)$$

Da mesma forma, a distribuição condicional completa para o parâmetro τ pode ser escrita da forma proporcional:

$$P(\tau | A, B, k, Y, \phi) \propto \tau^{\frac{n}{2}} \exp \left\{ \frac{-\tau}{2} \sum_{i=2}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} - \phi u_{i-1} \right)^2 \right\}$$

Assim, obtém-se o núcleo de uma distribuição gama com seguintes parâmetros:

$$Gama\left(1 + \frac{n}{2}; (1/2) \sum_{i=2}^n \left(Y_i - \frac{A}{1 + Be^{-kt_i}} - \phi u_{i-1}\right)^2\right)$$

Para os parâmetros B e k, também não foi possível encontrar uma distribuição conhecida, utilizando-se do algoritmo Metropolis-Hastings.

4.2.2 Obtenção das distribuições condicionais completas para os parâmetros do modelo Gompertz

A partir da função de verossimilhança do modelo Gompertz com estrutura de erro independente, a distribuição condicional para o parâmetro A, pode ser escrita como se segue:

$$P(A | B, k, \tau, Y) \propto \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n \left(Y_i - Ae^{-Be^{-kt_i}}\right)^2\right\}$$

Desenvolvendo o quadrado da expressão acima, tem-se:

$$\propto \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n \left(Y_i^2 - 2Y_i Ae^{-Be^{-kt_i}} + A^2 e^{-2Be^{-kt_i}}\right)\right\}$$

Retirando-se os termos que não dependem do parâmetro A e rearranjando os termos acima, segue que:

$$\propto \exp\left\{-\frac{\tau \sum_{i=1}^n e^{-2Be^{-kt_i}}}{2} \left(A - \frac{\sum_{i=1}^n Y_i e^{-Be^{-kt_i}}}{\sum_{i=1}^n e^{-2Be^{-kt_i}}}\right)^2\right\}$$

A expressão acima representa o núcleo de uma distribuição normal para o parâmetro A:

$$N \left(\frac{\sum_{i=1}^n Y_i e^{-Be^{-kt_i}}}{\sum_{i=1}^n e^{-2Be^{-kt_i}}}; \frac{\sigma^2}{\sum_{i=1}^n e^{-2Be^{-kt_i}}} \right)$$

Analogamente, a distribuição condicional completa para o parâmetro τ é dada:

$$P(\tau | A, B, k, Y) \propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \left(Y_i - Ae^{-Be^{-kt_i}} \right)^2 \right\}$$

A distribuição condicional completa acima representa o núcleo de uma distribuição gama com os seguintes parâmetros:

$$Gama \left(1 + \frac{n}{2}; (1/2) \sum_{i=1}^n \left(Y_i - Ae^{-Be^{-kt_i}} \right)^2 \right)$$

No caso dos parâmetros B e k não foi possível encontrar uma distribuição conhecida. Assim utilizou-se do algoritmo Metropolis-Hastings.

A partir da função de verossimilhança do modelo Gompertz com estrutura de erro autorregressivo AR(1), a distribuição condicional para o parâmetro A, pode ser escrita como se segue:

$$P(A | B, k, \tau, Y, \phi) \propto \exp \left\{ -\frac{\tau}{2} \sum_{i=2}^n \left(Y_i - Ae^{-Be^{-kt_i}} - \phi u_{i-1} \right)^2 \right\}$$

Fazendo $Y_i^* = Y_i - \phi u_{i-1}$ e desenvolvendo o quadrado acima, tem-se:

$$\propto \exp \left\{ -\frac{\tau}{2} \sum_{i=2}^n \left(Y_i^{*2} - 2Y_i^* Ae^{-Be^{-kt_i}} + \left(Ae^{-Be^{-kt_i}} \right)^2 \right) \right\}$$

Aplicando o somatório em cada um dos termos da expressão acima e retirando os termos que não dependem de A, que são termos constantes, obtém-se:

$$\propto \exp \left\{ \frac{-\tau}{2} \left(-2A \sum_{i=2}^n Y_i^* e^{-Be^{-kt_i}} + A^2 \sum_{i=2}^n e^{-2Be^{-kt_i}} \right) \right\}$$

Rearranjando os termos, segue que:

$$\propto \exp \left\{ \frac{-\tau \sum_{i=2}^n e^{-2Be^{-kt_i}}}{2} \left(A - \frac{\sum_{i=2}^n Y_i^* e^{-Be^{-kt_i}}}{\sum_{i=2}^n e^{-2Be^{-kt_i}}} \right)^2 \right\}$$

Substituindo $Y_i - \phi u_{i-1}$ por Y_i^* , fixado inicialmente, obtém-se:

$$\propto \exp \left\{ \frac{-\tau \sum_{i=2}^n e^{-2Be^{-kt_i}}}{2} \left(A - \frac{\sum_{i=2}^n (Y_i - \phi u_{i-1}) e^{-Be^{-kt_i}}}{\sum_{i=2}^n e^{-2Be^{-kt_i}}} \right)^2 \right\}$$

Resultando no núcleo de uma distribuição normal com média e variância dada a seguir:

$$N \left(\frac{\sum_{i=2}^n (Y_i - \phi u_{i-1}) e^{-Be^{-kt_i}}}{\sum_{i=2}^n e^{-2Be^{-kt_i}}}; \frac{\sigma^2}{\tau \sum_{i=2}^n e^{-2Be^{-kt_i}}} \right)$$

Analogamente, a distribuição condicional completa para o parâmetro ϕ pode ser escrita da forma proporcional, como se segue:

$$P(\phi | A, B, k, \tau, Y) \propto \exp \left\{ \frac{-\tau}{2} \sum_{i=2}^n \left(Y_i - A e^{-Be^{-kt_i}} - \phi u_{i-1} \right)^2 \right\}$$

Fazendo $Y_i^* = Y_i - A e^{-Be^{-kt_i}}$ e desenvolvendo o quadrado da expressão acima, tem-se:

$$\propto \exp \left\{ \frac{-\tau}{2} \sum_{i=2}^n \left(Y_i^{*2} - 2Y_i^* \phi u_{i-1} + \phi^2 u_{i-1}^2 \right) \right\}$$

Aplicando o somatório em cada termo, segue:

$$\propto \exp \left\{ \frac{-\tau}{2} \left(\sum_{i=2}^n Y_i^{*2} - 2\phi \sum_{i=2}^n Y_i^* u_{i-1} + \phi^2 \sum_{i=2}^n u_{i-1}^2 \right) \right\}$$

Retirando os termos que não dependem do parâmetro ϕ e rearranjando-os, obtém-se:

$$\propto \exp \left\{ \frac{-\tau \sum_{i=2}^n u_{i-1}^2}{2} \left(\phi - \frac{\sum_{i=2}^n Y_i^* u_{i-1}}{\sum_{i=2}^n u_{i-1}^2} \right)^2 \right\}$$

Substituindo $Y_i - Ae^{-Be^{-kt_i}}$ por Y_i^* , obtém-se a distribuição condicional completa dada por:

$$\propto \exp \left\{ \frac{-\tau \sum_{i=2}^n u_{i-1}^2}{2} \left(\phi - \frac{\sum_{i=2}^n \left(Y_i - Ae^{-Be^{-kt_i}} \right) u_{i-1}}{\sum_{i=2}^n u_{i-1}^2} \right)^2 \right\}$$

A expressão acima representa o núcleo de uma distribuição normal com média e variância dadas a seguir:

$$N \left(\frac{\sum_{i=2}^n \left(Y_i - Ae^{-Be^{-kt_i}} \right) u_{i-1}}{\sum_{i=2}^n u_{i-1}^2}; \frac{\sigma^2}{\sum_{i=2}^n u_{i-1}^2} \right)$$

Analogamente, a distribuição condicional completa para o parâmetro τ

pode ser escrita por:

$$P(\tau | A, B, k, \phi, Y) \propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=2}^n \left(Y_i - A e^{-B e^{-k t_i}} - \phi u_{i-1} \right)^2 \right\}$$

Assim, a distribuição condicional completa acima representa o núcleo de uma distribuição gama com os seguintes parâmetros:

$$Gama \left(1 + \frac{n}{2}; (1/2) \sum_{i=2}^n \left(Y_i - A e^{-B e^{-k t_i}} - \phi u_{i-1} \right)^2 \right)$$

No caso dos parâmetros B e k novamente foi utilizado o algoritmo Metropolis-Hastings pelo fato de não se obter uma distribuição conhecida.

4.3 Análise dos dados simulados

4.3.1 Modelo Logístico

Para ilustrar a metodologia na análise de dados simulados e reais, somente a característica diâmetro externo longitudinal (DEL) foi utilizada.

Os valores simulados do diâmetro externo longitudinal (DEL) em função do tempo de crescimento do diâmetro do fruto (em dias), considerando a estrutura de erro independente, podem ser observados pelo Gráfico 2.

Os valores paramétricos utilizados para o estudo de simulação foram assumidos com base nos valores dos parâmetros obtidos por meio do método de mínimos quadrados, ou seja, foram tomados valores próximos a estes, e estão apresentados na Tabela 3.

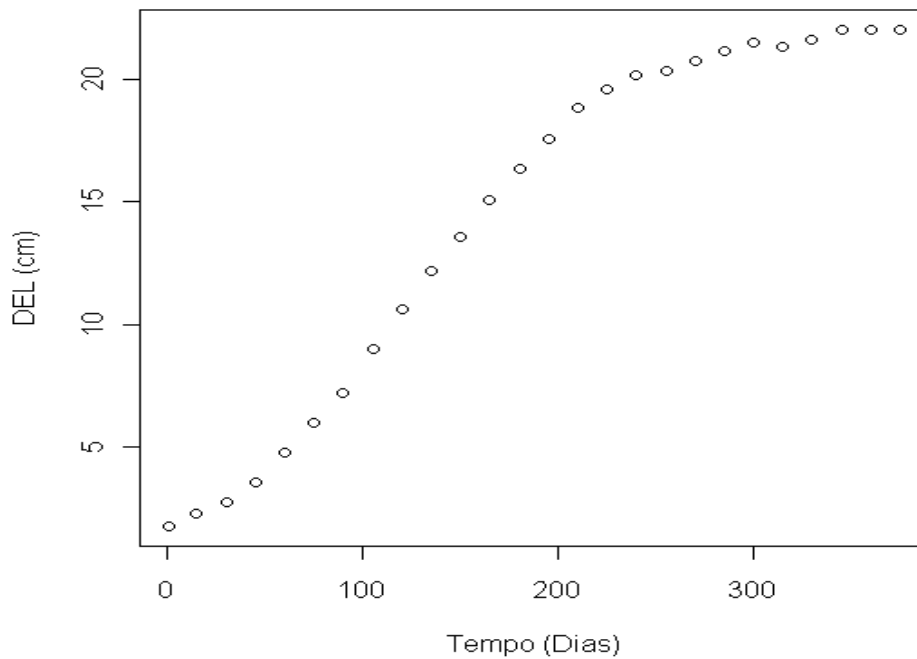


Gráfico 2 Valores simulados do diâmetro externo longitudinal (DEL) em função do tempo, considerando independência dos erros para o modelo Logístico

A convergência das cadeias de cada um dos parâmetros do modelo foi monitorada pela visualização gráfica da cadeia gerada pelo MCMC e pelos critérios de Geweke (1992), Heidelberger e Welch (1993) e Raftery e Lewis (1992).

A cadeia inicial amostrada foi de 4000 iterações. Por meio do critério de Raftery e Lewis (1992), foi sugerido 75000 iterações, *burn-in* de 60 e a distância mínima de uma iteração a outra (*thin*) de 12, para obtenção de uma amostra aproximadamente independente.

Na Tabela 3, visualizam-se os resultados para os critérios de convergência das cadeias. Pelo critério de Geweke (1992), o valor p foi sempre maior que o nível de significância pré-fixado de 5%, não apresentando evidências contra a convergência das cadeias geradas para os parâmetros do modelo Logístico.

Segundo o critério de Heidelberger e Welch (1993), que avalia a hipótese nula de estacionariedade da amostra gerada, não houve evidências para rejeitar a hipótese nula para todos os parâmetros, indicando que não é necessário um número maior de iterações das computadas na análise.

Ainda na Tabela 3, observa-se uma simetria da distribuição para o parâmetro B que também pode ser visualizada pelo gráfico 4 pelo histograma obtido. Neste caso, a média e a moda podem ser consideradas como os valores que maximizam a função de verossimilhança. Para os parâmetros A, k e τ obtém-se distribuições assimétricas (Gráficos 3, 5 e 6), adotando a moda como valor estimado.

Tabela 3 Valores paramétricos (VP), Média, Moda, Erro Monte Carlo (EMC), Intervalo de Máxima Densidade (HPD) e o Critério de Geweke (valor p) das cadeias amostradas

Parâmetros	VP	Média	Moda	EMC	HPD		Geweke (Valor p)
					LI	LS	
A	22,00	21,89	21,76	0,0500	20,33	23,68	0,1176
B	12,00	12,00	12,00	0,0001	11,98	12,02	0,3033
k	0,02	0,02	0,018	0,0001	0,017	0,024	0,1254
τ	8,00	14,04	13,23	0,0464	7,39	21,55	0,4932

Pelo valor obtido do parâmetro τ (Tabela 3), tem-se o valor da medida de dispersão σ^2 dos dados simulados de diâmetro externo longitudinal, que é de 0,0755 para o modelo Logístico com erros independentes.

Os valores encontrados para o Erro Monte Carlo (EMC) foram de baixa magnitude (Tabela 3), garantindo que as amostras geradas pelas distribuições condicionais completas podem ser consideradas como amostras das distribuições marginais.

Nos Gráficos de 3 a 6 tem-se a visualização gráfica da cadeia, para cada parâmetro, confirmando a convergência apresentada na tabela 3, após o descarte das iterações iniciais (*burn-in*) e o intervalo entre os pontos amostrados (*thin*) indicados pelo critério de Raftery e Lewis (1992).

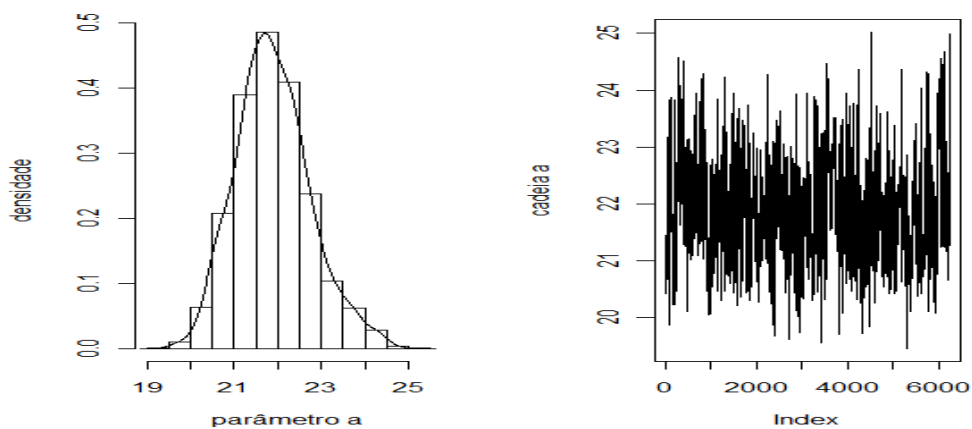


Gráfico 3 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro A

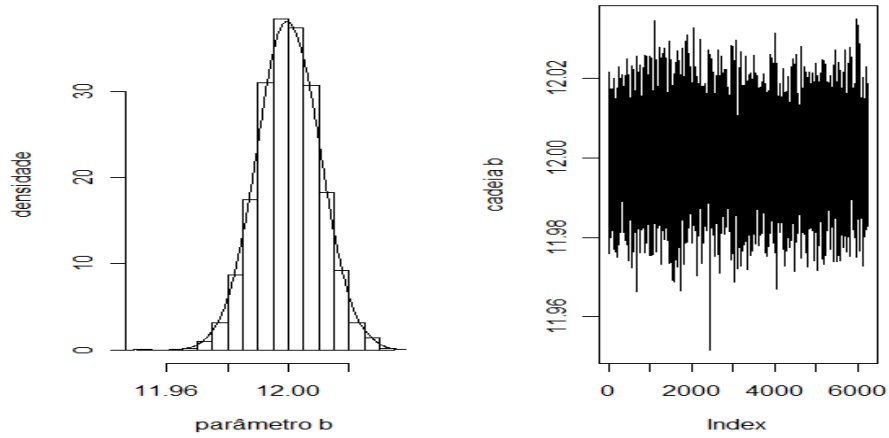


Gráfico 4 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro B

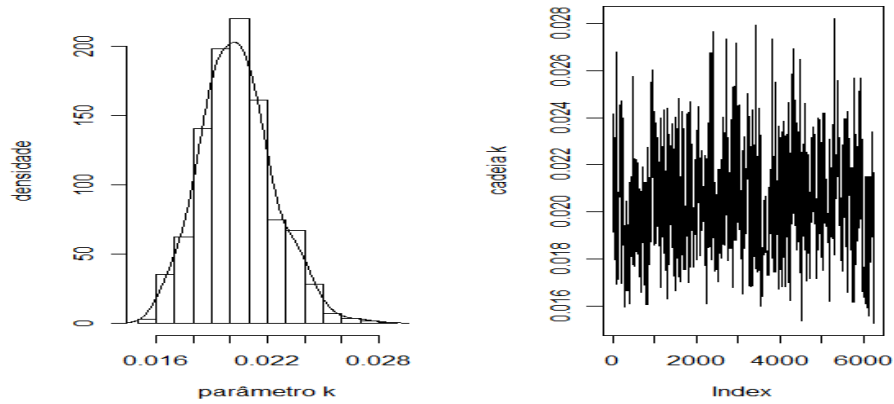


Gráfico 5 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro k

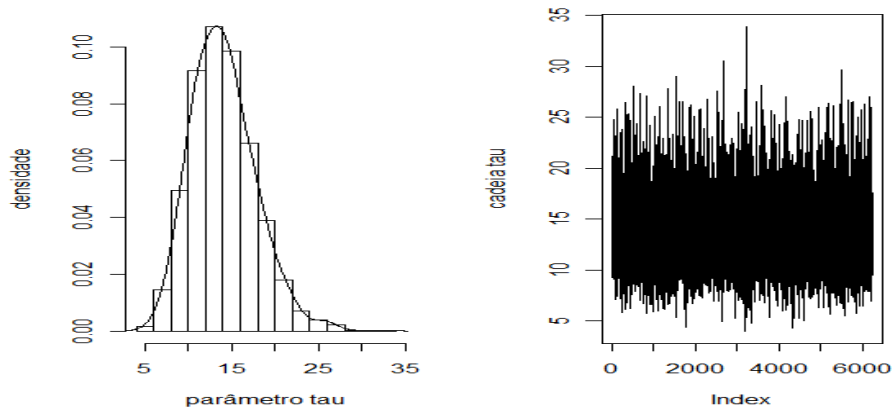


Gráfico 6 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro precisão

Os valores simulados do diâmetro externo longitudinal (DEL) em função do tempo de crescimento do diâmetro do fruto (em dias), considerando a estrutura de erro dependente AR(1), podem ser observados pelo Gráfico 7.

Após simular uma cadeia de tamanho 10000, o critério de Raftery e Lewis (1992), sugeriu 18000 iterações para o tamanho da cadeia, *bunr-in* de 20 e um passo (*thin*) de 5.

Para o critério de Geweke (1992), o valor p obtido também foi sempre maior que o nível de significância adotado de 5%, não apresentando evidências contra a convergência.

Segundo o critério de Heidelberger e Welch (1993), a hipótese nula de estacionariedade da amostra gerada não foi rejeitada para todos os parâmetros do modelo Logístico com AR(1).

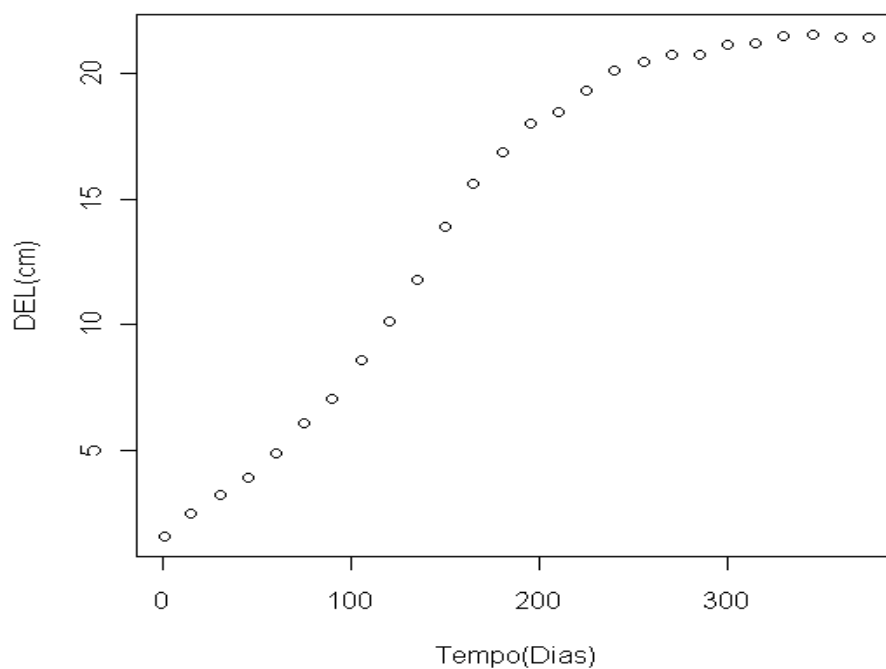


Gráfico 7 Valores simulados do diâmetro externo longitudinal (DEL) em função do tempo, considerando os erros autorregressivos AR(1) para o modelo Logístico

Por meio da tabela 4, observa-se uma simetria da distribuição para os parâmetros A e k que também pode ser visualizada pelos gráficos 8 e 10, respectivamente, onde são apresentados os histogramas. Pode-se observar que tanto a média como a moda podem ser consideradas como os valores que maximizam a função de verossimilhança. Para os parâmetros B, τ e ϕ obtém-se distribuições assimétricas (Gráficos 9, 11 e 12 respectivamente), considerando a moda como o melhor valor estimado.

Tabela 4 Valores paramétricos (VP), Média, Moda, Erro Monte Carlo (EMC), Intervalo de Máxima Densidade (HPD) e o Critério de Geweke (valor p) das cadeias amostradas

Parâmetros	VP	Média	Moda	EMC	HPD		Geweke (Valor p)
					LI	LS	
A	22,00	21,50	21,51	$1,38 \cdot 10^{-2}$	20,30	22,47	0,6805
B	12,00	13,90	14,18	$1,03 \cdot 10^{-1}$	10,07	17,05	0,5787
k	0,02	0,019	0,019	$4,90 \cdot 10^{-7}$	0,019	0,020	0,5242
τ	0,80	0,99	0,97	$1,55 \cdot 10^2$	0,10	1,66	0,1575
ϕ	0,50	0,43	0,40	$9,87 \cdot 10^{-3}$	-0,004	0,97	0,4713

Os resultados apresentados para o Erro Monte Carlo (EMC) foram de baixa magnitude (Tabela 4), garantindo a convergência das cadeias geradas para cada parâmetro.

O valor obtido para o parâmetro τ (Tabela 4) encontra-se o valor da variância σ^2 dos dados simulados que é de 1,0349 para o modelo Logístico com erros autorregressivos AR(1).

Os resultados da tabela 4 mostram a convergência das cadeias para cada parâmetro e podem ser confirmados por meio da visualização gráfica do traço da cadeia, nos gráficos de 8 a 12, após o descarte das iterações iniciais (*burn-in*) e o intervalo entre os pontos amostrados (*thin*).

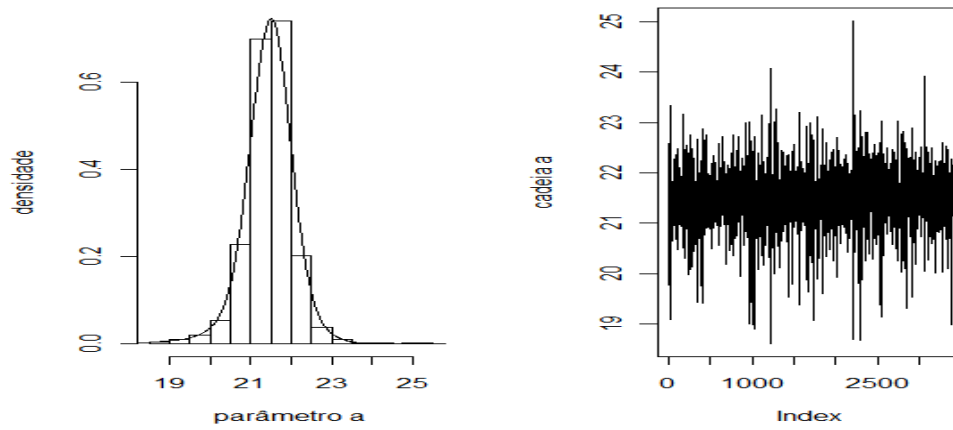


Gráfico 8 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro A

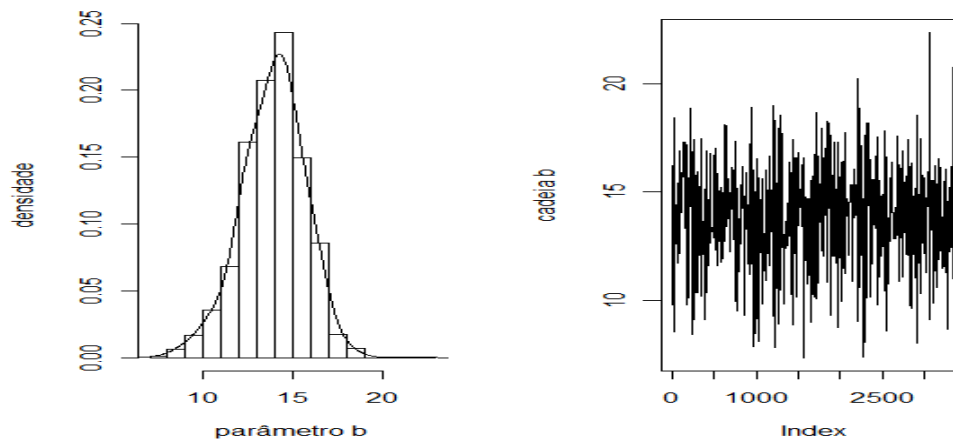


Gráfico 9 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro b

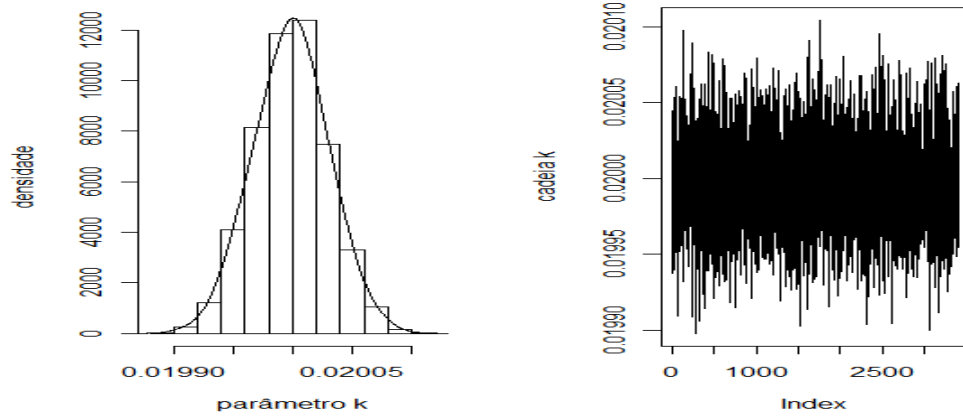


Gráfico 10 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro k

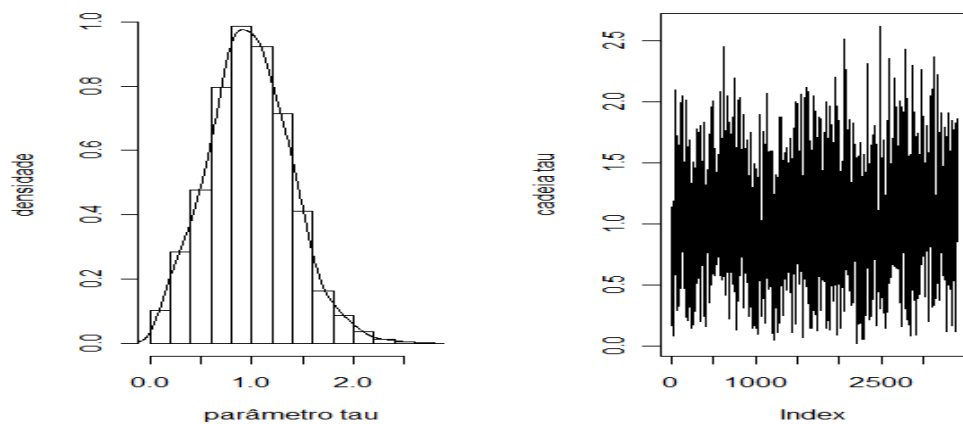


Gráfico 11 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro precisão

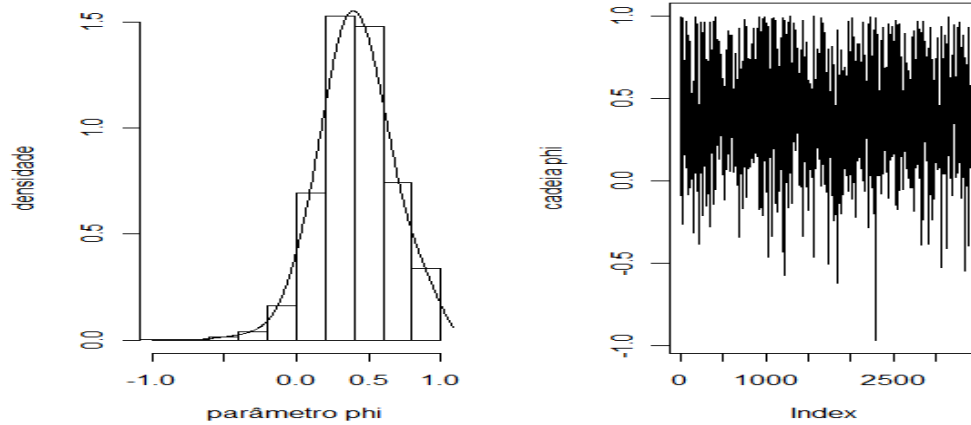


Gráfico 12 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro ϕ

4.3.2 Modelo Gompertz

Os valores simulados do diâmetro externo longitudinal (DEL) em função do tempo de crescimento do diâmetro do fruto (em dias), considerando a estrutura de erro independente para o modelo de Gompertz, podem ser observados pelo Gráfico 13.

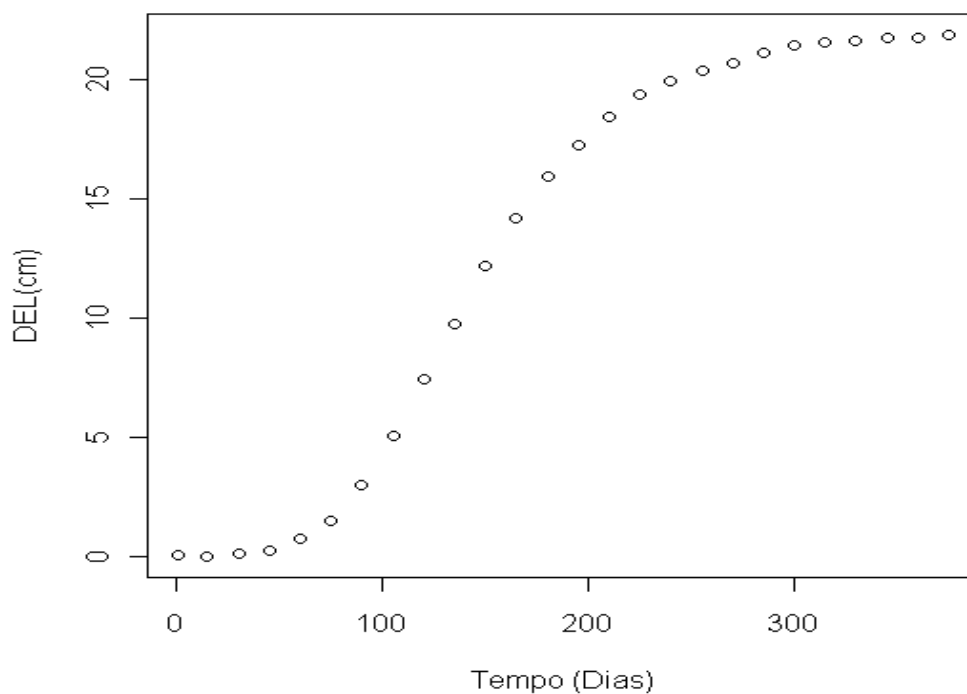


Gráfico 13 Valores simulados do diâmetro externo longitudinal (DEL) em função do tempo, considerando independência dos erros para o modelo Gompertz

Da mesma forma, a convergência das cadeias de cada um dos parâmetros do modelo foi monitorada pela visualização gráfica da cadeia gerada pelo MCMC e pelos critérios de Geweke (1992), Heidelberger e Welch (1993) e Raftery e Lewis (1992).

Pelo critério de Raftery e Lewis (1992), foi sugerido 46000 iterações para o tamanho da cadeia, *burn-in* de 32 e um passo (*thin*) de 8 após simular uma amostra inicial de tamanho 4000.

Na Tabela 5, visualizam-se os resultados para os critérios de convergência das cadeias. Pelo critério de Geweke (1992), o valor p foi sempre

maior que o nível de significância pré-fixado de 5%, não apresentando evidências contra a convergência para os parâmetros do modelo de Gompertz.

Segundo o critério de Heidelberger e Welch (1993), a aceitação da hipótese nula de estacionariedade da amostra gerada indica que não é necessário um número maior de iterações além das computadas na análise.

Pelo valor do parâmetro τ (Tabela 5), encontra-se o valor da variância σ^2 dos dados simulados que é de 0,077 para o modelo Gompertz considerando a estrutura de erros independentes.

Por meio da Tabela 5, observam-se uma simetria da distribuição para os parâmetros A, B e k podendo ser confirmados pelos gráficos 14, 15, e 16 respectivamente, em que contém os histogramas. A média como também a moda podem ser consideradas como os valores que maximizam a função de verossimilhança. Somente para o parâmetro τ obteve-se uma distribuição assimétrica (Gráfico 17), adotando a moda como valor estimado.

Os resultados obtidos para o Erro Monte Carlo (EMC) foram de baixa magnitude (Tabela 5) para todos os parâmetros, garantindo a convergência das cadeias.

Tabela 5 Valores paramétricos (VP), Média, Moda, Erro Monte Carlo (EMC), Intervalo de Máxima Densidade (HPD) e o Critério de Geweke (valor p) das cadeias amostradas

Parâmetros	VP	Média	Moda	EMC	HPD		Geweke (Valor p)
					LI	LS	
A	22,00	21,97	21,96	$6,48 \cdot 10^{-3}$	21,56	22,38	0,3657
B	3,00	2,99	2,98	$7,31 \cdot 10^{-3}$	2,59	3,39	0,5862
k	0,02	0,020	0,020	$3,35 \cdot 10^{-5}$	0,019	0,022	0,3726
τ	8,00	14,05	12,93	$5,51 \cdot 10^2$	7,45	21,55	0,1282

Nos Gráficos de 14 a 17 são apresentados os histogramas e as cadeias geradas para cada parâmetro e confirmam a convergência apresentada na Tabela 5 após a retirada das iterações iniciais (*burn-in*) e o passo (*thin*).

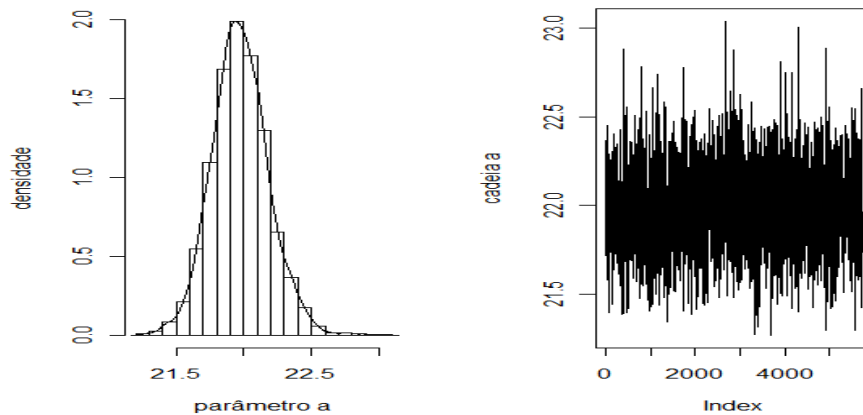


Gráfico 14 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro A

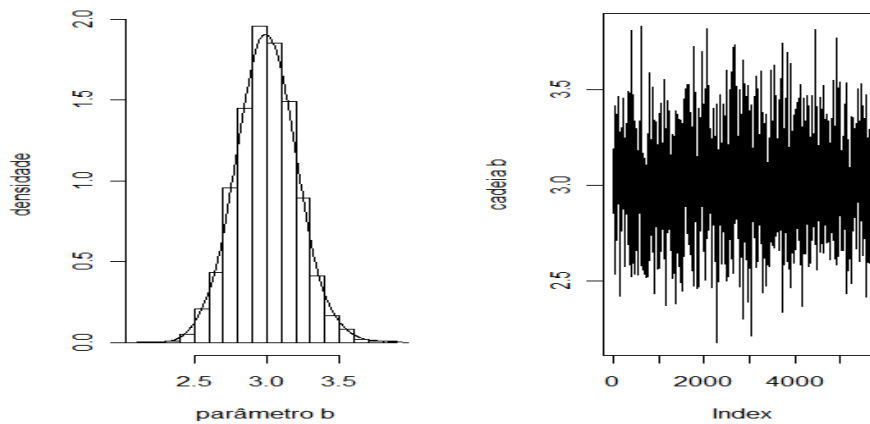


Gráfico 15 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro B

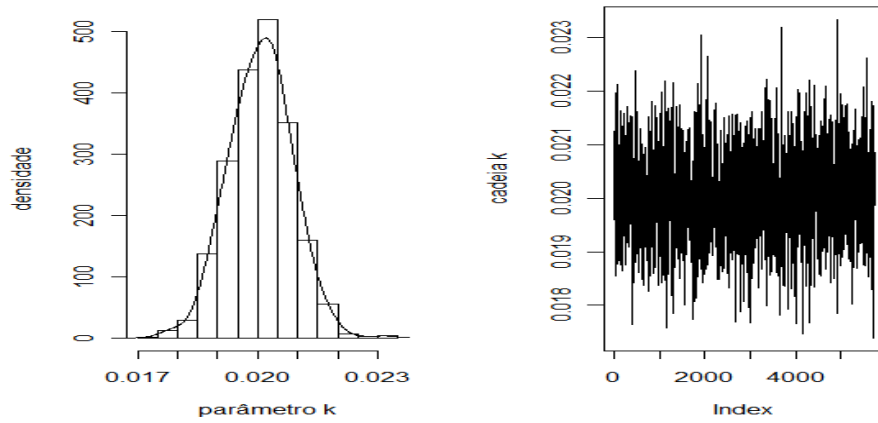


Gráfico 16 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro k

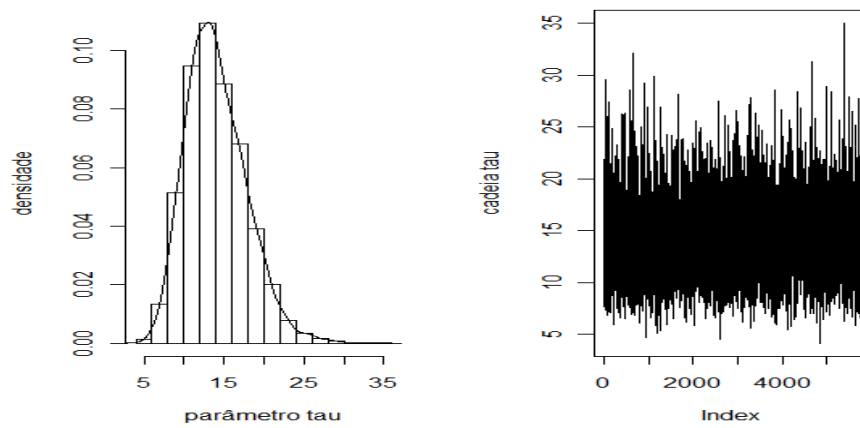


Gráfico 17 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro precisão

Os valores simulados do diâmetro externo longitudinal (DEL) em função do tempo de crescimento do diâmetro (em dias), considerando a estrutura de erro dependente AR(1), podem ser observados pelo Gráfico 18.

Inicialmente, foi simulada uma amostra de tamanho 10000. Pelo critério de Raftery e Lewis (1992), foi sugerido 80000 iterações para o tamanho da cadeia, *burn-in* de 60 e um passo (*thin*) de 15 para obter uma amostra aproximadamente independente.

Para o critério de Geweke (1992), o valor p obtido também foi sempre maior que o nível de significância adotado de 5%, concluindo não apresentar evidências contra a convergência das cadeias.

Segundo o critério de Heidelberger e Welch (1993), não foi necessário um número maior de iterações além das computadas, pois a hipótese nula de estacionariedade da amostra gerada não foi rejeitada para todos os parâmetros.

A medida de dispersão σ^2 dos dados simulados é encontrada pelo valor obtido do parâmetro τ (Tabela 6), que é de 1, 5850 para o modelo Gompertz considerando a estrutura de erro autorregressivo AR(1).

Por meio da Tabela 6, observam-se uma simetria da distribuição para os parâmetros A, B e k assim como pode-se visualizar pelos Gráficos 19, 20 e 21 respectivamente, os histogramas obtidos. Neste caso, tanto a média como a moda podem ser consideradas como os valores que maximizam a função de verossimilhança. Para os parâmetros τ e ϕ obteve-se uma distribuição assimétrica (Gráficos 22 e 23), adotando a moda como valor estimado.

Na Tabela 6, os valores obtidos para o Erro Monte Carlo (EMC) foram de baixa magnitude, garantindo a convergência das cadeias. Nos Gráficos de 19 a 23 a seguir, pode-se confirmar os resultados apresentados na Tabela 6, por meio da visualização gráfica da cadeia, após o descarte das iterações iniciais (*burn-in*) e o intervalo entre os pontos amostrados (*thin*).

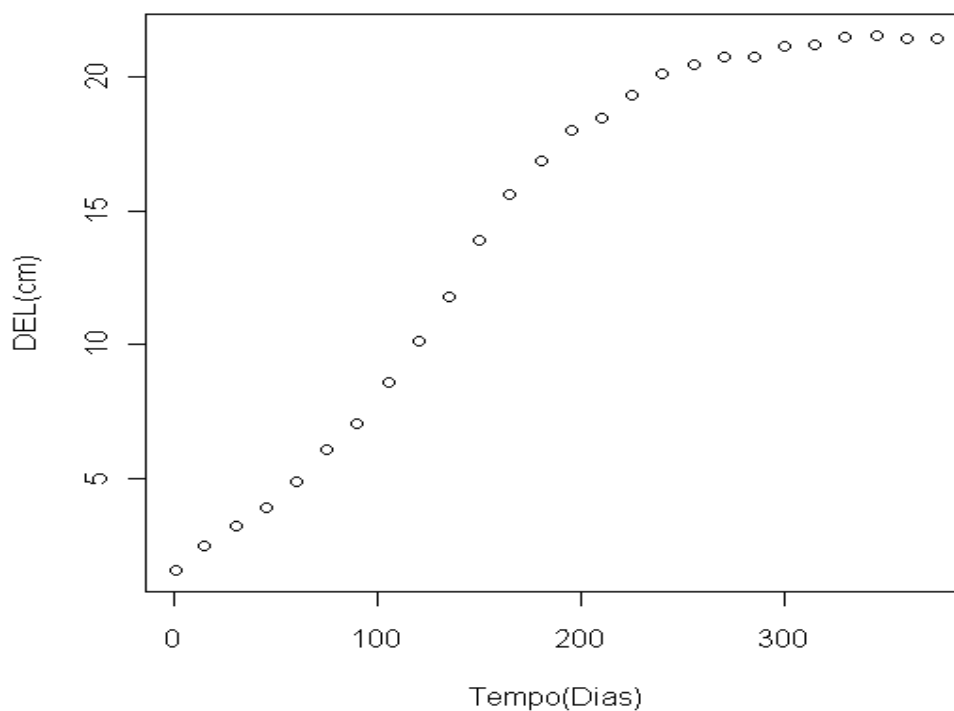


Gráfico 18 Valores simulados do diâmetro externo longitudinal (DEL) em função do tempo, considerando a estrutura de erros AR(1) para o modelo Gompertz

Tabela 6 Média, Moda, Erro Monte Carlo (EMC), Intervalo de Máxima Densidade (HPD) e o Critério de Geweke (valor p) das cadeias amostradas

Parâmetros	VP	Média	Moda	EMC	HPD		Geweke (Valor p)
					LI	LS	
A	22,00	21,12	21,11	$1,47 \cdot 10^{-2}$	20,22	21,97	0,7775
B	1,50	1,47	1,47	$8,90 \cdot 10^{-3}$	1,13	1,77	0,4856
k	0,02	0,020	0,019	$4,93 \cdot 10^{-7}$	0,019	0,020	0,6198
τ	0,50	0,66	0,63	$6,74 \cdot 10^{-3}$	0,25	1,11	0,9220
ϕ	0,50	0,57	0,55	$5,80 \cdot 10^{-3}$	0,16	0,99	0,3297

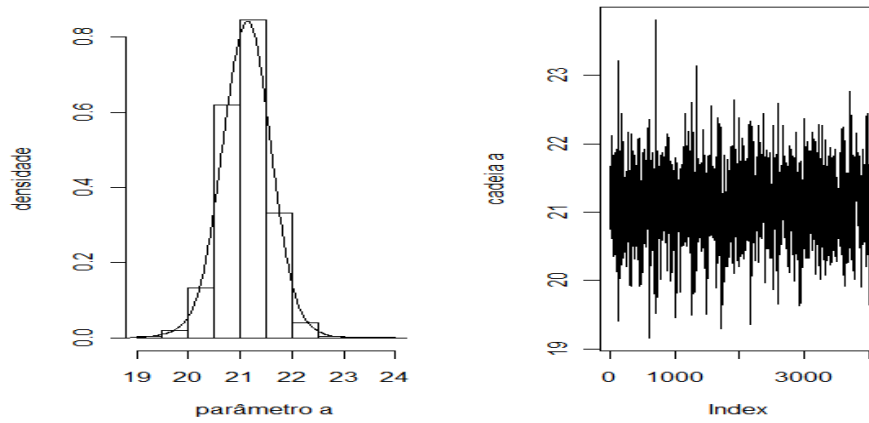


Gráfico 19 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro A

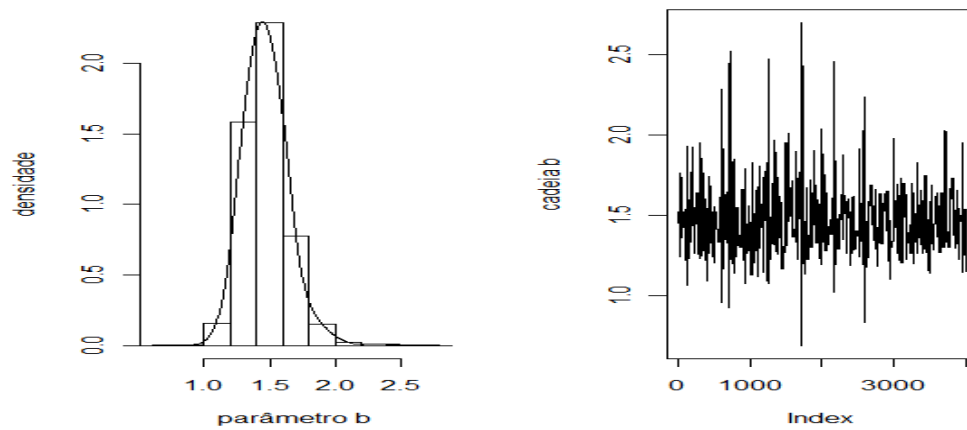


Gráfico 20 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro B

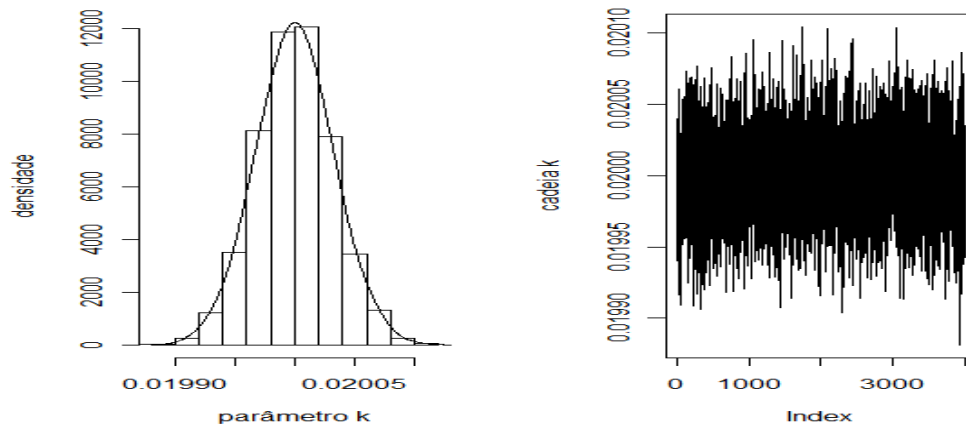


Gráfico 21 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro k

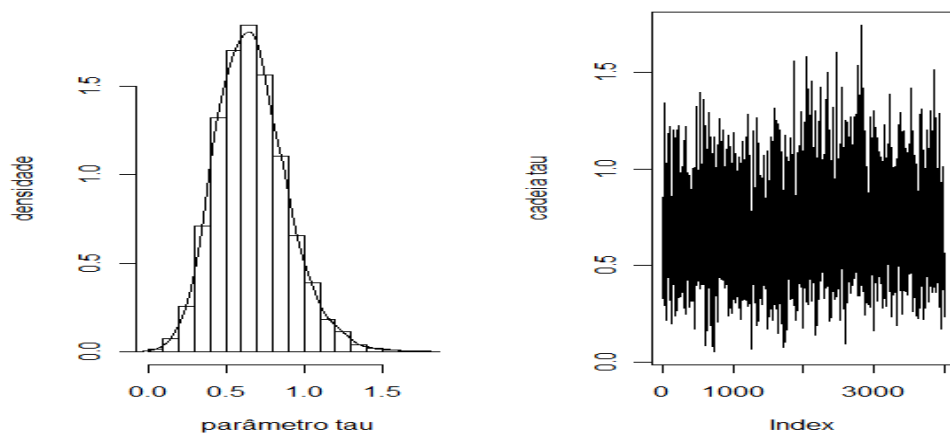


Gráfico 22 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro τ

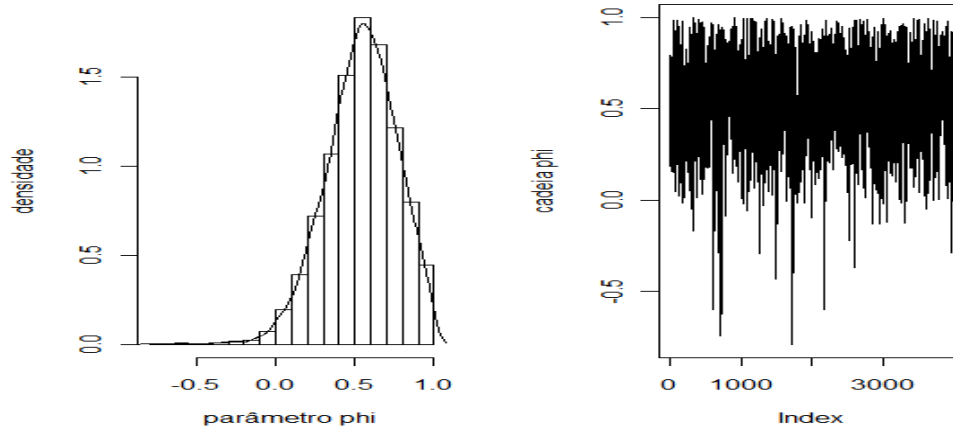


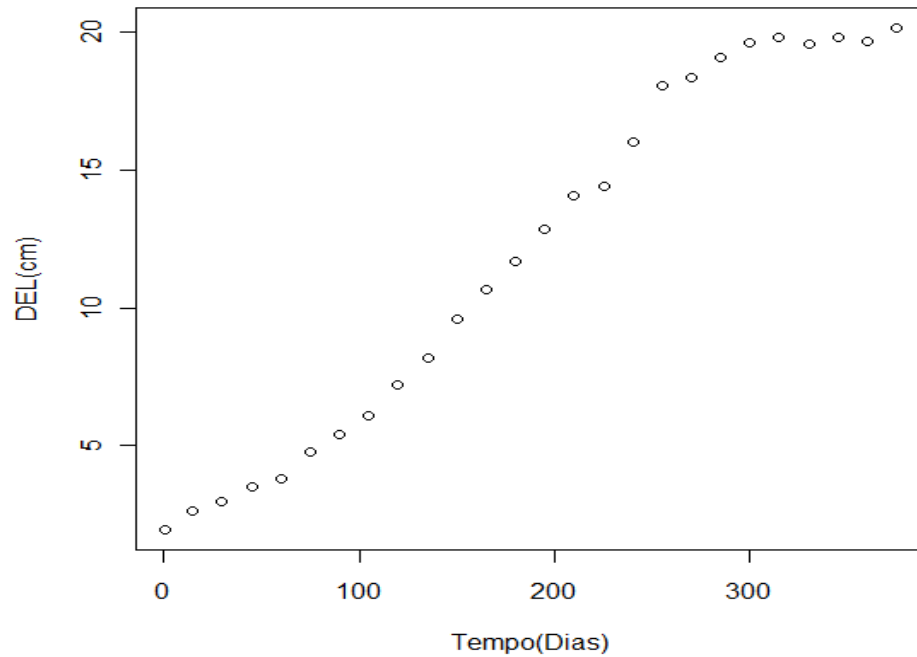
Gráfico 23 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro phi

4.4 Análise dos dados reais

4.4.1 Modelo Logístico

Utilizando os dados reais e a partir do resultado de que os resíduos são correlacionados na forma de um processo autorregressivo de primeira ordem AR(1), foi obtida a aproximação das distribuições marginais para cada um dos parâmetros em ambos modelos.

No Gráfico 24 são apresentados os valores reais do diâmetro externo longitudinal (DEL) em função do tempo de crescimento do diâmetro do fruto (em dias).



Pelo critério de Raftery e Lewis (1992), foi sugerido 50000 iterações para o tamanho da cadeia, um *burn-in* de 90 e a distância mínima de uma iteração a outra (*thin*) de 10, após simular uma amostra piloto de 10000.

Para o critério de Geweke (1992), o valor p obtido também foi sempre maior que o nível de significância adotado de 5%, garantindo a convergência das cadeias. Segundo o critério de Heidelberger e Welch (1993), a estacionariedade da amostra gerada não foi rejeitada para todos os parâmetros.

Tabela 7 Média, Moda, Erro Monte Carlo (EMC), Intervalo de Máxima Densidade (HPD) e o Critério de Geweke (valor p) das cadeias amostradas

Parâmetros	Média	Moda	EMC	HPD		Geweke (Valor p)
				LI	LS	
A	20,0398	20,0622	$1,72 \cdot 10^{-2}$	19,3384	20,7137	0,9736
B	21,2794	21,2472	$1,19 \cdot 10^{-1}$	17,6069	23,9234	0,9916
k	0,0199	0,0199	$5,37 \cdot 10^{-7}$	0,0199	0,0200	0,6624
τ	1,3994	1,3178	$8,19 \cdot 10^{-3}$	0,6657	2,2047	0,8258
ϕ	0,6519	0,8521	$3,98 \cdot 10^{-3}$	0,1565	0,9999	0,9663

Ainda na Tabela 7, pode-se observar uma simetria da distribuição para os parâmetros A, B e k, sendo também visualizada pelos Gráficos 25, 26 e 27 respectivamente, por meio dos histogramas obtidos. Assim, tanto a média como a moda podem ser consideradas como os valores que maximizam a função de verossimilhança. Para os parâmetros τ e ϕ obteve-se uma distribuição assimétrica (Gráficos 28 e 29, respectivamente) adotando-se a moda como o melhor valor estimado.

Pelo resultado do parâmetro τ na Tabela 7, encontra-se o valor da variância σ^2 que é de 0,7588 para os dados reais no modelo Logístico considerando a estrutura de erro autorregressivo (AR1).

Os valores encontrados para o Erro Monte Carlo (EMC) foram de baixa magnitude (Tabela 7), garantindo que houve a convergência das cadeias geradas e as amostras geradas se aproximam de amostras das distribuições marginais.

A visualização gráfica do traço da cadeia é apresentada nos gráficos de 25 a 29 bem como os histogramas, após o *burn-in* e o *thin* indicados pelo critério de Raftery e Lewis (1992), confirmando a convergência de cada cadeia.

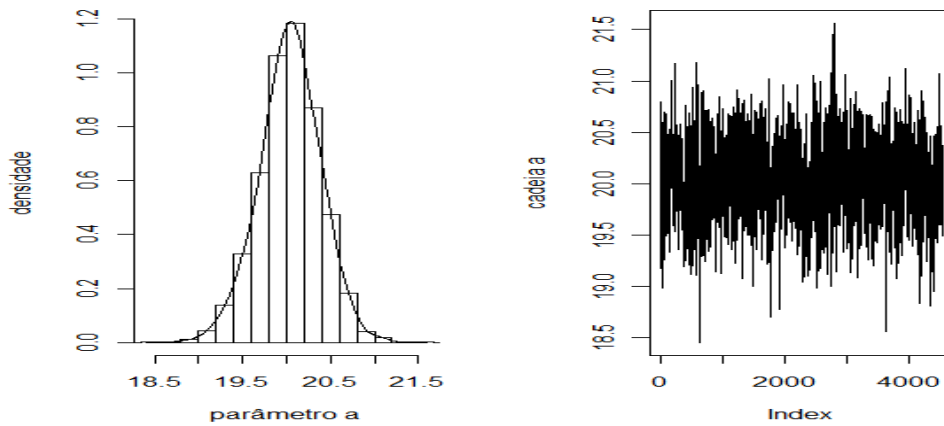


Gráfico 25 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro A

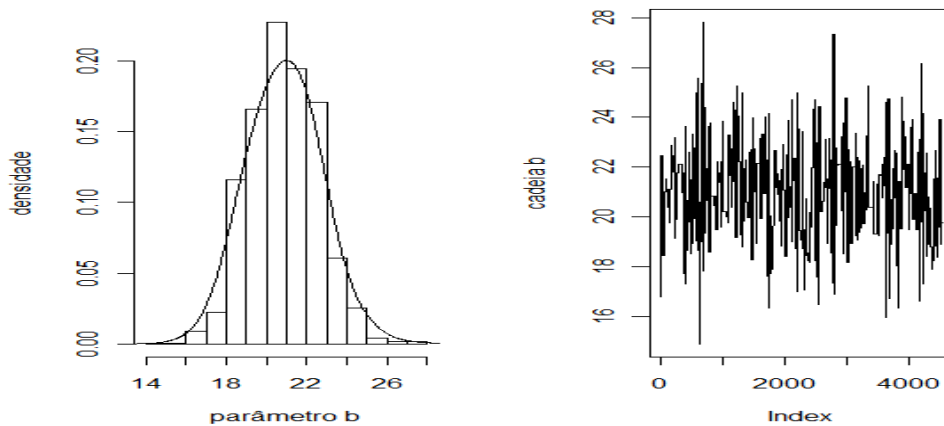


Gráfico 26 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro B

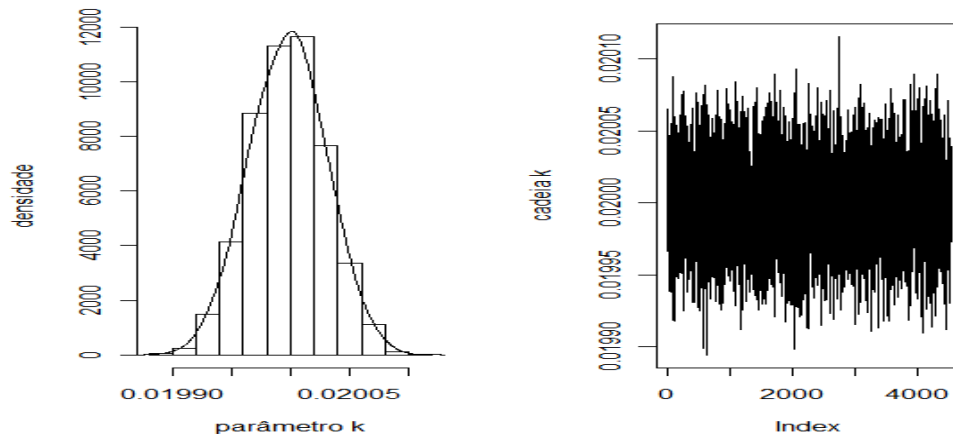


Gráfico 27 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro k

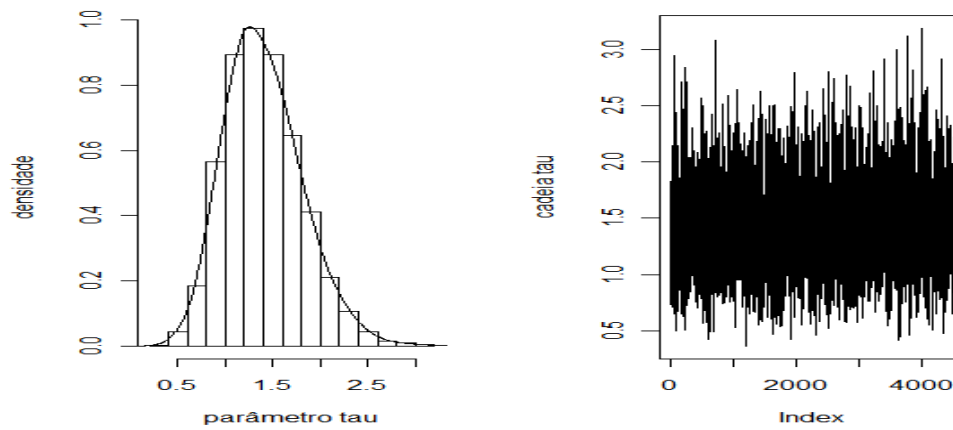


Gráfico 28 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro precisão

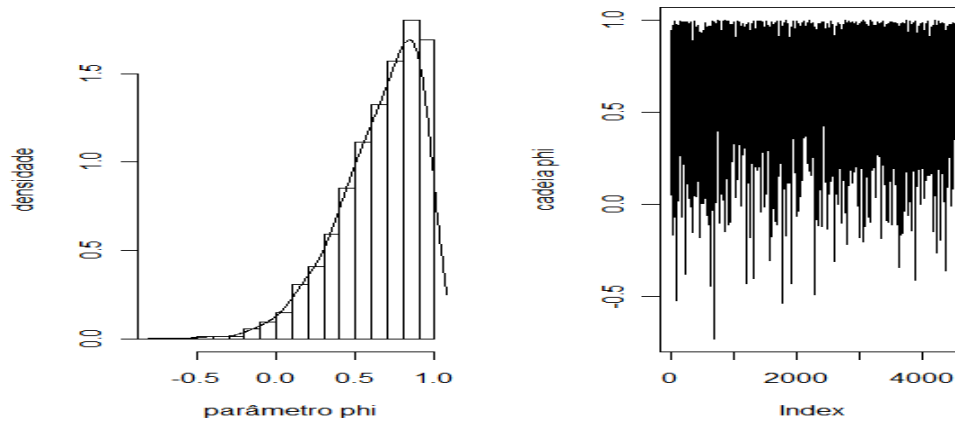


Gráfico 29 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro phi

4.4.2 Modelo Gompertz

Inicialmente, simulando uma amostra piloto de tamanho 10000, o critério de Raftery e Lewis (1992) sugeriu 63000 iterações para o tamanho da cadeia, um *burn-in* de 55 e a distância mínima de uma iteração a outra (*thin*) de 11 para obter a convergência. Segundo o critério de Heidelberger e Welch (1993), a hipótese nula de estacionariedade da amostra gerada não foi rejeitada para todos os parâmetros. Os valores para o critério de Geweke (1992) foram todos acima do nível de significância de 5% (Tabela 8), não apresentando indícios contra a convergência.

Por meio da Tabela 8, observam-se uma simetria da distribuição para os parâmetros A e k podendo ser confirmada pela visualização dos gráficos 30 e 32, onde mostra o histograma deste parâmetro. Assim, tanto a média como a moda podem ser consideradas como os valores que maximizam a função de verossimilhança.

Tabela 8 Média, Moda, Erro Monte Carlo (EMC), Intervalo de Máxima Densidade (HPD) e o Critério de Geweke (valor p) das cadeias amostradas

Parâmetros	Média	Moda	EMC	HPD		Geweke (Valor p)
				LI	LS	
A	17,2974	17,2947	$8,04 \cdot 10^{-3}$	16,7764	17,8109	0,4503
B	3,8679	3,7814	$2,11 \cdot 10^{-2}$	3,2815	4,4989	0,8443
k	0,0199	0,0199	$4,06 \cdot 10^{-7}$	0,0199	0,0200	0,7320
τ	1,1931	1,1331	$7,19 \cdot 10^{-3}$	0,5211	1,8681	0,7084
ϕ	0,9410	0,9786	$1,69 \cdot 10^{-3}$	0,8380	0,9999	0,7777

Para os parâmetros B, τ e ϕ obteve-se uma distribuição assimétrica observando-se a tabela 8 como também os gráficos 31, 33 e 34, respectivamente. Desta forma, adota-se a moda como o melhor valor estimado.

Pelo valor obtido para o parâmetro τ (Tabela 8) encontra-se o valor da variância σ^2 aos dados reais que é de 0,8825 para o modelo Gompertz considerando a estrutura de erro autorregressivo (AR1). Os valores encontrados para o Erro Monte Carlo (EMC) foram de baixa magnitude (Tabela 8) para os parâmetros, garantindo a convergência das cadeias.

Os resultados da tabela 8 podem ser confirmados por meio da visualização gráfica do traço da cadeia, nos Gráficos de 30 a 34, onde se obteve a convergência após o descarte das iterações iniciais (*burn-in*) e o intervalo entre os pontos amostrados (*thin*).

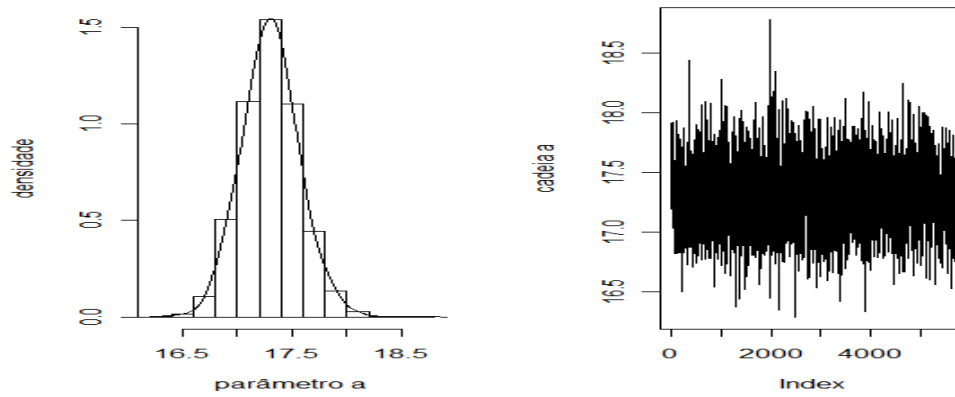


Gráfico 30 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro A

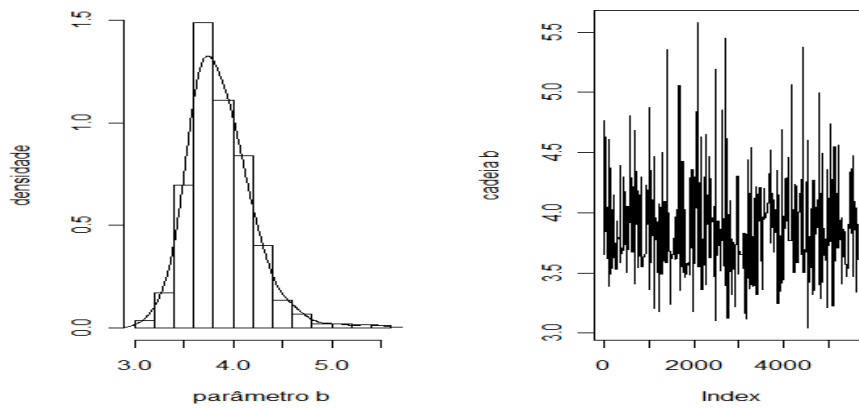


Gráfico 31 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro B

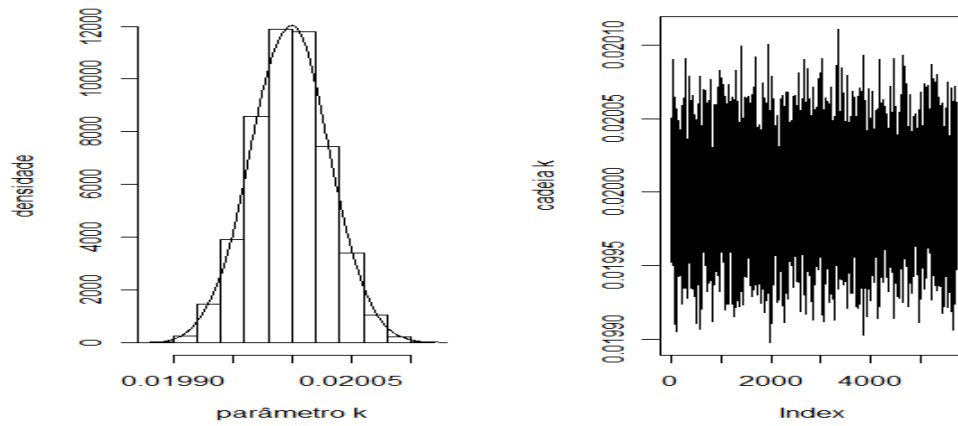


Gráfico 32 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro k

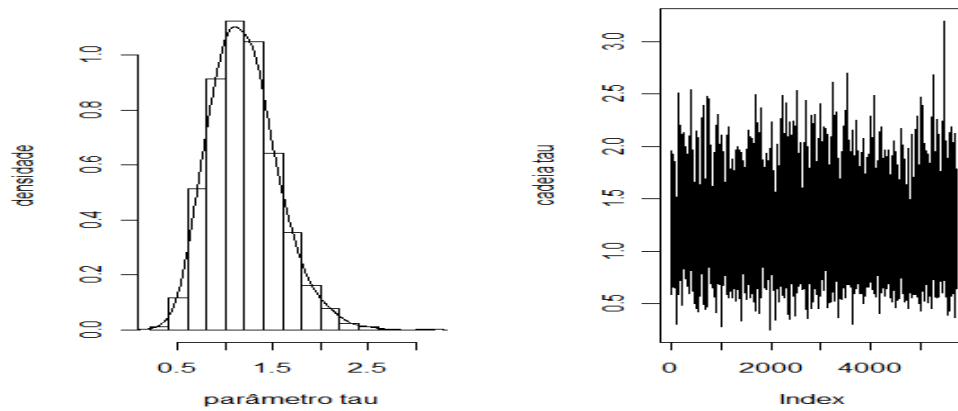


Gráfico 33 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro precisão

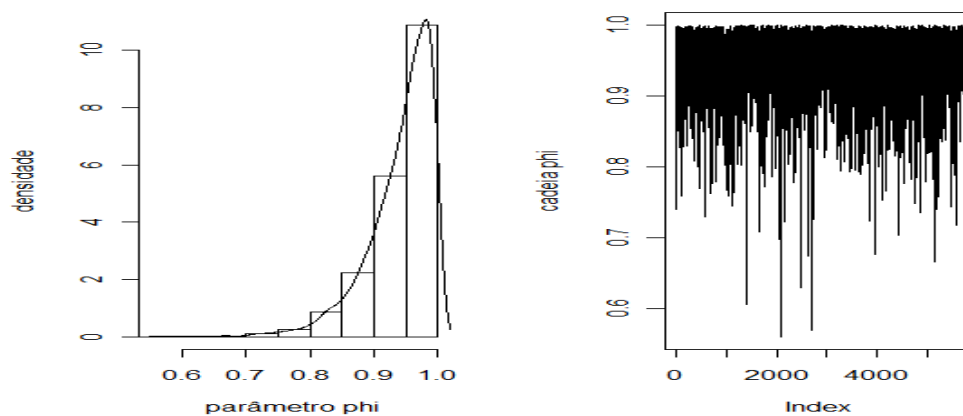


Gráfico 34 Histograma da distribuição marginal e traço da cadeia gerada pelo método MCMC, para o parâmetro phi

4.4.3 Comparação dos modelos

O resultado da comparação dos modelos Logístico e Gompertz, considerando a estrutura de erro dependente AR(1), feita por meio do AIC encontram-se na tabela 9. Os resultados mostram que o modelo Logístico é o mais adequado na descrição dos dados experimentais, pois apresentou menor valor para o AIC.

Tabela 9 AIC para a comparação dos modelos Logístico e Gompertz com AR(1) aos dados reais

Modelos	<i>AIC</i>
Logístico	0,2039
Gompertz	0,3550

5 CONCLUSÃO

Nas condições do presente estudo, e de acordo com os resultados obtidos, pode-se concluir que:

- a) Por meio do método de mínimos quadrados, o modelo Logístico foi o mais adequado para ajustar os dados experimentais de diâmetro externo longitudinal (DEL) e transversal (DET), considerando a estrutura autorregressiva de primeira ordem para os resíduos e resultou em estimativas dos parâmetros coerentes com os apresentados na literatura.
- b) A metodologia utilizada no procedimento de ajuste por meio do estudo de simulação dos dados de diâmetro externo longitudinal (DEL) mostrou-se eficiente, considerando as estruturas de erros independentes e autorregressiva dos resíduos, apresentando estimativas bem próximas ao valor verdadeiro.
- c) Para o ajuste dos modelos aos dados experimentais de diâmetro externo longitudinal (DEL), por meio do método iterativo de Monte Carlo via cadeias de Markov, considerando a estrutura autorregressiva de primeira ordem para os resíduos, constata-se que o modelo Logístico foi o mais adequado na descrição dos dados e resultou em estimativas dos parâmetros bastante coerentes com os relatados na literatura.

REFERÊNCIAS

ALVAREZ, I. A.; CASTRO, P. R. de C. Crescimento da parte aérea de cana crua e queimada. **Scientia Agrícola**, Piracicaba, v. 56, n. 4, p. 1069-1079, 1999.

BENASSI, A. C. **Caracterizações biométricas, química e sensorial de frutos de coqueiro variedade anã verde**. 2006. 49 f. Tese (Doutorado em Produção Vegetal) - Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 2006.

BENASSI, A. C. et al. Caracterização biométrica de frutos de coqueiro, *Vocos nucifera* L. variedade anã-verde, em diferentes estádios de desenvolvimento. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 29, n. 2, p. 302-307, mar./abr. 2007.

BERGAMASCO, A. F. et al. Ajuste de modelos não-lineares a dados de crescimento de fêmeas da raça holandesa. **Ciência e Agrotecnologia**, Lavras, v. 25, n. 2, p. 235-241, mar./abr. 2001.

BIALE, J. B.; YOUNG, R. E. Growth, maturation and senescence in fruits. **Science**, Washington, v. 146, n. 3646, p. 880-888, 1964.

BROWN, J. E.; BROWN, C. J.; BUTTS, W. T. Relationships among weights, gains and earliness of maturing in Hereford and Angus females. **Journal of Animal Science**, Champaign, v. 35, n. 3, p. 507-517, 1972.

CAMBOIM NETO, L. F. **Coqueiro anão verde**: influência de diferentes lâminas de irrigação e de porcentagens de área molhada no desenvolvimento, na produção e nos parâmetros físicoquímicos do fruto. 2002. 113 f. Tese (Doutorado em Engenharia Agrícola) - Universidade Federal de Viçosa, Viçosa, MG, 2002.

CARDOSO, G. D. et al. Uso da análise de crescimento não destrutiva como ferramenta para avaliação de cultivares. **Revista de Biologia e Ciências da Terra**, João Pessoa, v. 6, n. 2, p. 79-84, 2006.

CARLESSO, V. O. et al. Avaliação de modelos de secagem em camada fina de sementes de maracujá amarelo. **Revista Brasileira de Sementes**, Brasília, v. 29, n. 2, p. 28-37, mar./abr. 2007.

CAVALINI, F. C. et al. Maturity indexes for 'Kumagai' and 'Paluma' guavas. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 28, n. 2, p. 176-179, ago. 2006.

COOMBE, B. G. The development of fleshy fruits. **Annual Review of Plant Physiology**, Palo Alto, v. 27, p. 507-528, 1976.

CORRÊA, L. S. et al. Comportamento do coqueiro anão verde (*Cocos nucifera* L.) em Selvíria, MS. In: CONGRESSO BRASILEIRO DE FRUTICULTURA, 17., 2002, Belém. **Anais...** Belém: UFAM, 2002. 1 CD-ROM.

CORRÊA, P. C. et al. Modelagem matemática para a descrição do processo de secagem do feijão (*Phaseolus vulgaris* L.) em camadas delgadas. **Revista Engenharia Agrícola**, Jaboticabal, v. 27, n. 2, p. 501-510, 2007.

CORSATO, C. E. **Fenologia e carboidratos de reserva do caqui (Diospyros kaki L.) 'Rama Forte' em clima tropical no Brasil**. 2004. 54 p. Tese (Doutorado em Fitotecnia) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2004.

DENISE, R. S. K.; BRINKS, J. S. Genetic and environmental aspects of the growth curve parameters in beef cows. **Journal of Animal Science**, Champaign, v. 61, n. 6, p. 1431-1440, July 1985.

DRAPER, N. R.; SMITH, N. **Applied regression analysis**. New York: J. Wiley, 1998. 593 p.

EHLERS, R. S. **Introdução à inferência bayesiana**. São Paulo: USP, 2007. 68 p.

GALLANT, A. R. **Nonlinear statistical models**. New York: J. Wiley, 1987. 610 p.

GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches for calculating marginal densities. **Journal of the American Statistical Association**, London, v. 85, n. 410, p. 398-409, 1990.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-472, Nov. 1992.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Los Alamitos, v. 6, n. 6, p. 721-741, 1984.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: BERNARDO, J. M. et al. (Ed.). **Bayesian statistics**. New York: Oxford University, 1992. p. 625-631.

GOMES, R. P. **O coqueiro-da-baía**. 5. ed. São Paulo: Nobel, 1984. 111 p.

GUEDES, M. H. P. et al. Análise Bayesiana da curva de crescimento de cordeiros da raça Santa Inês. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 57, n. 3, p. 415-417, maio/jun. 2005.

_____. 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. 383-390, mar./abr. 2004.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, London, v. 57, p. 97-109, 1970.

HEIDELBERG, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations Research**, Baltimore, v. 31, n. 6, p. 1109-1114, Nov./Dec. 1983.

HOFFMANN, R.; VIEIRA, S. **Análise de regressão: uma introdução à econometria**. 3. ed. São Paulo: Hucitec, 1998. 379 p.

JESUS, N. de; MARTINS, A. B. G.; BARBOSA, J. C. Desenvolvimento de frutos de Longan (*Dimocarpus longan* Lour) na região de Jaboticabal, SP. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 30, n. 1, p. 159-164, fev. 2008.

KAYS, S. J. **Postharvest physiology of perishable plant products**. Athens: AVI, 1997. 532 p.

KVET, J. et al. Methods of growth analysis. In: SESTÁK, Z.; CATSKÝ, J.; JARVIS, P. G. (Ed.). **Plant photosynthetic production: manual of methods**. The Hague: W. Junk, 1971. p. 343-391.

MACHADO, E. J. **Uso do 'Bootstrap' na estimação de parâmetros em modelos não-lineares: uma aplicação em mecanismos cinéticos de Michaelis-Menten**. 2006. 131 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2006.

MAGALHÃES, J. A. S. de. **Características físicas e químicas da água e do albúmen de cultivares de coco (*Cocos nucifera* L.) provenientes da estação experimental da PESAGRO RJ/Campos**. 1999. 60 f. Dissertação (Mestrado em Produção Vegetal) - Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, 1999.

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, set. 2008.

MAZZINI, A. R. de A. et al. Curva de crescimento de novilhos Hereford: heterocedasticidade e resíduos autoregressivos. **Ciência Rural**, Santa Maria, v. 35, n. 2, p. 422-427, mar./abr. 2005.

MENDES, P. N. et al. Análise da curva de crescimento difásica de fêmeas Hereford por meio da função não-linear de Gompertz. **Ciência Animal Brasileira**, Goiânia, v. 10, n. 2, p. 454-461, abr./jun. 2009.

METROPOLIS, N. et al. Equations of state calculations by fast computing machines. **Journal of Chemical Physics**, Woodbury, v. 21, n. 6, p. 1087-1092, 1953.

MORETTIN, P. A.; TOLOI, C. M. de C. **Previsão de séries temporais**. 3. ed. São Paulo: Atual, 2004. 436 p.

NETER, J.; WASSERMAN, W.; KUTNER, M. H. **Applied linear statistical models**: regression, analysis of variance, and experimental designs. Homewood: R. D. Irwin, 1985. 1127 p.

NOGUEIRA, D. A.; SÁFADI, T.; FERREIRA, D. F. Avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov. **Revista Brasileira de Estatística**, Rio de Janeiro, v. 65, n. 224, p. 59-88, 2004.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 429 p.

PEREIRA, A. R.; MACHADO, E. C. **Análise quantitativa do crescimento de comunidades vegetais**. Campinas: IAC, 1987. 33 p. (Boletim Técnico, 114).

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.

PEREIRA, M. C. T. et al. Crescimento e produção de primeiro ciclo da bananeira 'Prata Anã'(AAB) em sete espaçamentos. **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 7, p. 1377-1387, jul. 2000.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, 2008. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 out. 2010.

RAFTERY, A. L.; LEWIS, S. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. **Statistical Science**, Hayward, v. 7, n. 4, p. 493-497, Aug. 1992.

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.

SEARLE, S. R. **Linear models**. New York: J. Wiley, 1971. 532 p.

SILVA, N. A. M. et al. Modelo hierárquico Bayesiano aplicado na avaliação genética de curvas de crescimento de bovinos de corte. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 62, n. 2, p. 409-418, abr. 2010.

STATISTICAL ANALYSIS SYSTEM INSTITUTE. **SAS procedures guide for computers**. Cary, 1999. v. 3, 373 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 (*Phoenix roebelenii* O'BRIEN). **Magistra**, Cruz das Almas, v. 22, n. 1, p. 1-7, 2010.

TORALLES, R. P. et al. Caracterização parcial do escurecimento enzimático pela polifenoloxidase em pêssegos das cv. Granada, Jade, Esmeralda e Maciel. **Revista Brasileira de Agrociência**, Pelotas, v. 10, n. 1, p. 241-244, mar. 2004.

URCHEI, M. A.; RODRIGUES, J. D.; STONE, L. F. Análise de crescimento de duas cultivares de feijoeiro sob irrigação, em plantio direto e preparo convencional. **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 3, p. 497-506, mar. 2000.

VIEIRA, S.; MISCHAN, M. M. A logística e a Gompertz: duas funções alternativas no estudo de dados de crescimento. **Ciência e Cultura**, São Paulo, v. 28, n. 8, p. 950-952, ago. 1976.

WARRINGTON, L. J. et al. Apple fruit growth and maturity are affected by early season temperatures. **Journal American Society Horticultural Science**, Alexandria, v. 124, n. 5, p. 468-477, Sept. 1999.