



PATRÍCIA NEVES MENDES

**AJUSTE DOS MODELOS GOMPERTZ E
LOGÍSTICO NA DESCRIÇÃO DAS CURVAS DE
CRESCIMENTO DE TOMATES: UMA
ABORDAGEM BAYESIANA**

**LAVRAS – MG
2011**

PATRÍCIA NEVES MENDES

**AJUSTE DOS MODELOS GOMPERTZ E LOGÍSTICO NA DESCRIÇÃO
DAS CURVAS DE CRESCIMENTO DE TOMATES: UMA
ABORDAGEM BAYESIANA**

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

Orientador

Joel Augusto Muniz

Co-orientadora

Thelma Sáfadi

LAVRAS – MG

2011

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

Mendes, Patrícia Neves.

Ajuste dos Modelos Gompertz e Logístico na descrição das
curvas de crescimento de tomates : uma abordagem bayesiana /
Patrícia Neves Mendes. – Lavras : UFLA, 2011.

82 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2011.

Orientador: Joel Augusto Muniz.

Bibliografia.

1. Modelos não-lineares. 2. Metodologia bayesiana. 3. Tomate
Santa Clara. I. Universidade Federal de Lavras. II. Título.

CDD – 519.542

PATRÍCIA NEVES MENDES

**AJUSTE DOS MODELOS GOMPERTZ E LOGÍSTICO NA DESCRIÇÃO
DAS CURVAS DE CRESCIMENTO DE TOMATES: UMA
ABORDAGEM BAYESIANA**

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

APROVADA em 11 de fevereiro de 2011.

Dr. Augusto Ramalho de Moraes UFLA

Dr. Janser Moura Pereira UFU

Dra. Taciana Villela Savian Esalq USP

Dra. Thelma Sáfyadi UFLA

Joel Augusto Muniz

Orientador

LAVRAS – MG

2011

AGRADECIMENTOS

A Deus, pela vida e família maravilhosas que tenho.

À Universidade Federal de Lavras, especialmente ao Departamento de Ciências Exatas, pela oportunidade de realizar este curso.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

Ao meu orientador, professor Joel Augusto Muniz, pela orientação, amizade, ensinamentos e apoio durante todo o curso.

À professora Thelma que sempre me ajudou nas horas difíceis e muito contribuiu com este trabalho.

Aos meus amigos incondicionais de Lavras, os quais nunca serão esquecidos, tanto pelas horas e horas de estudo, quanto pelas festas e risadas, que juntos me ajudaram a chegar ao final deste trabalho.

Aos amigos Taciana e Anderson, aos quais tenho muito a agradecer pela imensa contribuição profissional que sempre me deram.

Aos meus queridos avós Elba e José Belchior, pelo amor e compreensão, nas muitas vezes em que estive ausente devido aos estudos.

Ao meu irmão, Talles, que juntamente com a Juliana sempre me apoiaram. E hoje eles têm o maior presente da vida, a Camilinha.

Ao meu namorado Omar, que me acompanhou na fase mais difícil do trabalho e me deu muito apoio, carinho e amor. E ainda me ajudou a ter um tempo de diversão e alegria, mesmo com tantas obrigações e pressões.

Aos meus pais, Mara e Bolivar, aos quais não tenho palavras para agradecer todo o suporte, carinho, força, compreensão e amor oferecidos durante a minha vida. Sempre presentes nos momentos mais difíceis sem medir esforços para a minha felicidade. Essa vitória é de vocês.

RESUMO

Conhecer o crescimento do tomateiro e de seus frutos, medido através do acúmulo de biomassa ao longo do tempo, é fundamental para o manejo adequado e detecção de problemas no desenvolvimento da cultura. Este crescimento pode ser estudado por meio de vários modelos de regressão não-linear que facilitam a interpretação dos processos envolvidos no sistema de produção vegetal. Entre os modelos empíricos usados para estimar o crescimento de plantas, e de seus componentes, encontra-se a função de Gompertz e a logística. Um dos métodos utilizados para estimar os parâmetros das curvas de crescimento é o método bayesiano. Este estudo teve como objetivo aplicar a metodologia bayesiana na descrição dos dados - reais e simulados - de crescimento do diâmetro de tomates, utilizando os modelos de Gompertz e logístico. Os algoritmos para o Amostrador de Gibbs e o Metropolis-Hastings foram implementados no software livre R e por meio deles foram obtidas as amostras das distribuições marginais *a posteriori* dos parâmetros. A condição de convergência das cadeias foi verificada por meio dos critérios sugeridos por Nogueira, Sáfiadi e Ferreira (2004) disponíveis no pacote BOA do software R. Os modelos foram comparados pelo Fator de Bayes. A metodologia bayesiana mostrou-se eficiente, uma vez que foi avaliada e comprovada pelo processo de simulação, apresentando estimativas bem próximas do valor paramétrico, e as estimativas mostraram-se coerentes com os valores relatados na literatura. O modelo logístico foi considerado mais adequado que o modelo de Gompertz na descrição do crescimento de tomates.

Palavras-chave: Metodologia bayesiana. Modelo de Gompertz e logístico. Curvas de crescimento. Tomates.

ABSTRACT

Knowing the growth of tomato and its fruits, as measured by biomass accumulation over time is essential for the proper handling and detection of problems in the development of crops. This growth can be studied using various models of non-linear regression that can be used to facilitate interpretation of the processes involved in plant production system. Among the empirical models often used to estimate the growth of plants and their components is the function of Gompertz and Logistic. One method used to estimate the parameters of the growth rate is the Bayesian method. The study objective to apply the Bayesian approach in describing the data – real and simulated – the diameter growth of tomatoes, using the models of Gompertz and Logistic. The algorithms for the Gibbs Sampler and Metropolis – Hastings were implemented using the R language, and through them were obtained the samples of posterior marginal distributions for the parameters. The condition of convergence of the chains was checked using the criteria suggested by Nogueira, Sáfadi and Ferreira (2004) available on the R software package BOA. The models were compared by using Bayes factor. The Bayesian approach was efficient, since it was evaluated and verified by the simulation process, with very close estimates of the parametric value, and estimates were shown to be consistent with the values reported in literature. The Logistic model was considered more likely than the Gompertz model in describing the growth of tomatoes.

Keywords: Bayesian methodology. Gompertz and Logistic models. Growth curves. Tomatoes.

LISTA DE TABELAS

Tabela 1	Equações para os casos especiais da função de Richards.....	22
Tabela 2	Interpretação do fator de Bayes, segundo Jeffreys (1961).....	35
Tabela 3	Estimativas da média e moda <i>a posteriori</i> , intervalo de máxima densidade <i>a posteriori</i> (HPD) para os dados simulados segundo o modelo de Gompertz	50
Tabela 4	Crítérios de Geweke (valor p) e Gelman & Rubin (R) utilizados na detecção da convergência das cadeias amostradas para os dados simulados do modelo de Gompertz.....	52
Tabela 5	Estimativas da média e moda <i>a posteriori</i> , intervalo de máxima densidade <i>a posteriori</i> (HPD) para os dados simulados segundo o modelo logístico.....	55
Tabela 6	Crítérios de Geweke (valor p) e Gelman & Rubin (R) utilizados na detecção da convergência das cadeias amostradas para os dados simulados do modelo logístico.....	56
Tabela 7	Estimativas da média e moda <i>a posteriori</i> , intervalo de máxima densidade <i>a posteriori</i> (HPD) para os dados reais segundo o modelo de Gompertz.....	59
Tabela 8	Crítérios de Geweke (valor p) e Gelman & Rubin (R) utilizados na detecção da convergência das cadeias amostradas para os dados reais do modelo de Gompertz.	62

Tabela 9	Estimativas da média e moda <i>a posteriori</i> , intervalo de máxima densidade <i>a posteriori</i> (HPD) para os dados reais segundo o modelo logístico.....	65
Tabela 10	Crítérios de Geweke (valor p) e Gelman & Rubin (<i>R</i>) utilizados na detecção da convergência das cadeias amostradas para os dados reais do modelo logístico.....	68
Tabela 11	Fator de Bayes (FB) para a comparação dos modelos de Gompertz e logístico no ajuste dos dados reais.....	72

LISTA DE FIGURAS

Figura 1	Varição do diâmetro longitudinal dos tomates durante 77 dias de desenvolvimento.....	41
Figura 2	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>a</i> do modelo de Gompertz, referente ao cenário 1.....	52
Figura 3	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>b</i> do modelo de Gompertz, referente ao cenário 1.....	53
Figura 4	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>k</i> do modelo de Gompertz, referente ao cenário 1.....	53
Figura 5	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>tau</i> do modelo de Gompertz, referente ao cenário 1.....	54
Figura 6	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>a</i> do modelo logístico, referente ao cenário 1.....	57
Figura 7	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>b</i> do modelo logístico, referente ao cenário 1.....	57
Figura 8	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>k</i> do modelo logístico, referente ao cenário 1.....	58
Figura 9	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>tau</i> do modelo logístico, referente ao cenário 1.....	58

Figura 10	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>a</i> do modelo de Gompertz referente aos dados reais.....	63
Figura 11	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>b</i> do modelo de Gompertz referente aos dados reais.....	63
Figura 12	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>k</i> do modelo de Gompertz referente aos dados reais.....	64
Figura 13	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>tau</i> do modelo de Gompertz referente aos dados reais.....	64
Figura 14	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>a</i> do modelo logístico referente aos dados reais.....	69
Figura 15	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>b</i> do modelo logístico referente aos dados reais.....	69
Figura 16	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>k</i> do modelo logístico referente aos dados reais.....	70
Figura 17	Densidade <i>a posteriori</i> e traço da cadeia gerada pelo método MCMC para o parâmetro <i>tau</i> do modelo logístico referente aos dados reais.....	70

SUMÁRIO

1	INTRODUÇÃO.....	13
2	REFERENCIAL TEÓRICO.....	16
2.1	Modelos não-lineares.....	16
2.2	Modelos de crescimento vegetal.....	19
2.3	Inferência Bayesiana.....	23
2.3.1	Teorema de Bayes.....	26
2.3.2	Métodos de Monte Carlo via Cadeias de Markov.....	28
2.3.2.1	Amostrador de Gibbs.....	28
2.3.2.2	Metropolis-Hastings.....	30
2.3.3	Avaliação da convergência.....	31
2.3.3.1	Monitoração informal da convergência	32
2.3.3.2	Monitoração formal da convergência.....	32
2.3.4	Seleção de modelos.....	34
3	MATERIAL E MÉTODOS.....	37
3.1	Modelos estudados.....	37
3.2	Dados simulados.....	39
3.3	Dados reais.....	40
3.4	Implementação dos algoritmos MCMC.....	41
4	RESULTADOS E DISCUSSÃO.....	43
4.1	Modelo de Gompertz.....	43
4.1.1	Distribuições condicionais completas para o modelo de Gompertz.....	44
4.2	Modelo logístico.....	46
4.2.1	Distribuições condicionais completas para o modelo logístico..	47
4.3	Resultados obtidos para os dados simulados.....	49
4.3.1	Modelo de Gompertz.....	50

4.3.2	Modelo logístico.....	54
4.4	Resultados obtidos para os dados reais.....	59
4.4.1	Modelo de Gompertz.....	59
4.4.2	Modelo logístico.....	65
4.5	Comparação dos modelos estudados.....	71
5	CONCLUSÕES.....	73
	REFERÊNCIAS.....	74

1 INTRODUÇÃO

O tomate é um fruto - uma vez que é o produto do desenvolvimento do ovário e do óvulo da flor, formando o pericarpo e as sementes, respectivamente, após a fecundação - embora coloquialmente considerado como legume pelos leigos. De sua família (Família Solanáceas), fazem parte também as batatas, as berinjelas, as pimentas e os pimentões, além de algumas espécies não-comestíveis. Originário da América Central e do Sul, era amplamente cultivado e consumido pelos povos pré-colombianos, sendo atualmente cultivado e consumido no mundo todo.

Este é hoje a mais importante hortaliça do mundo todo, tanto por área cultivada, como por valor comercial. O Brasil figura como um de seus maiores produtores. Em 2008, o Brasil destacou-se como o sexto produtor mundial de tomate, com produção de cerca de 3 milhões de toneladas, em aproximadamente 57,6 mil hectares (ANUÁRIO..., 2008).

Em relação às suas características nutricionais, o fruto é importante fonte de vitaminas A, B e C, e de sais minerais como Fósforo, Ferro, Potássio e Magnésio. A vitamina A é indispensável para a normalidade da vista, mucosas e pele, auxilia o crescimento e evita infecções. As vitaminas do Complexo B ajudam na regularização do sistema nervoso e aparelho digestivo, tonificam o músculo cardíaco, colaboram para a pele e para o crescimento. Já a vitamina C, principal componente do tomate, dá resistência aos vasos sanguíneos, vitalidade às gengivas, evita a fragilidade dos ossos e má formação dos dentes, contribuindo no combate a infecções e cicatrização de ferimentos. O fruto é contraindicado para pessoas que sofram de fermentações gástricas e acidez no estômago.

Diante do exposto, o conhecimento do crescimento do tomateiro e de seus frutos, medido pelo acúmulo de biomassa ao longo do tempo, é

fundamental para o manejo adequado e a detecção de problemas no desenvolvimento das culturas. Na prática, estes problemas podem ser caracterizados pela ocorrência de deficiências hídricas e nutricionais, competição de plantas daninhas, solos compactados ou mal drenados. Assim, o desenvolvimento e o aperfeiçoamento de metodologias que permitam estudar de forma prática e eficiente este crescimento constituem uma importante fonte de pesquisa na área de Fitotecnia.

Uma destas metodologias pode ser representada pela análise de curvas de crescimento, a qual engloba principalmente o ajuste de vários modelos de regressão não-linear que podem ser usados, de acordo com suas habilidades, para facilitar a interpretação dos processos envolvidos no sistema de produção vegetal. Entre os modelos de crescimento, destacam-se o logístico e de Gompertz, os quais fornecem informações úteis fundamentadas particularmente nas estimativas de seus parâmetros, e apresentam interpretação biológica, tais como taxas de crescimento e acúmulo máximo de biomassa.

A avaliação dos processos de desenvolvimento de uma planta ou 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. Parâmetros relativos à maturação e idade de frutos e hortaliças são importantes indicativos de que o produto encontra-se pronto para ser retirado do campo e apto para o consumo.

Nem sempre, porém, é uma tarefa fácil ajustar modelos aos conjuntos de dados de crescimento, pois o método de estimação frequentista geralmente utilizado necessita de processos iterativos cuja convergência às vezes não é observada e, além disso, pressuposições básicas a respeito da validade do modelo como a independência dos resíduos quase nunca é considerada.

Entre os métodos utilizados para estimar os parâmetros das curvas de crescimento, o bayesiano mostra-se como um dos mais eficientes. Este consiste

de informações referentes aos dados amostrais do experimento (função de verossimilhança), do conhecimento prévio do pesquisador com respeito aos parâmetros do modelo (distribuição *a priori*) e do cálculo da densidade *a posteriori* dos parâmetros. Na inferência bayesiana, considera-se que tanto as observações como os parâmetros de um modelo estatístico são variáveis aleatórias.

Tendo em vista que no Brasil há poucos relatos de trabalhos sobre curvas de crescimento que tenham abordado, conjuntamente, os modelos não-lineares e a análise bayesiana na área de Fitotecnia, este trabalho vem como um método alternativo para o estudo de curvas de crescimento de tomates, o qual poderá ser estendido às mais variadas culturas.

Desta maneira, o presente trabalho foi realizado com o objetivo de estudar as curvas de crescimento referentes ao diâmetro longitudinal de tomates por meio dos modelos não-lineares logístico e de Gompertz e de apresentar um procedimento bayesiano para a análise e a comparação entre esses modelos.

2 REFERENCIAL TEÓRICO

2.1 Modelos não-lineares

Uma atividade muito comum em análises estatísticas é a avaliação da relação entre a variável dependente e uma ou mais variáveis independentes. Essa análise pode ser feita por meio dos modelos de regressão. Segundo Draper e Smith (1998), os modelos de regressão 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).

Os modelos de crescimento vegetal são classificados como modelos de regressão linear e não-linear e apresentam algumas peculiaridades quanto ao procedimento de ajuste. Sob o ponto de vista da inferência 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 mais simples, assume-se 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 Toloi (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, quando se trabalha com a estatística clássica, o método dos quadrados mínimos generalizado é mais eficiente do que o método dos quadrados mínimos ponderados e ordinários.

Pereira, Muniz e Silva (2005), estudando a predição do nitrogênio mineralizado em latossolo pelo uso de modelos não-lineares, incorporaram a autocorrelação residual e encontraram estimativas mais precisas, ou seja, com menor erro, para os parâmetros envolvidos nos modelos.

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.

Os modelos não-lineares são largamente estudados no ajuste de curvas de crescimento animal (GUEDES et al., 2005; MAZZINI et al., 2003, 2005; MENDES; AQUINO; MUNIZ, 2001; MENDES et al., 2008, 2009; SILVA et al., 2007), bem como na descrição de cinética de digestão de animais ruminantes (SAVIAN et al., 2007a, 2007b, 2007c, 2009); são utilizados também na avaliação genética de curvas de crescimento de bovinos (SILVA et al., 2010) nas curvas de lactação de animais (MUNIZ et al., 2007; SILVA et al., 2005) e no estudo da técnica de platô de resposta para avaliar a exigência de zinco em frangos de corte (REZENDE et al., 2007). Os modelos não-lineares são aplicados também no estudo da extração de zinco de um lodo de esgoto (SOUZA et al., 2010).

Na área de produção vegetal, em diversas situações experimentais, os fenômenos também 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 - (MACHADO, 2006); 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, pode ser utilizado como indicador de possível escurecimento em polpa. Machado (2006) estudou diferentes parametrizações para o processo de estimação de parâmetros em mecanismos cinéticos de Michaelis-Menten.

Alguns autores já têm trabalhado com modelos não-lineares em detrimento aos lineares na área vegetal. Entre eles pode-se citar: Vieira Júnior et al. (1999), estudando o potencial total de água em sementes; Corrêa, Araújo e Afonso Júnior (2003), na determinação dos parâmetros de secagem de sementes de milho doce (*Zea mays* L.); Paixão et al. (2004), ao estimarem a infiltração de água no solo por meio de ajustes de funções não-lineares e de modelos empíricos; Fideles Filho et al. (2005), ao avaliarem o índice de massa foliar e biomassa aérea da cultura do feijoeiro irrigado; Linhares, Silva e Lucas (2005), no ajuste da curva de degradação de inseticida em frutos de berinjela, utilizando o modelo exponencial de primeira ordem; Carlesso et al. (2007), comparando cinco modelos não-lineares com o objetivo de determinar o modelo que melhor descreve a secagem por convecção em camada fina de sementes de maracujá-amarelo; Pereira et al. (2009) e Pereira, Muniz e Silva (2005), no estudo da dinâmica e disponibilização de nutrientes no sistema solo-planta e na

comparação entre modelos para predição do nitrogênio mineralizado utilizando a metodologia bayesiana, respectivamente; Martins Filho et al. (2008) comparando as curvas de crescimento de duas cultivares de feijoeiro; Terra, Muniz e Savian (2010), estudando o crescimento dos frutos da tamareira-anã.

O estudo do crescimento de frutos e vegetais tem uma larga aplicação na pesquisa agrícola, pois possibilita ao pesquisador comparar o comportamento de uma planta, ou de suas partes de interesse, frente às mais diversas situações empregadas. Desta maneira, têm sido utilizados vários modelos não-lineares para descrever curvas de crescimento das mais variadas espécies. A diversidade de modelos e a necessidade de se comparar curvas originadas de diferentes tratamentos fazem com que pesquisas nesta área tenham grande demanda de métodos estatísticos.

2.2 Modelos de crescimento vegetal

Segundo Fitzhugh Junior (1976), uma curva sigmóide descrevendo uma sequência de medidas de tamanho, frequentemente peso corporal, pode denominar o termo “curvas de crescimento” ou “curvas tamanho-idade”. Este tipo de dado pode ser caracterizado de três formas: estático, seccional-cruzado e longitudinal. No tipo estático, para um determinado grupo de indivíduos com mesma idade ou estágio de desenvolvimento, o tamanho é observado uma única vez em cada indivíduo. No tipo seccional-cruzado, embora cada indivíduo seja pesado apenas uma vez, assim como no tipo estático, as mesmas medidas são feitas em outras idades em outros indivíduos vindos de amostras da mesma população. Por fim, no tipo longitudinal, todas as medidas utilizadas são obtidas do mesmo indivíduo, em todas as idades avaliadas. Este último tipo de dados abrange todas as informações contidas nos dois primeiros tipos.

O interesse pelas curvas de crescimento vem aumentando nos últimos anos, principalmente pelo desenvolvimento de novas técnicas computacionais que permitem maior rapidez e precisão das análises (SILVA; AQUINO; OLIVEIRA, 2001).

Segundo Kvet et al. (1971), a análise de crescimento de comunidades vegetais é um dos primeiros passos na análise de produção primária, caracterizando-se como o 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 análise de 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.

O estágio de desenvolvimento dos frutos no momento da colheita tem influência direta 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. Os frutos colhidos muito maduros deterioram-se rapidamente, não podendo ser armazenados e/ou comercializados em locais distantes (KAYS, 1997).

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 à época

da colheita. Porém, estes índices podem variar consideravelmente, dependendo do local de cultivo, cultivares e condições climáticas do ano de crescimento. Devem-se manter os frutos na planta até atingirem a completa maturação, permitindo um desenvolvimento normal e completo das sementes (MELO, 2007).

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, período no qual se fazem as medidas ou avaliações das características que se quer avaliar (WARRINGTON et al., 1999).

Gomes et al. (2005), ao estudarem a influência do manejo do solo e intensidade de raleio de frutas no crescimento e qualidade de cultivares de pêssago, destacam que conhecer o crescimento da fruta e os fatores que o afetam é um requisito necessário para entender o efeito da poda, fertilização do solo e raleio das frutas, bem como para prever o seu tamanho final. Neste trabalho, as curvas de crescimento foram ajustadas por polinômios ortogonais às medidas do diâmetro sutural das frutas, em cada semana.

No estudo de crescimento de feijoeiro, 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. Neste mesmo trabalho, os autores ajustaram uma equação exponencial quadrática em função do tempo, para a matéria seca total da planta.

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, Scarpare Filho e Verdial (2005), estudando a fenologia do caquizeiro “Rama Forte” em clima tropical no Brasil, encontraram para as curvas de crescimento do fruto um padrão sigmoidal duplo.

Ao estudarem o crescimento e desenvolvimento de tomates “Santa Clara”, Moura et al. (2004) encontraram um comportamento sigmoidal de crescimento tanto para o diâmetro transversal quanto para o diâmetro longitudinal.

No estudo de curvas de crescimento de duas cultivares de feijoeiro, Martins Filho et al. (2008) utilizaram o modelo não-linear logístico e a metodologia bayesiana na estimação dos parâmetros.

Vários tipos de modelos estatísticos podem ser usados, de acordo com suas habilidades, para facilitar a interpretação dos processos envolvidos no sistema de produção vegetal. Os principais modelos estão na Tabela 1:

Tabela 1 Equações para os casos especiais da função de Richards.

Função	Equações	Valor de M
Brody	$a[1-b\exp(-kt)]$	1
Von Bertalanffy	$a[1-b\exp(-kt)]^3$	3
Logística	$a[1+b\exp(-kt)]^{-1}$	-1
Gompertz	$a\exp[-b\exp(-kt)]$	∞
Richards	$a[1-b\exp(-kt)]^M$	Variável

Fonte: Fitzhugh Junior (1976)

De acordo com Silva, Aquino e Oliveira (2001), esses modelos não-lineares apresentam dois parâmetros interpretáveis biologicamente e um outro que se identifica como uma constante matemática. O parâmetro “a”, definido como peso assintótico ou peso adulto, representa a estimativa de peso à maturidade, independente de flutuações de pesos devido a efeitos genéticos e

ambientais. O parâmetro “k” corresponde ao índice de maturidade ou à estimativa de precocidade de maturidade e determina a eficiência do crescimento de um vegetal. Quanto maior for o valor desse parâmetro, mais precoce é o vegetal e vice-versa. O parâmetro “b” é denominado de parâmetro de integração ou interceptação com o eixo y, não possui significado biológico e é utilizado apenas para adequar o valor inicial do peso, fazendo com que a curva passe pela origem quando $y \neq 0$ e/ou $t \neq 0$.

O modelo Richards apresenta um parâmetro a mais que os outros. Esse parâmetro “M” refere-se ao ponto de inflexão da curva que representa o grau de maturidade do fruto. Freitas e Costa (1983 citados por RODRIGUES; MUNIZ; PEREIRA, 1992), explicam que o ponto de inflexão ocorre quando a alteração na taxa de crescimento é máxima, ou seja, quando os acréscimos da taxa de crescimento diário, em função da idade, passam de crescentes para decrescentes.

Segundo Vieira e Mischán (1976 citados por GUEDES et al., 2004), 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, 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.

Segundo Mazzini et al. (2005), a função de Gompertz foi inicialmente utilizada por Gompertz, Winsor e Laird para descrever a taxa de mortalidade numa população. A aplicação desta função em modelos sigmodais de crescimento foi sugerida por Duarte (1975).

2.3 Inferência Bayesiana

Segundo Paulino, Turkman e Murteira (2003), a abordagem clássica – em que os parâmetros desconhecidos são considerados como fixos e toda a

análise se restringe àquelas informações contidas na amostra dos dados – foi adotada, de forma quase unânime, pelos estatísticos, durante a primeira metade do século XX. Mais tarde, com os avanços computacionais, surgiu como alternativa a inferência bayesiana, que trata o parâmetro desconhecido como quantidade aleatória.

Durante um grande período, a metodologia bayesiana ficou resguardada por necessitar de resoluções matemáticas, mais precisamente de integrações, inviáveis de serem feitas algebricamente. Por volta da década de 60, ela ressurgiu em alguns trabalhos teóricos como o de Jeffreys (1961), mas somente em 1990, com o trabalho de Gelfand e Smith (1990), é que a proposta para a resolução do problema de integrais complexas foi solucionado atraindo a atenção de muitos outros pesquisadores. Atualmente, principalmente devido ao avanço computacional, a metodologia bayesiana constitui um dos principais assuntos da comunidade científica envolvida com o desenvolvimento e aplicação de procedimentos estatísticos.

Os artigos que abordam a estimação de parâmetros nos modelos estatísticos por este método têm tido grande sucesso devido ao fato de que todos os parâmetros podem ser usados dentro do modelo como variáveis aleatórias, o que requer um menor número de dados, pois os conceitos probabilísticos envolvidos diminuem a dependência do ajuste do modelo em relação ao número de observações utilizadas (SILVA et al., 2006).

Os modelos não-lineares que descrevem o crescimento de frutas são discutidos na literatura por meio da abordagem de inferência clássica, que tem por característica uma forte dependência dos dados experimentais. A metodologia bayesiana pode ser empregada principalmente em situações em que se utilizam experimentos pequenos, com poucas medidas no tempo, portanto mais econômicos e que seriam inviáveis de se analisar com estatística clássica.

Em estudos recentes envolvendo ajuste de modelos não-lineares, a inferência bayesiana foi utilizada com sucesso, reduzindo o número de estimativas viesadas, mesmo utilizando poucas informações (MARTINS FILHO et al., 2008).

A eficiência do método Bayesiano foi verificada por Simões (1981), no estudo de modelos de regressão linear, usando a técnica de simulação de dados. O autor concluiu que as estimativas obtidas tendem mais rapidamente ao valor verdadeiro do parâmetro quando comparadas ao método frequentista.

Para verificar o comportamento desse método em estudos de regressão linear em ensaios biológicos, Mendoza (1990) também utilizou a técnica de simulação, e destacou que a estimativa de índices biológicos tem sido bastante controversa quando a análise é conduzida sob uma abordagem frequentista.

A metodologia bayesiana consiste de informação referente aos dados amostrais (função de verossimilhança), do conhecimento prévio a respeito dos parâmetros (distribuição *a priori*) e do cálculo da densidade a posteriori dos parâmetros.

A função de verossimilhança é a densidade conjunta de um grupo de observações y_1, \dots, y_n , examinada como uma função dos parâmetros.

A utilização de informação *a priori* em inferência bayesiana requer a especificação de uma distribuição *a priori* para os parâmetros, devendo representar probabilisticamente o conhecimento que se tem sobre ele antes da realização do experimento. Este conhecimento a respeito do parâmetro pode ser obtido de análises anteriores, experiência do pesquisador na área em questão ou publicações sobre o assunto que se deseja pesquisar ou estudar. Quando em determinado estudo o pesquisador tem pouca ou nenhuma informação para se incorporar à distribuição *a priori*, pode-se considerar o que é chamado de *priori* não-informativa, a qual é representada principalmente pela *priori* de Jeffreys (JEFFREYS, 1961).

Generalizando, a distribuição *a priori* é o único fator que diferencia a estatística clássica da bayesiana, e por isso ela é alvo de crítica para muitos estatísticos clássicos, que alegam que a situação *a priori* é um processo subjetivo, já que dois pesquisadores podem ter graus de incertezas diferentes sobre uma determinada quantidade (PAULINO; TURKMAN; MURTEIRA, 2003).

2.3.1 Teorema de Bayes

O teorema de Bayes é o precursor da inferência bayesiana e pode ser considerado como um mecanismo de atualização da opinião do pesquisador sobre o parâmetro θ . Por meio deste teorema, combina-se a distribuição *a priori* com a informação contida nos dados amostrais (função de verossimilhança), obtendo-se uma distribuição *a posteriori*. Sendo assim, toda inferência é realizada usando-se a distribuição *a posteriori* e esta informação pode ser resumida pela média, mediana, moda ou pelos intervalos de credibilidade (PAULINO; TURKMAN; MURTEIRA, 2003).

O teorema de Bayes, apresentado nas expressões (1) e (2) a seguir, é uma proposição sobre probabilidades condicionadas. Para o caso em que o parâmetro θ é contínuo, o teorema é dado por:

$$P(\theta | Y) = \frac{L(\theta | Y)P(\theta)}{\int L(\theta | Y)P(\theta)d\theta} \quad (1)$$

sendo $Y = \{y_1, y_2, \dots, y_n\}$ um vetor de observações, $P(\theta)$ *a priori*, $L(\theta | Y)$ a verossimilhança e no caso em que θ é discreto, tem-se:

$$P(\theta | Y) = \frac{L(\theta | Y)P(\theta)}{\sum L(\theta | Y)P(\theta)} \quad (2)$$

Como nas expressões (1) e (2), o denominador não depende do

parâmetro θ , este pode ser considerado como uma constante, e então, estas expressões podem ser representadas por:

$$P(\theta|Y) \propto L(\theta|Y)P(\theta) \quad (3)$$

ou seja, a expressão (3) pode ser entendida como:

distribuição a posteriori \propto verossimilhança x distribuição a priori ,
em que: \propto representa proporcionalidade.

Segundo Paulino, Turkman e Murteira (2003), a função de verossimilhança tem importante papel no teorema de Bayes, pois representa o meio através do qual os dados (Y) transformam o conhecimento *a priori* sobre o parâmetro. Em resumo, a distribuição *a posteriori* incorpora, via Teorema de Bayes, toda a informação disponível sobre o parâmetro, isto é, a informação inicial mais a informação da experiência ou amostra.

Rosa (1998) afirma que, para se inferir com relação a qualquer elemento de um vetor de parâmetros θ , a distribuição *a posteriori* conjunta dos parâmetros, $P(\theta|Y)$, deve ser integrada em relação a todos os outros elementos que a constituem. Assim, se o interesse do pesquisador se concentra sobre determinado elemento de θ , por exemplo, θ_1 , tem-se a necessidade da obtenção da distribuição $P(\theta_1|Y)$, denominada de distribuição marginal *a posteriori*, a qual é dada por:

$$P(\theta_1|Y) = \int_{\theta \neq \theta_1} P(\theta|Y) d\theta$$

A integração dessa distribuição, geralmente, não é analítica, necessitando de métodos numéricos para a sua resolução. Em geral, utilizam-se algoritmos iterativos especializados denominados de algoritmos MCMC (*Markov Chain Monte Carlo*). Para a utilização desses algoritmos, é necessário que se obtenha, a partir da distribuição *a posteriori*, um conjunto de

distribuições chamadas de distribuições condicionais completas (SILVA et al., 2008).

2.3.2 Métodos de Monte Carlo via cadeias de Markov

Uma cadeia de Markov é um processo estocástico no qual o estado futuro da cadeia θ^{k+1} depende somente do estado atual θ^k e dos dados, e não da história passada da cadeia (GAMERMAN, 1997). Para o autor, as primeiras iterações são influenciadas pelo estado inicial $\theta^{(1)}$ e são descartadas. Este período é conhecido como aquecimento da cadeia ou *burn-in*. Também são consideradas dependências entre as observações subsequentes da cadeia e, para se obter uma amostra independente, as observações finais devem ser obtidas a cada k iterações, sendo este valor conhecido como salto, *thin* ou intervalo de amostragem (PAULINO; TURKMAN; MURTEIRA, 2003).

Os métodos MCMC são uma alternativa aos métodos não iterativos, tendo em vista a resolução de integrais complexas. A ideia destes métodos é obter uma amostra das distribuições marginais *a posteriori* dos parâmetros de interesse, por meio de um processo iterativo, utilizando as distribuições condicionais completas de cada parâmetro. Por sua vez, esses valores gerados são considerados amostras aleatórias de uma determinada distribuição de probabilidade, caracterizando, assim, o método de simulação Monte Carlo. Dessa forma, tem-se uma ação conjunta dos métodos, que resulta no processo MCMC, cujos principais algoritmos são o Amostrador de Gibbs e o Metropolis-Hastings.

2.3.2.1 Amostrador de Gibbs

O Amostrador de Gibbs (A.G.) tem sido extremamente útil na resolução

de problemas que envolvem a estimação de mais de um parâmetro, porém para sua utilização exige-se que as distribuições condicionais completas *a posteriori* dos parâmetros tenham formas conhecidas.

O A.G. é, essencialmente, um esquema iterativo de amostragem de uma cadeia de Markov, cujo núcleo de transição é formado pelas distribuições condicionais completas. É uma técnica para gerar variáveis aleatórias de uma distribuição marginal sem que se conheça a sua densidade (GAMERMAN, 1997). À medida que o número k de iterações aumenta, a sequência de valores gerados se aproxima da distribuição de equilíbrio, ou seja, da densidade marginal desejada para cada parâmetro, quando se assume que a convergência foi atingida.

Abaixo são apresentados os passos para o A.G.:

i) Especificar os valores iniciais $\theta^{(0)} = \theta_1^{(0)}, \dots, \theta_S^{(0)}$ e inicializar o contador de iterações $k = 1$;

ii) Obtenha um novo valor para $\theta^{(k)}$ a partir de $\theta^{(k-1)}$ através de sucessivas gerações de valores. Para $i = 1$ até S , faça:

Gere um valor para $\theta_i^{(k)}$ de:

$$\theta_i^{(k)} \sim P(\theta_i | \theta_1^{(k)}, \dots, \theta_{i+1}^{(k-1)}, \dots, \theta_S^{(k-1)}, Y)$$

iii) Faça $k = k + 1$ e volte para o passo **ii**.

iv) Repita o procedimento até alcançar a convergência.

O conjunto dos i valores amostrados representa amostras da distribuição marginal *a posteriori* de θ . A partir da amostra desta distribuição, obtêm-se as estimativas pontuais (média, mediana e moda) e estimativas por região (intervalo de credibilidade HPD) para os parâmetros amostrados. Segundo Paulino, Turkman e Murteira (2003), o HPD é denominado intervalo de credibilidade de máxima densidade *a posteriori*. E quanto menor for o tamanho do intervalo, mais concentrada é a distribuição do parâmetro e a amplitude do

intervalo informa sobre a dispersão de θ . Para distribuições unimodais e simétricas, o intervalo de credibilidade e o HPD são idênticos. Havendo assimetria, os dois critérios produzem intervalos diferentes, sendo essa diferença dependente do grau de assimetria (ANDRADE, 2008).

2.3.2.2 Metropolis-Hastings

O algoritmo de Metropolis-Hastings é utilizado para a obtenção da amostra da distribuição marginal *a posteriori* quando o Amostrador de Gibbs não se mostra eficiente, ou seja, a distribuição condicional dos parâmetros não se caracteriza como uma distribuição de probabilidade conhecida. Neste caso, geram-se valores do parâmetro a partir de uma distribuição proposta e esse é aceito ou não com uma certa probabilidade de aceitação (CHIB; GREENBERG, 1995). O conjunto final de valores assim obtidos irá representar uma amostra de observações da distribuição condicional *a posteriori*.

Este algoritmo é bastante geral, e pode, pelo menos em princípio, ser implementado com qualquer distribuição condicional completa *a posteriori* e para qualquer distribuição. Entretanto, sob o ponto de vista prático, a escolha da distribuição é crucial para o bom desenvolvimento do algoritmo, influenciando na convergência para a distribuição marginal *a posteriori* (CHIB; GREENBERG, 1995).

Os passos a seguir especificam o algoritmo de Metropolis-Hasting:

- i) Inicialize o contador de iterações $k = 0$ e especifique valores iniciais $\theta^{(0)}$;
- ii) Gere um novo valor θ' da distribuição proposta $q(\cdot | \theta)$;
- iii) Calcule a probabilidade de aceitação $\alpha(\theta, \theta')$ e gere $u \sim U(0,1)$;

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

- iv) Se $u \leq \alpha$, então aceite o novo valor, θ' , e faça $\theta_1^{(k+1)} = \theta'$. Caso contrário, rejeite e faça $\theta^{k+1} = \theta$;
- v) Incremente o contador de k para $k + 1$ e volte ao passo **ii** até atingir a convergência.

À medida que o número de iterações aumenta, $k \rightarrow \infty$, a sequência se aproxima de sua condição de equilíbrio, atingindo a característica de estacionariedade com um comportamento assintótico (NOGUEIRA; SÁFADI; FERREIRA, 2004).

Segundo Gamerman (1997), um caso particular é o algoritmo Metropolis que considera apenas distribuições propostas do tipo simétricas, isto é, $q(\theta'|\theta) = q(\theta|\theta')$, para todos os valores de θ, θ' . Neste caso, a probabilidade de aceitação é reduzida a: $\alpha(\theta, \theta') = \min\left(1, \frac{P(\theta')}{P(\theta)}\right)$.

2.3.3 Avaliação da convergência

Os algoritmos Amostrador de Gibbs e Metropolis-Hastings são iterativos, ou seja, necessitam da constatação de convergência para que realmente se possa inferir sobre seus resultados como sendo valores das distribuições marginais dos parâmetros do modelo considerado. A monitoração dessa convergência é importante, portanto, evitando um número excessivo ou insuficiente de iterações no processo de amostragem. Entre os testes de avaliação da convergência desses algoritmos encontram-se os testes de diagnóstico de Gelman e Rubin (1992), Geweke (1992), Heidelberg e Welch (1983) e Raftery e Lewis (1992).

2.3.3.1 Monitoração informal da convergência

Segundo Gelfand e Smith (1990) e Nogueira, Sáfadi e Ferreira (2004) sugeriram, inicialmente, uma maneira informal de verificação da convergência, utilizando técnicas gráficas. O procedimento consiste em processar um número de iterações que se acredita serem suficientes (N). Feito isso, um histograma é gerado com estes resultados. Em seguida, mais Z iterações são necessárias e, sendo assim, cria-se um novo histograma ($N + Z$). Faz-se a comparação do gráfico anterior com o último. Verifica-se visualmente se há diferenças entre eles; caso não existam, significa que a sequência já estabilizou e o processo convergiu. Se o gráfico, após um período inicial, apresenta repetidamente o mesmo comportamento, então se pode concluir pela convergência da cadeia.

A utilização de apenas técnicas gráficas para a verificação da convergência pode não ser adequada, uma vez que a escala dos gráficos pode nos fornecer uma falsa impressão de igualdade. Para Gamerman (1997), outro ponto a ser discutido é sobre qual o número de iterações Z seria suficiente. Acredita-se que um número não muito grande seria necessário, dado que existe a suspeita de convergência, e esta pode estar próxima, não necessitando desta maneira de muitas iterações a mais.

2.3.3.2 Monitoração formal da convergência

O critério de Raftery e Lewis (1992) fornece a estimativa do número de iterações necessárias para se obter a convergência, bem como as estimativas do número de iterações que devem ser descartadas (*burn-in*) e a distância mínima de uma iteração à outra (*thin*) para se obter uma amostra independente. O Fator de Dependência (FD) é responsável pelo acréscimo multiplicativo ao número de iterações necessárias para se alcançar à convergência. Quando o fator de

dependência é menor que 5, a cadeia atingiu a convergência (GAMERMAN, 1997).

Outro critério é o de Gelman e Rubin (1992) que propõe a utilização de várias cadeias em paralelo, começando de valores iniciais distintos. Para cada parâmetro de interesse, compara-se a variabilidade dentro e entre as cadeias amostradas, sendo cada um desses parâmetros denotados por θ ; e uma vez que a estacionariedade tenha sido atingida, por exemplo, na i -ésima iteração, consideram-se as realizações $\theta_j, \theta_{j+h}, \dots, \theta_{j+Nh}$, $j \geq i$ como uma amostra aleatória da distribuição desejada. A convergência é monitorada através do fator R . Se o valor de R for distante de um, será necessário considerar mais iterações para obter uma melhor estimativa dos parâmetros. Quando R aproxima-se de um, a convergência ocorre e a amostra selecionada das iterações é independente e identicamente distribuída (iid). A lentidão na convergência pode estar relacionada com a alta correlação entre as componentes de θ .

O critério de Geweke (1992) é fundamentado em técnicas de análise espectral e fornece um diagnóstico para a ausência de convergência. Este critério propõe uma avaliação de convergência baseada no teste de igualdade de médias da primeira e da última parte da cadeia, geralmente, dos primeiros 10% e dos últimos 50%. A cadeia convergirá quando a maioria dos dados estiver entre os limites de uma distribuição normal padrão.

O critério de Heidelberger e Welch (1983) faz uso de testes estatísticos para avaliar a hipótese nula de estacionariedade da amostra gerada. Se a hipótese nula for rejeitada para um dado valor, o teste é repetido após serem descartados os 10% dos valores iniciais da amostra. Caso a hipótese seja rejeitada novamente, descarta-se mais 10% dos valores iniciais, e assim sucessivamente, até serem descartados 50% dos valores iniciais. Se a hipótese nula for rejeitada novamente, isso indica que é necessário um número maior de iterações. Caso

contrário, o número de iterações descartadas é indicado como o valor do *burn-in* (NOGUEIRA; SÁFADI; FERREIRA, 2004).

Após um estudo de avaliação dos critérios de convergência para os métodos MCMC, Nogueira, Sáfadi e Ferreira (2004) concluíram que o seguinte procedimento deveria ser utilizado para que se obtivesse uma avaliação mais precisa da convergência:

- i) aplicar Raftery e Lewis (1992) em uma amostra piloto e determinar o tamanho ideal da sequência;
- ii) monitorar a convergência das sequências nas proximidades do tamanho ideal, indicado pelo critério de Raftery e Lewis (1992), por meio dos critérios de Gelman e Rubin (1992) e Geweke (1992);
- iii) determinar o tamanho do “burn-in”, ou seja, do número de iterações desprezadas para se retirar o efeito dos valores iniciais do parâmetro, pelo critério de Heidelberger e Welch (1983).

2.3.4 Seleção de modelos

O fator de Bayes é um critério bayesiano para a comparação e a seleção de modelos, que envolve a procura de um modelo que seja, simultaneamente, o mais simples e que descreva da melhor forma possível os dados.

Desta forma, o fator de Bayes seleciona o melhor modelo comparando as probabilidades *a posteriori*, ou seja:

$$FB_{(M_1, M_2)} = \frac{P(Y | M_1)}{P(Y | M_2)},$$

em que $P(Y | M_1)$ e $P(Y | M_2)$ são as verossimilhanças marginais de cada modelo e definidas por:

$$P(Y | M_i) = \int L(Y | \theta, M_i) P(\theta | M_i) d\theta, \quad i = 1, 2 \quad (4)$$

em que: $L(Y|\theta, M_i)$ é a função de verossimilhança para o modelo M_i , $P(\theta|M_i)$ é a distribuição *a priori* assumida para o modelo M_i e θ é o parâmetro.

Em algumas situações, as quantidades $P(Y|M_i)$ e $P(Y|M_j)$ podem ser calculadas analiticamente. Mas, em geral, essas integrais são de difícil solução e os métodos MCMC são usados para obter soluções aproximadas (KASS; RAFTERY, 1995). Assim, gera-se uma amostra de tamanho $t(\theta^1, \theta^2, \dots, \theta^t)$, na qual a verossimilhança marginal dada em (4) pode ser

calculada por: $P(Y|M) = \frac{1}{T} \sum_{t=1}^T L(Y|\theta^t, M)P(\theta^t|M)$, em que T é o tamanho

da amostra final.

Na Tabela 2 é apresentada uma interpretação para o fator de Bayes dada por Jeffreys (1961), a qual contém categorias de evidências em relação aos modelos M_i e M_j e pode ser utilizada na escolha do melhor modelo.

Tabela 2 Interpretação do fator de Bayes, segundo Jeffreys (1961).

Valores de $FB_{(M_i, M_j)}$	Conclusão
$FB_{(M_i, M_j)} < 1$	Evidência a favor de M_j
$1 \leq FB_{(M_i, M_j)} < 3,2$	Evidência muito fraca a favor de M_i
$3,2 \leq FB_{(M_i, M_j)} < 10$	Evidência fraca a favor de M_i
$10 \leq FB_{(M_i, M_j)} < 100$	Evidência forte a favor de M_i
$FB_{(M_i, M_j)} \geq 100$	Evidência muito forte a favor de M_i

Segundo Silva et al. (2008), o fator de Bayes é eficiente quando não se utilizam *prioris* impróprias, como a *priori* de Jeffreys (JEFFREYS, 1961). Para casos assim, uma boa alternativa seria o uso do pseudofator de Bayes

(GELFAND, 1996). Este último é de aplicação mais prática, sendo calculado por meio de distribuições preditivas, as quais podem ser obtidas mediante valores gerados a cada iteração dos métodos MCMC.

3 MATERIAL E MÉTODOS

3.1 Modelos estudados

Foram utilizados os modelos de Gompertz e logístico com estrutura de erros independentes, os quais seguem abaixo:

- i. Modelo de Gompertz com erros independentes:

$$Y_i = a \exp\left\{-b\left[\exp(-kt_i)\right]\right\} + \varepsilon_i, \text{ com } i = 1, \dots, n$$

em que:

Y_i é o valor da variável dependente (exemplo: peso de frutos, comprimento de planta, teor de água em sementes);

a é a assíntota do modelo, indica o valor de estabilização da variável dependente em relação ao tempo t ;

b corresponde ao parâmetro de locação, sem interpretação biológica;

k determina a taxa de expansão do crescimento;

t refere-se ao tempo de análise da variável dependente (dias após a antese);

ε_i é o resíduo no tempo i , considerado independente, com distribuição

$$N(0, \sigma_\varepsilon^2);$$

n é o número de observações.

Para obtenção da distribuição *a posteriori* conjunta dos parâmetros do modelo de Gompertz, assumiram-se *prioris* uniformes para os parâmetros a , b e k . Desta forma tem-se:

$$p(a) p(b) p(k) \propto w$$

Supondo que os $\varepsilon_i = Y_i - a \exp\left\{-b\left[\exp(-kt_i)\right]\right\}$ são independentes e identicamente distribuídos (i.i.d.) e seguem uma distribuição normal, isto é:

$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, assumiu-se como *priori* para a precisão, τ , que é o inverso da

variância σ_e^2 , uma distribuição Gama com hiperparâmetros α e β , isto é:

$$p(\tau | \alpha, \beta) \propto \tau^{\alpha-1} \exp(-\beta\tau).$$

A verossimilhança para os erros do modelo é dada por:

$$L(Y | a, b, k, \tau, \alpha, \beta) \propto \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n [y_i - a \exp\{-b \exp(-kt)\}]^2\right\},$$

em que $Y = \{y_1, \dots, y_n\}$.

Por meio do teorema de Bayes, a distribuição *a posteriori* conjunta é dada por:

$$p(a, b, k, \tau | y, \alpha, \beta) \propto L(Y | a, b, k, \tau, \alpha, \beta) p(a) p(b) p(k) p(\tau | \alpha, \beta)$$

e substituindo as devidas expressões, fica:

$$p(a, b, k, \tau | y, \alpha, \beta) \propto \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n [y_i - a \exp\{-b \exp(-kt)\}]^2\right\} \tau^{\alpha-1} \exp(-\beta\tau)$$

ii. Modelo logístico com erros independentes:

$$Y_i = a [1 + b \exp(-kt_i)]^{-1} + \varepsilon_i, \text{ com } i = 1, \dots, n$$

em que os parâmetros (a, b, k) têm a mesma interpretação do modelo **i**.

Para obtenção da distribuição *a posteriori* conjunta dos parâmetros do modelo Logístico com erros independentes também assumiram-se *prioris* uniformes para os parâmetros a , b e k . Desta forma, tem-se:

$$p(a) p(b) p(k) \propto w$$

Sob a mesma suposição de que os erros $\varepsilon_i = Y_i - a [1 + b \exp(-kt_i)]^{-1}$ são i.i.d. assumiu-se como *priori* para a precisão, τ , uma distribuição Gama com hiperparâmetros α e β , isto é: $p(\tau | \alpha, \beta) \propto \tau^{\alpha-1} \exp(-\beta\tau)$.

A verossimilhança para os erros do modelo é dada por:

$$L(y | a, b, k, \tau, \alpha, \beta) \propto \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n [y_i - a [1 + b \exp(-kt)]^{-1}]^2\right\},$$

em que $Y = \{y_1, \dots, y_n\}$.

Por meio do teorema de Bayes, a distribuição *a posteriori* conjunta é dada por:

$$p(a, b, k, \tau | y, \alpha, \beta) \propto L(Y | a, b, k, \tau, \alpha, \beta) p(a) p(b) p(k) p(\tau | \alpha, \beta)$$

e substituindo as devidas expressões, fica:

$$p(a, b, k, \tau | y, \alpha, \beta) \propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \left[y_i - a \left[1 + b \exp(-kt) \right]^{-1} \right]^2 \right\} \tau^{\alpha-1} \exp(-\beta\tau)$$

Antes do ajuste dos modelos aos dados reais, foi feito um estudo com dados simulados, com o objetivo de avaliar o recurso computacional empregado e testar a metodologia utilizada.

3.2 Dados simulados

Foram simulados 77 valores de diâmetros longitudinal de tomates, correspondendo a uma coleta por dia, para ambos os modelos. Considerando o modelo de Gompertz, os valores assumidos como verdadeiros para o modelo foram: a (valor de estabilização do crescimento do diâmetro dos tomates) igual a 50; b (constante, sem interpretação biológica) igual a 3; k (taxa de crescimento da variável dependente) igual a 0,05 e ε_i seguindo uma distribuição normal com média zero e variância 2, ou seja, $\varepsilon_i \sim N(0, 2)$. Para o modelo logístico, foram utilizados os mesmos valores paramétricos para a e k e o parâmetro b , que não possui interpretação biológica, também ficou fixado no valor 3. A mesma distribuição para os erros, $\varepsilon_i \sim N(0, 2)$, foi utilizada neste modelo.

Os valores paramétricos foram escolhidos com base em valores encontrados na literatura para a variedade estudada. Foi escolhido para os parâmetros a e k , um valor próximo aos relatados por outros autores, a fim de

observar a convergência das cadeias. O parâmetro b manteve-se fixo num valor próximo ao encontrado na literatura especializada.

3.3 Dados reais

No presente trabalho foram utilizados dados de um experimento conduzido por Moura (1999). O delineamento experimental utilizado foi em blocos casualizados, com quatro repetições, sendo 11 tempos de amostragem para a variedade de tomate “Santa Clara”. Foram coletados dois frutos por repetição para as análises. Em intervalos de sete dias, a contar da antese, os frutos eram colhidos, imediatamente acondicionados em sacos de polietileno de baixa densidade e transportados para o laboratório. No laboratório foram feitas várias avaliações nos frutos, das quais, sem perda de generalidade, foi escolhida a variável diâmetro longitudinal dos frutos, para ser estudada neste trabalho.

Os dados foram analisados por Moura (1999) e Moura et al. (2004), por meio da estatística clássica, utilizando modelos não-lineares para o ajuste das curvas e estimação dos parâmetros. Não foi mencionado o modelo utilizado, bem como a interpretação de seus parâmetros; mas, o modelo utilizado tem a forma correspondente à parametrização do modelo logístico.

Na Figura 1 pode ser observado o comportamento do diâmetro longitudinal dos tomates nos 11 tempos de amostragem aos quais foram submetidos, iniciando aos 7 dias após a antese.

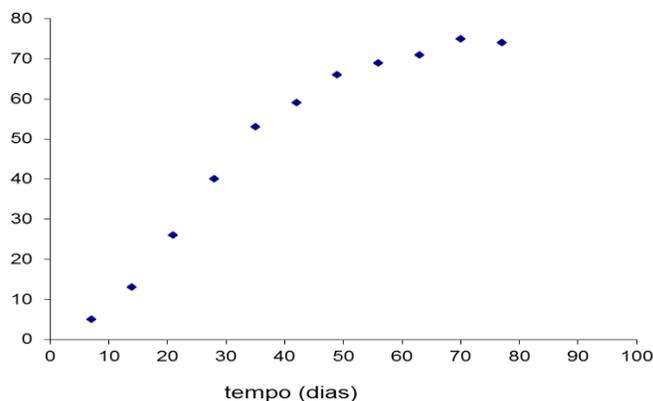


Figura 1 Variação do diâmetro longitudinal dos tomates durante 77 dias de desenvolvimento

3.4 Implementação dos algoritmos MCMC

Para a análise dos dados utilizou-se o software R (R DEVELOPMENT CORE TEAM, 2010), no qual foi implementado o algoritmo para o Amostrador de Gibbs e o Metropolis-Hastings. Na análise dos dados simulados foram consideradas 40.000 iterações, tanto para o modelo de Gompertz quanto para o modelo logístico.

Para ambos os modelos foram eliminadas (*burn-in*) 10% das iterações, ou seja, 4.000 delas foram descartadas a fim de evitar os efeitos dos valores iniciais dos parâmetros adotados. Além disso, considerou-se um espaçamento entre os pontos amostrados (*thin*) de tamanho 10. Desta forma, obteve-se uma cadeia final de 3.600 observações para cada parâmetro em cada amostra. A verificação final da convergência foi feita de acordo com o procedimento proposto por Nogueira, Sáfyadi e Ferreira (2004), utilizando o pacote BOA do software R.

Na análise dos dados reais tanto para o modelo de Gompertz quanto para o logístico foram consideradas 70.000 iterações, das quais se descartaram

as 1.000 primeiras e para assegurar a independência da amostra, utilizou-se um salto de tamanho 10 entre as observações da cadeia. Sendo assim, a cadeia final ficou composta por 6.900 observações. Para a verificação final da convergência das cadeias foram utilizados os mesmos critérios dos dados simulados.

Para a comparação entre os dois modelos - Gompertz e logístico - foi utilizado o fator de Bayes (FB), uma vez que este é um critério para a seleção de modelos o qual busca o modelo mais simples e que descreva os dados da melhor maneira, simultaneamente.

4 RESULTADOS E DISCUSSÃO

Primeiramente foi realizado o teste de Durbin-Watson (DW) para verificar a presença de autocorrelação residual de primeira ordem nos resíduos. O teste citado testa a hipótese nula de que os resíduos são não correlacionados contra a hipótese alternativa de que os resíduos seguem um processo autorregressivo de primeira ordem (DRAPER; SMITH, 1998). De acordo com os resultados obtidos no teste feito no software R por meio da função *dwt*, não foi detectada a autocorrelação residual em nenhum dos modelos propostos, pois o teste foi não significativo para ambos os modelos. Para o modelo de Gompertz, o valor da estatística de DW foi de 1,38 e para o modelo Logístico, 1,34. Ambos os valores situam-se acima do valor crítico superior ($d_{u,\alpha} = 1,32$) dado pela tabela “d” de Durbin-Watson, indicando ausência de autocorrelação residual ao nível de 5%. Sendo assim, toda a análise foi feita baseada no ajuste dos modelos com erros independentes.

4.1 Modelo de Gompertz

A distribuição *a posteriori* conjunta dos parâmetros do modelo é dada por:

$$p(a, b, k, \tau | y, \alpha, \beta) \propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n [y_i - a \exp[-b \exp(-kt)]]^2 \right\} \tau^{\alpha-1} \exp(-\beta\tau)$$

A partir desta distribuição foram obtidas as distribuições condicionais completas para cada parâmetro, a fim de obter aproximações das distribuições marginais dos parâmetros. Quando as condicionais completas possuem formas conhecidas, utiliza-se o Amostrador de Gibbs para determinar as marginais; já no caso das condicionais possuírem distribuições desconhecidas, o algoritmo de Metropolis-Hastings é utilizado.

4.1.1 Distribuições condicionais completas para o modelo de Gompertz

De posse da *posteriori* conjunta, a distribuição condicional completa a *posteriori* para o parâmetro a do modelo de Gompertz é dada por:

$$p(a | b, k, \tau, y, \alpha, \beta) = \frac{p(a, b, k, \tau | y, \alpha, \beta)}{\int p(a, b, k, \tau | y, \alpha, \beta) da} \quad (5)$$

A expressão (5) pode ser reescrita como:

$$p(a | b, k, \tau, y, \alpha, \beta) \propto \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^n [y_t - a \exp[-b \exp(-kt)]]^2 \right\} \quad (6)$$

Desenvolvendo o quadrado em (6), tem-se que $p(a | b, k, \tau, y, \alpha, \beta)$ é proporcional à seguinte expressão:

$$\propto \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^n [y_t^2 - 2y_t a \exp[-b \exp(-kt)] + a^2 (\exp[-b \exp(-kt)])^2] \right\} \quad (7)$$

Desenvolvendo o somatório em (7) e considerando que os termos que não dependem de a são termos constantes para este parâmetro, obtém-se a proporcionalidade abaixo:

$$\propto \exp \left\{ -\frac{\tau}{2} \left[\sum_{t=1}^n y_t^2 - 2a \sum_{t=1}^n y_t \exp[-b \exp(-kt)] + a^2 \sum_{t=1}^n (\exp[-b \exp(-kt)])^2 \right] \right\} \quad (8)$$

Colocando-se o termo que acompanha a^2 em evidência na expressão (8) e utilizando o completamento de quadrado, tem-se que a última expressão é proporcional à:

$$\propto \exp \left\{ -\frac{\tau}{2} \left(\sum_{t=1}^n (\exp[-b \exp(-kt)])^2 \right) \left[-2a \frac{\sum_{t=1}^n y_t \exp[-b \exp(-kt)]}{\sum_{t=1}^n (\exp[-b \exp(-kt)])^2} + a^2 \right] \right\} \quad (9)$$

A expressão (9) representa o núcleo de uma distribuição normal com média e variância dadas, respectivamente, por:

$$\mu = \frac{\sum_{t=1}^n y_t \exp[-b \exp(-kt)]}{\sum_{t=1}^n (\exp[-b \exp(-kt)])^2}; \sigma^2 = \frac{\sigma_e^2}{\sum_{t=1}^n (\exp[-b \exp(-kt)])^2}$$

A distribuição condicional completa *a posteriori* para o parâmetro b do modelo de Gompertz pode ser escrita, proporcionalmente, da seguinte forma:

$$p(b | a, k, \tau, y, \alpha, \beta) \propto \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^n [y_t - a \exp[-b \exp(-kt)]]^2 \right\} \quad (10)$$

Considerando o parâmetro k do modelo de Gompertz, sua distribuição condicional completa *a posteriori* é dada por:

$$p(k | a, b, \tau, y, \alpha, \beta) \propto \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^n [y_t - a \exp[-b \exp(-kt)]]^2 \right\} \quad (11)$$

Para os parâmetros b e k , as distribuições condicionais completas *a posteriori* não têm forma conhecida, quando comparadas a alguma distribuição padrão. Desta maneira, há necessidade de se utilizar o algoritmo de Metropolis-Hastings para realizar a análise dos parâmetros. A distribuição auxiliar utilizada para o parâmetro b , nos dados de simulação, foi uma distribuição normal com média 3 e variância 1. Para os dados reais, também foi considerada uma distribuição normal como candidata, porém com parâmetros 5 e 1, representando, respectivamente, a média e variância da distribuição. Já para o parâmetro k , a distribuição candidata foi uma gama, sendo que os valores dos parâmetros alfa e beta foram 5.000 e 100.000, respectivamente; e para os dados reais os parâmetros utilizados para alfa e beta foram 700 e 10.000, respectivamente. Tais valores foram escolhidos desta maneira no intuito de garantir que a média dessa distribuição ficasse próxima do valor paramétrico utilizado na ocasião.

A distribuição condicional completa *a posteriori* para o parâmetro τ , pode ser escrita de forma proporcional, da seguinte maneira:

$$p(\tau | a, b, k, y, \alpha, \beta) \propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n [y_i - a \exp[-b \exp(-kt)]]^2 \right\} \tau^{\alpha-1} \exp(-\beta\tau) \quad (12)$$

Reescrevendo os termos da expressão (12) de maneira adequada, isto é, agrupando-se alguns termos, a seguinte proporcionalidade é obtida:

$$\propto \tau^{\frac{n+2\alpha}{2}-1} \exp \left\{ -\tau \left[\beta + \frac{1}{2} \sum_{i=1}^n \{y_i - a \exp[-b \exp(-kt)]\}^2 \right] \right\} \quad (13)$$

A expressão (13) representa o núcleo de uma distribuição gama com os seguintes parâmetros:

$$\alpha^* = \frac{n+2\alpha}{2}; \beta^* = \beta + \frac{1}{2} \sum_{i=1}^n \{y_i - a \exp[-b \exp(-kt)]\}^2,$$

sendo α e β os hiperparâmetros da distribuição gama utilizada como *priori* para o parâmetro τ , a qual assumiu os valores de 25.000 e 50.000, respectivamente, no ajuste dos dados simulados; já no ajuste dos dados reais, estes valores foram, respectivamente, 500 e 1.000 para os hiperparâmetros alfa e beta.

4.2 Modelo logístico

A distribuição *a posteriori* conjunta dos parâmetros do modelo logístico é dada por:

$$p(a, b, k, \tau | y, \alpha, \beta) \propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \left[y_i - a [1 + \exp(-kt)]^{-1} \right]^2 \right\} \tau^{\alpha-1} \exp(-\beta\tau)$$

A partir desta distribuição foram obtidas as distribuições condicionais completas para cada parâmetro, a fim de obter aproximações das distribuições

marginais dos parâmetros, utilizando sempre os algoritmos Amostrador de Gibbs e Metropolis-Hastings, dependendo das necessidades.

4.2.1 Distribuições condicionais completas para o modelo logístico

De posse da *posteriori* conjunta, a distribuição condicional completa *a posteriori* para o parâmetro a do modelo logístico é dada por:

$$p(a | b, k, \tau, y, \alpha, \beta) = \frac{p(a, b, k, \tau | y, \alpha, \beta)}{\int p(a, b, k, \tau | y, \alpha, \beta) da} \quad (14)$$

A expressão (14) pode ser reescrita como:

$$p(a | b, k, \tau, y, \alpha, \beta) \propto \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \left[y_i - a [1 + b \exp(-kt)]^{-1} \right]^2 \right\} \quad (15)$$

Desenvolvendo o quadrado em (15), tem-se que $p(a | b, k, \tau, y, \alpha, \beta)$ é proporcional à seguinte expressão:

$$\propto \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \left[y_i^2 - 2y_i a [1 + b \exp(-kt)]^{-1} + a^2 \left([1 + b \exp(-kt)]^{-1} \right)^2 \right] \right\} \quad (16)$$

Desenvolvendo o somatório em (16) e considerando que os termos que não dependem de a são termos constantes para este parâmetro, obtém-se a proporcionalidade abaixo:

$$\propto \exp \left\{ -\frac{\tau}{2} \left[\sum_{i=1}^n y_i^2 - 2a \sum_{i=1}^n y_i [1 + b \exp(-kt)]^{-1} + a^2 \sum_{i=1}^n \left([1 + b \exp(-kt)]^{-1} \right)^2 \right] \right\} \quad (17)$$

Colocando-se o termo que acompanha a^2 em evidência na expressão (17) e utilizando o completamento de quadrado, tem-se que a última expressão é proporcional à:

$$\propto \exp \left\{ -\frac{\tau}{2} \left(\sum_{t=1}^n \left([1 + b \exp(-kt)]^{-1} \right)^2 \right) \left[-2a \frac{\sum_{t=1}^n y_t [1 + b \exp(-kt)]^{-1}}{\sum_{t=1}^n \left([1 + b \exp(-kt)]^{-1} \right)^2} + a^2 \right] \right\} \quad (18)$$

A expressão (18) representa o núcleo de uma distribuição normal com média e variância dadas, respectivamente, por:

$$\mu = \frac{\sum_{t=1}^n y_t [1 + b \exp(-kt)]^{-1}}{\sum_{t=1}^n \left([1 + b \exp(-kt)]^{-1} \right)^2}; \sigma^2 = \frac{\sigma_e^2}{\sum_{t=1}^n \left([1 + b \exp(-kt)]^{-1} \right)^2}$$

A distribuição condicional completa *a posteriori* para o parâmetro b do modelo Logístico pode ser escrita, proporcionalmente, da seguinte forma:

$$p(b | a, k, \tau, y, \alpha, \beta) \propto \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^n \left[y_t - a [1 + b \exp(-kt)]^{-1} \right]^2 \right\} \quad (19)$$

Considerando o parâmetro k do modelo Logístico, sua distribuição condicional completa *a posteriori* é dada por:

$$p(k | a, b, \tau, y, \alpha, \beta) \propto \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^n \left[y_t - a [1 + b \exp(-kt)]^{-1} \right]^2 \right\} \quad (20)$$

Assim como no modelo de Gompertz, para os parâmetros b e k , as distribuições condicionais completas *a posteriori* não têm forma conhecida, quando comparada a alguma distribuição padrão. Desta maneira, utiliza-se o algoritmo de Metropolis-Hastings na realização das análises dos parâmetros. Considerando a análise dos dados simulados, as distribuições candidatas, bem como seus valores paramétricos, foram os mesmos utilizados no modelo de Gompertz, isto é, a distribuição auxiliar para o parâmetro b foi uma distribuição normal com média 3 e variância 1; para o parâmetro k , a distribuição candidata foi uma gama com os valores dos parâmetros alfa e beta iguais a 5.000 e

100.000, respectivamente. Na análise dos dados reais, considerou-se como candidata para o parâmetro b uma distribuição normal com parâmetros 5 e 1 representando a média e variância da distribuição, respectivamente; e para o parâmetro k , considerou-se, assim como no modelo de Gompertz, uma distribuição gama de parâmetros 700 e 10.000 para alfa e beta, respectivamente.

A distribuição condicional completa *a posteriori* para o parâmetro τ pode ser escrita, de forma proporcional, da seguinte maneira:

$$p(\tau | a, b, k, y, \alpha, \beta) \propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n \left[y_i - a \left[1 + b \exp(-kt) \right]^{-1} \right]^2 \right\} \tau^{\alpha-1} \exp(-\beta\tau) \quad (21)$$

Reescrevendo os termos da expressão (21) de maneira adequada, isto é, agrupando-se alguns termos, a seguinte proporcionalidade é obtida:

$$\propto \tau^{\frac{n+2\alpha}{2}-1} \exp \left\{ -\tau \left[\beta + \frac{1}{2} \sum_{i=1}^n \left\{ y_i - a \left[1 + b \exp(-kt) \right]^{-1} \right\}^2 \right] \right\} \quad (22)$$

A expressão (22) representa o núcleo de uma distribuição gama com os seguintes parâmetros:

$$\alpha^* = \frac{n+2\alpha}{2}; \beta^* = \beta + \frac{1}{2} \sum_{i=1}^n \left\{ y_i - a \left[1 + b \exp(-kt) \right]^{-1} \right\}^2,$$

sendo α e β os hiperparâmetros da distribuição gama utilizada como *priori* para o parâmetro τ , a qual assumiu os valores de 25.000 e 50.000, para a análise dos dados simulados; e os valores de 500 e 1.000 quando analisaram-se os dados reais.

4.3 Resultados obtidos para os dados simulados

Nesta seção serão apresentados os resultados de simulação relativos aos dois modelos estudados.

4.3.1 Modelo de Gompertz (MG)

Na Tabela 3 são apresentados os valores paramétricos utilizados no estudo de simulação, assim como as estimativas da média e moda *a posteriori*, intervalo de máxima densidade *a posteriori* (HPD) para os parâmetros do modelo, considerando o cenário estudado.

Na Tabela 4 estão os valores obtidos pelos critérios de convergência propostos por Gelman e Rubin (1992) e Geweke (1992), os quais foram utilizados para verificar a convergência das cadeias. Além desses dois critérios, também foram utilizados o critério de Heidelberger e Welch (1993) e os gráficos gerados pelo método MCMC.

Os gráficos do traço da cadeia e da densidade *a posteriori* para os parâmetros do modelo referentes ao cenário de simulação estão apresentados nas Figuras de 2 a 5.

Tabela 3 Estimativas da média e moda *a posteriori*, intervalo de máxima densidade *a posteriori* (HPD) para os dados simulados segundo o modelo de Gompertz.

	Valores paramétricos			
	<i>a</i>	<i>b</i>	<i>k</i>	τ
	50,0	3,0	0,05	0,5
Média	49,43	2,95	0,05	0,49
Moda	50,43	3,05	0,05	0,49
HPD 95%	[38,82;57,77]	[1,01;4,95]	[0,04;0,05]	[0,45;0,51]

Tendo em vista os resultados apresentados na Tabela 3, pode-se perceber que as estimativas da média e moda *a posteriori* para o parâmetro *k* são exatamente iguais, indicando uma distribuição do tipo simétrica para este parâmetro. Já para os outros parâmetros, essa proximidade entre as estimativas da média e moda *a posteriori* não é muito acentuada, demonstrando para estes parâmetros distribuições do tipo assimétricas. Em situações como esta, o melhor

valor pontual para representar os dados é a moda *a posteriori*. Este resultado também pode ser observado nas Figuras de 2 a 5.

Ainda em relação à Tabela 3, observa-se que os intervalos HPD são todos estatisticamente significativos, pois nesse intervalo todos os valores têm a mesma probabilidade de ocorrência. Além disso, nota-se que a metodologia bayesiana foi eficaz na estimação dos parâmetros do modelo de Gompertz, uma vez que todos os valores paramétricos utilizados no cenário da simulação encontram-se dentro dos intervalos de máxima densidade *a posteriori* estimados. Sendo assim, pode-se considerar que a metodologia utilizada adapta-se às diversas situações encontradas em estudos de crescimento de frutos.

De acordo com os resultados da Tabela 4, pode-se dizer que todas as cadeias geradas pelo método MCMC conseguiram atingir a convergência, no cenário simulado para o modelo Gompertz. Os valores-p obtidos para o critério de Geweke (1992) indicam que as cadeias convergiram, uma vez que este critério aponta para a não convergência da cadeia quando o valor p for menor que um nível de significância pré-estabelecido na pesquisa (no presente trabalho fixou-se em 5%). Os valores obtidos para o fator de redução R do critério de Gelman e Rubin (1992) também apontam para a convergência das cadeias, pois, segundo o critério, a convergência será alcançada quando R está próximo de 1.

O critério de Heidelberger e Welch (1993) também foi utilizado na verificação da convergência das cadeias e indicou a estacionariedade das mesmas como resultado, ou seja, o número de iterações realizadas é suficiente.

Tabela 4 Critérios de Geweke (valor p) e Gelman & Rubin (R) utilizados na detecção da convergência das cadeias amostradas para os dados simulados do modelo de Gompertz.

	Valores paramétricos			
	a	b	k	τ
	50,0	3,0	0,05	0,5
Geweke	0,2776	0,2593	0,7589	0,9951
R	1,0003	0,9994	1,0007	0,9997

Os traços das cadeias para os parâmetros do modelo de Gompertz, e suas respectivas densidades *a posteriori* estão apresentados nas Figuras de 2 a 5.

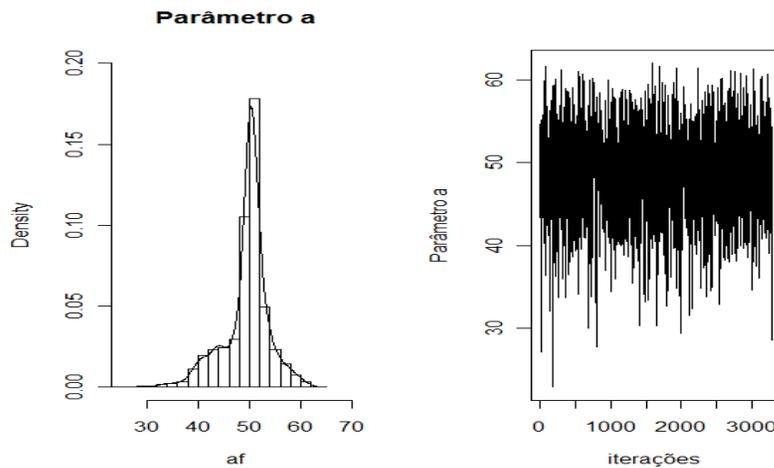


Figura 2 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro a do modelo de Gompertz, referente ao cenário 1

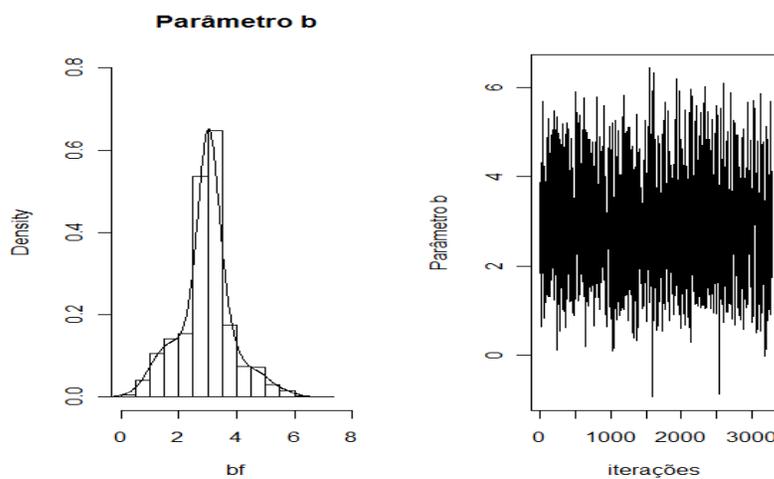


Figura 3 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro b do modelo de Gompertz, referente ao cenário 1

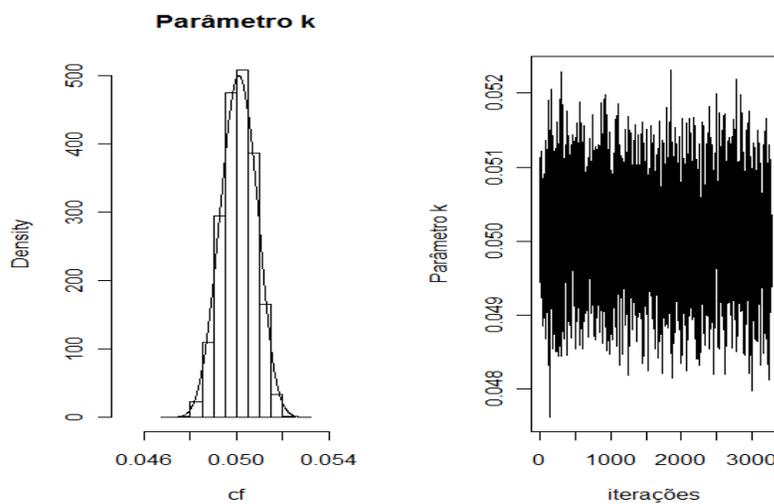


Figura 4 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro k do modelo de Gompertz, referente ao cenário 1

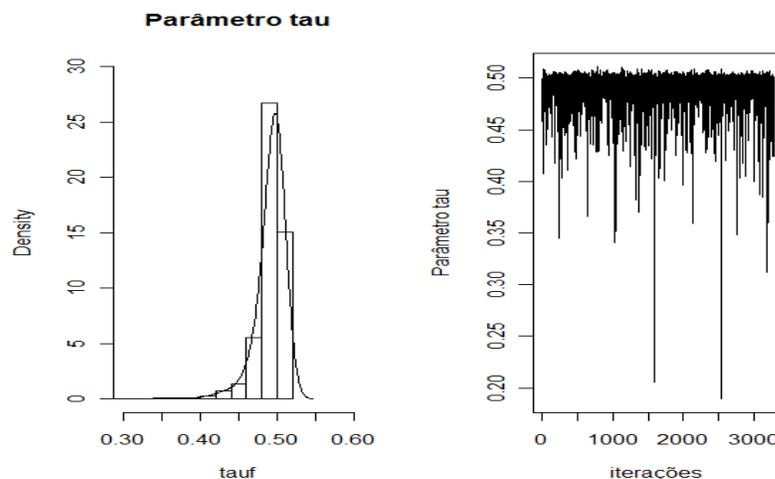


Figura 5 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro τ do modelo de Gompertz, referente ao cenário 1

Observa-se nas Figuras de 2 a 5 que as cadeias geradas para os parâmetros apresentam um comportamento que oscila em torno das médias estimadas, sem picos ao longo da cadeia. Nota-se também a simetria da densidade do parâmetro k (Figura 4), o que não ocorre com os outros parâmetros do modelo, tendo estes um comportamento assimétrico.

4.3.2 Modelo logístico (ML)

Na Tabela 5 os valores paramétricos utilizados nos estudo de simulação do modelo logístico são apresentados, bem como as estimativas da média e moda *a posteriori*, intervalo de máxima densidade *a posteriori* (HPD) para os parâmetros do modelo, considerando o mesmo cenário estudado para o modelo de Gompertz.

Os valores obtidos pelos critérios de convergência propostos por Gelman e Rubin (1992) e Geweke (1992) que foram utilizados para verificar a

convergência das cadeias, estão dispostos na Tabela 6. Também foi utilizado critério de Heidelberger e Welch (1993) e os gráficos gerados pelo método MCMC na verificação da convergência das cadeias.

Os gráficos do traço da cadeia e da densidade *a posteriori* para os parâmetros do modelo referentes ao cenário de simulação são apresentados nas Figuras de 6 a 9.

Tabela 5 Estimativas da média e moda *a posteriori*, intervalo de máxima densidade *a posteriori* (HPD) para os dados simulados segundo o modelo logístico.

	Valores paramétricos			
	<i>a</i>	<i>b</i>	<i>k</i>	τ
	50,0	3,0	0,05	0,5
Média	49,36	2,86	0,05	0,49
Moda	49,71	2,91	0,05	0,49
HPD 95%	[41,62;56,42]	[0,75;4,49]	[0,04;0,05]	[0,47;0,508]

De acordo com os resultados apresentados na Tabela 5, percebe-se que as estimativas da média e moda *a posteriori* para o parâmetro *k* são iguais, indicando uma distribuição do tipo simétrica para este parâmetro. O mesmo fato foi observado com este parâmetro no modelo de Gompertz, estudado anteriormente. Com relação aos outros parâmetros, a proximidade entre as estimativas da média e moda *a posteriori* é menor, porém, quando comparadas ao modelo de Gompertz, nota-se que estas estimativas estão menos distantes entre si, aproximando-se mais de uma distribuição simétrica que os mesmos parâmetros do outro modelo. Este resultado também pode ser observado nas Figuras de 6 a 9.

Nota-se ainda na Tabela 5 que os intervalos de credibilidade estimados são todos estatisticamente significativos. Percebe-se também que a metodologia bayesiana foi mais uma vez eficaz na estimação dos parâmetros do modelo logístico, estando todos os valores paramétricos utilizados no cenário da

simulação dentro dos intervalos de máxima densidade *a posteriori* estimados. Reis et al. (2011) encontraram intervalos de credibilidade contendo os valores reais de cada parâmetro, para quase todos os cenários estudados, num trabalho de simulação feito para comparar modelos aplicados ao equilíbrio de Hardy-Weinberg utilizando o coeficiente de desequilíbrio.

Observa-se na Tabela 6 que todas as cadeias geradas pelo método MCMC atingiram a convergência para o cenário simulado no modelo logístico.

De acordo com o critério de Geweke (1992), as cadeias relativas a todos os parâmetros convergiram, uma vez que os valores-p encontrados foram maiores que 5% (valor mínimo permitido). Para o parâmetro a do cenário 1, o valor p encontrado foi de 9,03%, sendo este resultado ainda assim superior ao mínimo estabelecido e suficiente para indicar a convergência desta cadeia.

Os valores obtidos para o fator de redução R do critério de Gelman e Rubin (1992) também apontam para a convergência das cadeias, estando todos eles muito próximo de 1. Por fim, mas não menos importante, o critério de Heidelberger e Welch (1993) também indicou a convergências das cadeias, por meio da estacionariedade das mesmas.

Tabela 6 Critérios de Geweke (valor p) e Gelman & Rubin (R) utilizados na detecção da convergência das cadeias amostradas para os dados simulados do modelo logístico.

	Valores paramétricos			
	a	b	k	τ
	50,0	3,0	0,05	0,5
Geweke	0,0903	0,8776	0,5763	0,7691
R	0,9996	0,9994	0,9994	0,9994

Os traços das cadeias para os parâmetros do modelo logístico, e suas respectivas densidades *a posteriori* estão apresentados nas Figuras de 6 a 9.

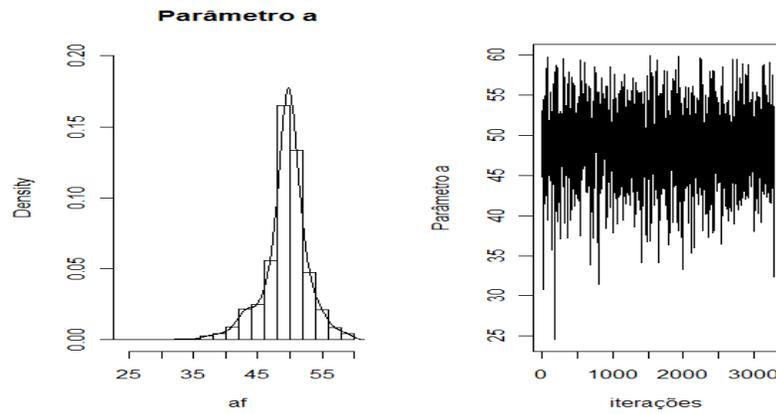


Figura 6 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro a do modelo logístico, referente ao cenário 1

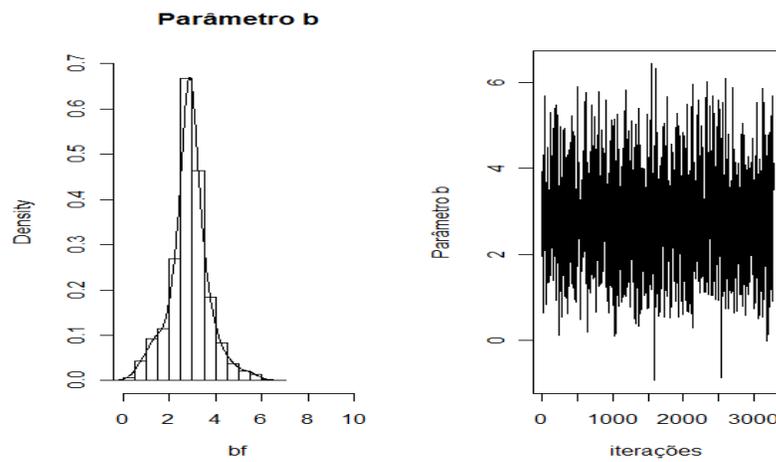


Figura 7 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro b do modelo logístico, referente ao cenário 1

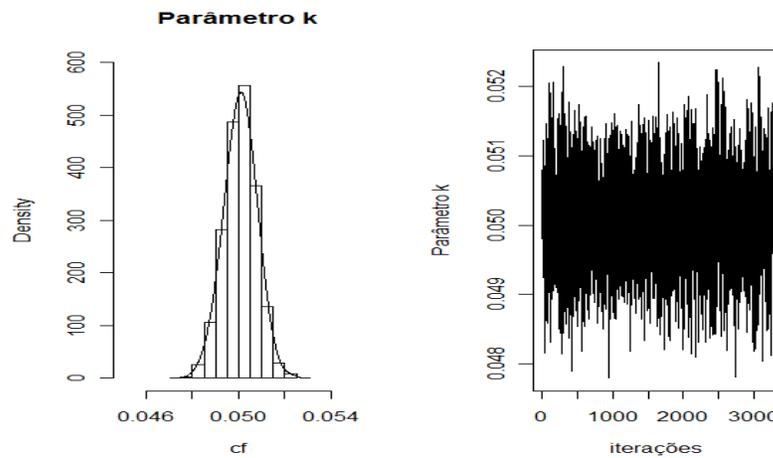


Figura 8 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro k do modelo logístico, referente ao cenário 1

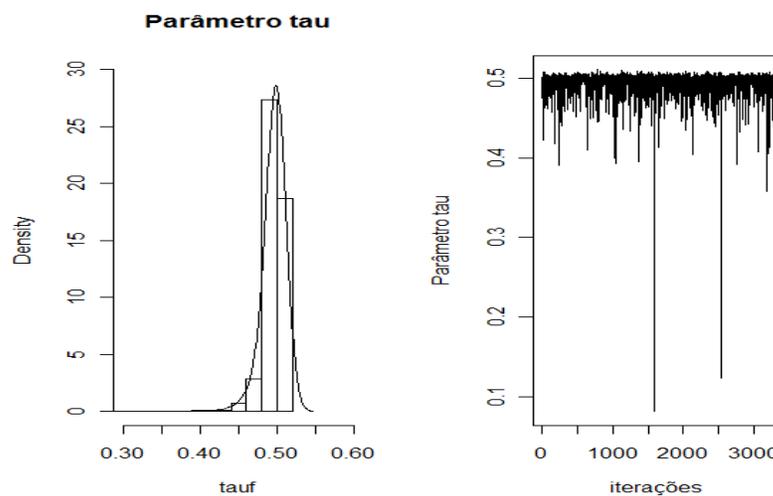


Figura 9 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro τ do modelo logístico, referente ao cenário 1

Na Figura 8, percebe-se a simetria da cadeia, fato já indicado e relatado anteriormente, quando foi feita a análise das médias e modas *a posteriori*. Nas Figuras 6, 7 e 9 pode ser observado que as cadeias variaram pouco em torno de seus respectivos valores paramétricos, e também a assimetria destes parâmetros.

4.4 Resultados obtidos para os dados reais

Nesta seção serão apresentados os resultados obtidos na análise dos dados reais referentes ao diâmetro longitudinal dos frutos de tomateiro da variedade “Santa Clara”, considerando os dois modelos estudados.

4.4.1 Modelo de Gompertz

Na Tabela 7, são apresentadas as estimativas da média e moda *a posteriori*, intervalo de máxima densidade *a posteriori* (HPD) para os parâmetros do modelo de Gompertz.

Na Tabela 8 estão os valores obtidos pelos critérios de convergência propostos por Gelman e Rubin (1992) e Geweke (1992), os quais foram utilizados para verificar a convergência das cadeias. Também fizeram parte da verificação da convergência das cadeias, o critério de Heidelberger e Welch (1993) e os gráficos gerados pelo método MCMC.

Os gráficos do traço da cadeia e da densidade *a posteriori* para os parâmetros do modelo estão apresentados nas Figuras de 10 a 13.

Tabela 7 Estimativas da média e moda *a posteriori*, intervalo de máxima densidade *a posteriori* (HPD) para os dados reais segundo o modelo de Gompertz.

Parâmetros	Média	Moda	Intervalo HPD (95%)	
			LI	LS
<i>a</i>	76,0075	75,9907	74,3342	77,7084
<i>b</i>	4,6612	4,6470	4,2075	5,1551
<i>k</i>	0,0707	0,0705	0,0666	0,0749
τ	0,5019	0,5008	0,4589	0,5466

Na análise dos dados reais, pode-se observar, na Tabela 7, que as estimativas da média e moda *a posteriori* para o ajuste do modelo de Gompertz, apresentam-se bastante próximas entre si, evidenciando distribuições do tipo simétricas para todos os parâmetros envolvidos no modelo. Nota-se que a diferença entre os valores das estimativas se dá a partir da segunda casa decimal, mostrando o quão próximos estão estes valores. Os valores da média e moda *a posteriori*, para o parâmetro a , se encontram em torno de 76 mm, isto significa que os frutos de tomateiro tendem a estabilizar o crescimento próximos de 76 mm. Estas estimativas estão de acordo com os valores encontrados por Moura et al. (2004), no ajuste do diâmetro de tomates “Santa Clara”, no qual o valor da estimativa para o parâmetro a , ou seja, o diâmetro máximo atingido pela fruto do tomateiro, foi de 72 mm.

No estudo de Terra, Muniz e Savian (2010), analisando o diâmetro dos frutos da tamareira-anã, pode-se observar que o diâmetro estimado na maturidade do fruto foi de aproximadamente 6,9771 mm aos 103 dias após a antese, tanto para o modelo de Gompertz quanto para o Logístico.

Segundo Bino et al. (1998), após a antese das flores, deve-se conhecer o processo de maturação tanto dos frutos quanto das sementes no intuito de se obter um material de melhor qualidade.

Observando o resultado estimado da média e moda *a posteriori* para o parâmetro k na Tabela 7 e comparando-o com a Figura 12, percebe-se claramente que quanto mais próximos estão os valores pontuais da média e moda, mais simétrica é a distribuição do parâmetro. O valor médio estimado para este parâmetro e apresentado na tabela acima é de 0,0707, indicando a taxa de expansão de crescimento do fruto até atingir a maturidade. Terra, Muniz e Savian (2010) encontraram para este parâmetro o valor de 0,0287 no ajuste do modelo de Gompertz aos dados de diâmetro dos frutos de tamareira-anã. De

acordo com Carvalho e Nakagawa (2000), as sementes, de modo geral, crescem rapidamente em tamanho, atingindo o máximo desenvolvimento em curto período de tempo, antes mesmo de completar o processo de maturação dos frutos.

Ainda na Tabela 7, pode-se observar que, de maneira geral, a metodologia utilizada foi eficiente, uma vez que os intervalos de máxima densidade *a posteriori* (HPD), para todos os parâmetros, foram significativos. Além disso, percebe-se também o quanto as estimativas dispersaram pouco, pois os intervalos obtidos estão bem concentrados, ou seja, a amplitude dos intervalos é pequena. Martins Filho et al. (2008) também obtiveram resultados satisfatórios utilizando a metodologia bayesiana na comparação de curvas de crescimento de cultivares de feijoeiro.

Os resultados relacionados à convergência das cadeias estão apresentados na Tabela 8.

Pode-se observar, por meio dos resultados da Tabela 8, que todas as cadeias geradas pelo método MCMC atingiram a convergência no ajuste do modelo de Gompertz aos dados reais. Os resultados obtidos para o critério de Geweke (1992) apontam para a convergência das cadeias, encontrando valores-p maiores que 5%, valor mais comumente utilizado. Os valores alcançados para o fator de redução R do critério de Gelman e Rubin (1992) também indicam que as cadeias convergiram, uma vez que este critério trabalha com resultados de R satisfatórios próximos do valor 1.

A convergência também foi verificada e obteve resultados satisfatórios por meio do critério de Heidelberger e Welch (1993) o qual indicou a estacionariedade das mesmas, isto é, o número de iterações realizadas no estudo é suficiente.

Tabela 8 Critérios de Geweke (valor p) e Gelman & Rubin (R) utilizados na detecção da convergência das cadeias amostradas para os dados reais do modelo de Gompertz.

Parâmetros	Geweke	R
a	0,4786	1,0032
b	0,4471	1,0008
k	0,3038	1,0013
τ	0,7994	1,0014

As Figuras 10, 11, 12 e 13 mostram as distribuições marginais dos parâmetros do modelo de crescimento de Gompertz. De maneira geral, estas figuras retratam de forma prática a utilização da metodologia Bayesiana, a qual está fundamentada em distribuições de probabilidade para estimar os parâmetros de interesse.

Através das Figuras supracitadas, os resultados relacionados à convergência das cadeias podem ser confirmados por meio da visualização gráfica apresentada.

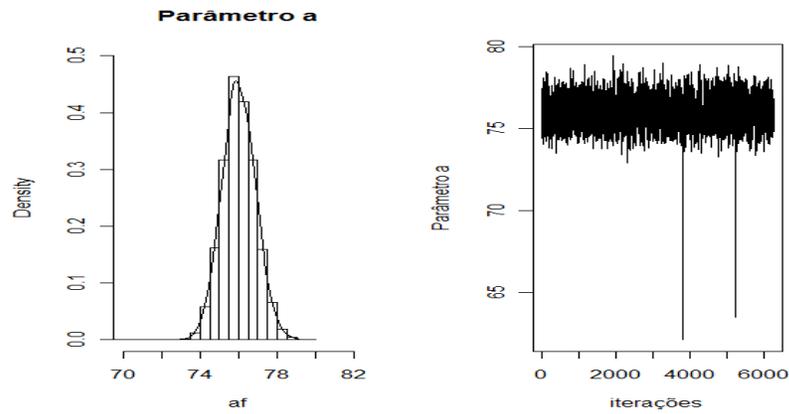


Figura 10 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro a do modelo de Gompertz referente aos dados reais

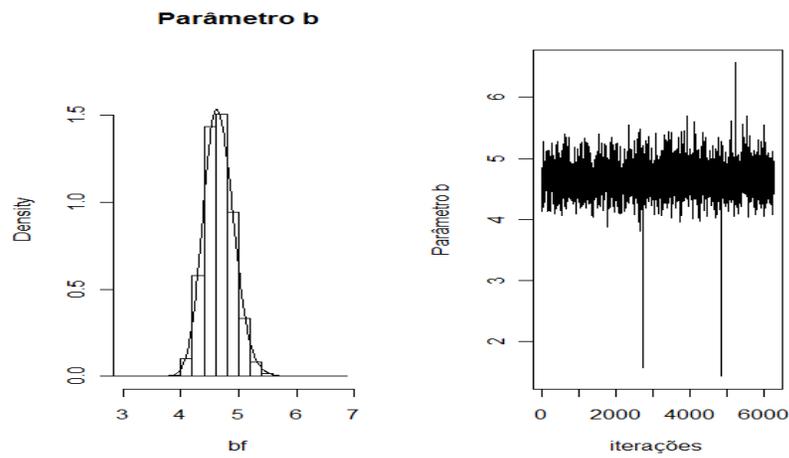


Figura 11 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro b do modelo de Gompertz referente aos dados reais

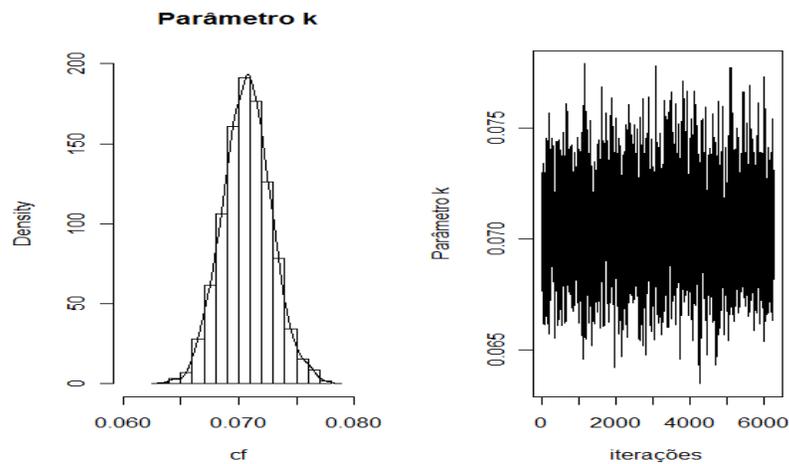


Figura 12 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro k do modelo de Gompertz referente aos dados reais

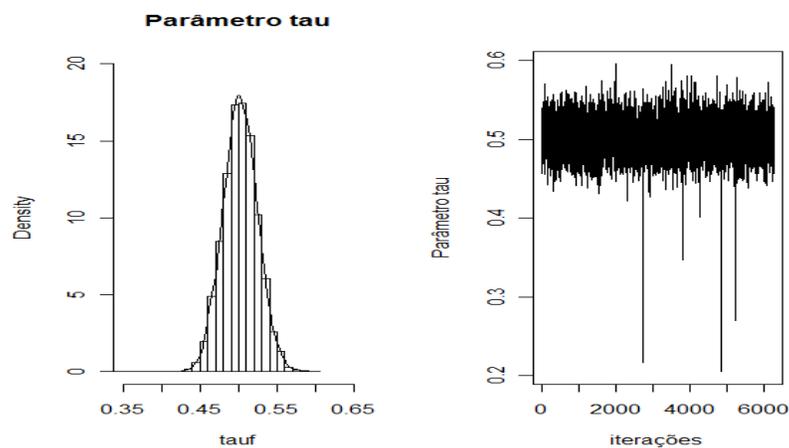


Figura 13 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro τ do modelo de Gompertz referente aos dados reais

Observa-se, pela análise das Figuras 10 a 13, que as cadeias geradas para todos os parâmetros do modelo de Gompertz oscilaram em torno das médias estimadas, não apresentando valores discrepantes ao longo das cadeias. Desta

maneira, em estudos futuros, podem-se propor distribuições *a priori* do tipo simétrica para os parâmetros deste modelo.

4.4.2 Modelo logístico

Nesta seção será apresentada a Tabela 9, na qual se encontram os valores das estimativas da média e moda *a posteriori*, intervalo de máxima densidade *a posteriori* (HPD) para os parâmetros do modelo logístico.

Serão apresentados, na Tabela 10, os valores obtidos pelos critérios de convergência propostos por Gelman e Rubin (1992) e Geweke (1992), utilizados para verificar a convergência das cadeias. O critério de Heidelberger e Welch (1993) e os gráficos gerados pelo método MCMC também contribuiram como critérios na verificação da convergência das cadeias.

A monitoração informal da convergência, obtida por meio da análise dos gráficos do traço da cadeia e da densidade *a posteriori* para os parâmetros do modelo, está apresentada nas Figuras de 14 a 17.

Tabela 9 Estimativas da média e moda *a posteriori*, intervalo de máxima densidade *a posteriori* (HPD) para os dados reais segundo o modelo logístico.

Parâmetros	Média	Moda	Intervalo HPD (95%)	
			LI	LS
<i>a</i>	77,0934	77,1723	75,1001	79,2287
<i>b</i>	7,5739	7,6656	6,5029	8,8094
<i>k</i>	0,0736	0,0741	0,0693	0,0781
τ	0,4608	0,4631	0,4176	0,5042

De acordo com os resultados apresentados na Tabela 9, percebe-se que as estimativas da média e moda *a posteriori* para o ajuste do modelo de logístico

mostram-se bem próximas entre si, sugerindo distribuições do tipo simétricas para os parâmetros envolvidos no modelo. Com base nos valores estimados para o parâmetro a pode-se dizer que o modelo logístico estimou o diâmetro longitudinal dos tomates em valores condizentes com a realidade da cultivar estudada (em torno de 77 mm). No trabalho de Terra, Muniz e Savian (2010), o parâmetro a foi estimado em 6,9000 mm para o modelo logístico, resultado muito parecido aos valores estimados do modelo de Gompertz, no estudo do crescimento do diâmetro dos frutos de tamareira-anã.

O valor médio para o parâmetro k foi estimado em 0,0736, ou seja, essa é a taxa de expansão do crescimento dos frutos de tomateiro. Já no estudo de Moura et al. (2004), que utilizou a metodologia clássica para ajustar os modelos, esse parâmetro foi estimado em 0,1142. Isto significa que no presente estudo, utilizando a metodologia bayesiana, os frutos demoraram mais tempo a atingir a maturidade, isto é, seu crescimento foi mais lento. Terra, Muniz e Savian (2010) estimaram o parâmetro k em 0,0372, valor este ligeiramente superior ao encontrado pelo modelo de Gompertz no ajuste dos mesmos dados, referentes ao crescimento de frutos de tamareira-anã. No estudo de curvas de crescimento de duas cultivares de feijoeiro, Martins Filho et al. (2008) verificaram que a cultivar “carioca” apresentou maior precocidade que a cultivar “neguinho”, pois os valores encontrados para o parâmetro k na primeira cultivar foram maiores que os valores da segunda cultivar, no ajuste do modelo logístico aos dados de altura das plantas.

No estudo realizado por Pereira et al. (2000) sobre o crescimento da bananeira “Prata-Anã”, o modelo logístico foi ajustado aos dados de altura das plantas e observou-se que as bananeiras tinham cerca de 0,16 m de altura à época do plantio. Por volta do sexto mês, as plantas atingiram a metade da altura máxima, e aos nove meses, se encontravam com 2,65 m, tendendo à estabilidade

a partir daí. A taxa de crescimento atingiu seu valor máximo por volta do sexto mês após o plantio, com o valor de 0,48 m/mês.

Da mesma forma que no modelo de Gompertz, pode-se observar que a metodologia utilizada foi eficiente, uma vez que todos os intervalos de máxima densidade *a posteriori* (HPD) foram significativos. Se forem comparados, porém, aos intervalos de máxima densidade *a posteriori* do modelo estudado anteriormente, estes aqui apresentam uma maior variabilidade nas estimativas encontradas para os parâmetros, pois seus intervalos possuem maior amplitude, ou seja, estão mais dispersos que os intervalos encontrados para os parâmetros do modelo de Gompertz. A maior dispersão pode ser notada no intervalo encontrado para o parâmetro τ do modelo logístico.

Pelos valores encontrados nas estimativas dos parâmetros, uma informação pontual pode ser obtida tanto pelo valor da média, quanto pelo valor da moda *a posteriori*, pois ambos os valores estão muito próximos entre si.

Na Tabela 10 são apresentados os resultados relacionados à convergência das cadeias do modelo logístico. Observa-se que todas as cadeias geradas pelo método MCMC atingiram a convergência no ajuste do modelo logístico aos dados reais. Mais uma vez, os valores obtidos para o critério de Geweke (1992) apontam para a convergência das cadeias, encontrando valores-p maiores que 5%, valor este pré-fixado e abaixo do qual se considera que as cadeias dos parâmetros não convergiram. Os valores alcançados para o fator de redução R do critério de Gelman e Rubin (1992) também indicam que as cadeias convergiram, uma vez que este critério trabalha com resultados de R satisfatórios próximos do valor 1.

Tabela 10 Critérios de Geweke (valor p) e Gelman & Rubin (R) utilizados na detecção da convergência das cadeias amostradas para os dados reais do modelo logístico.

Parâmetros	Geweke	R
a	0,1919	1,0050
b	0,1509	1,0107
k	0,7928	1,0046
τ	0,6312	1,0029

A convergência também foi verificada e obteve resultados satisfatórios por meio do critério de Heidelberger e Welch (1993) que indicou a estacionariedade das mesmas, isto é, o número de iterações realizadas no estudo é suficiente.

Nas Figuras de 14 a 17 são mostradas as distribuições marginais *a posteriori* dos parâmetros do modelo de crescimento logístico. A visualização das figuras a seguir possibilita a confirmação dos resultados relacionados à convergência das cadeias apresentados na Tabela 10.

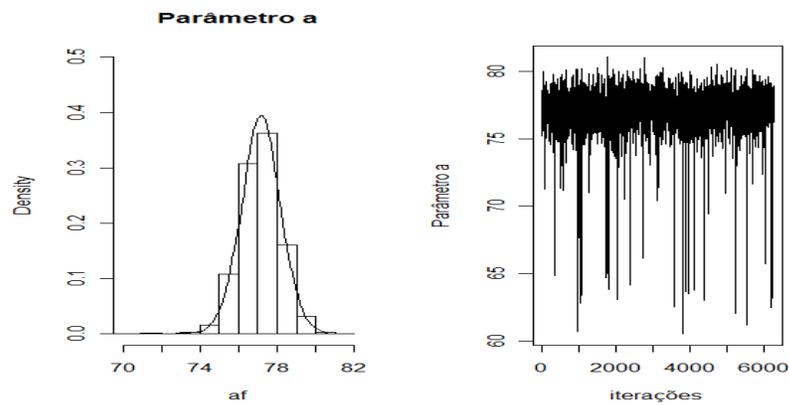


Figura 14 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro a do modelo logístico referente aos dados reais

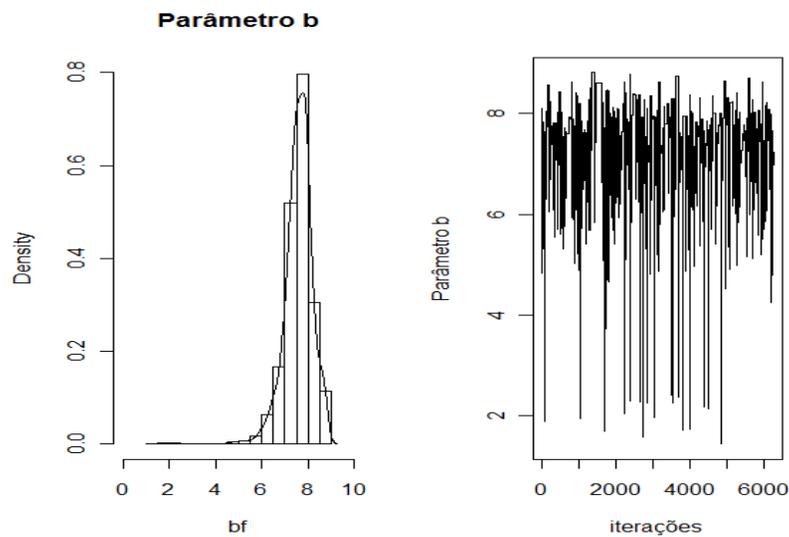


Figura 15 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro b do modelo logístico referente aos dados reais

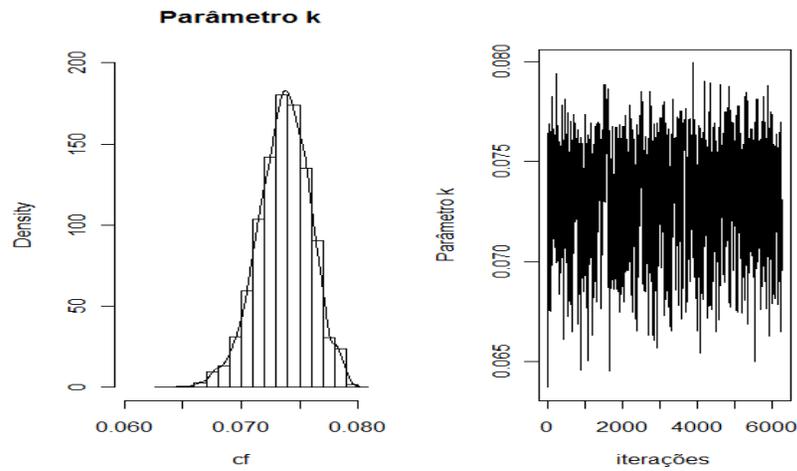


Figura 16 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro k do modelo logístico referente aos dados reais

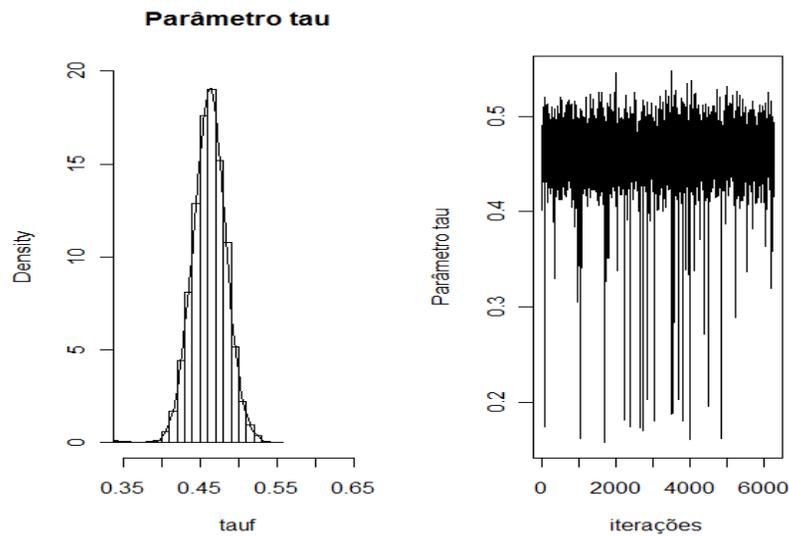


Figura 17 Densidade *a posteriori* e traço da cadeia gerada pelo método MCMC para o parâmetro τ do modelo logístico referente aos dados reais

Observando as Figuras 14 a 17, pode-se perceber que as cadeias geradas para todos os parâmetros do modelo logístico oscilaram mais em torno das médias estimadas se comparadas às cadeias geradas para os mesmos parâmetros do modelo de Gompertz. Este fato confirma o que já foi observado e relatado na discussão sobre a amplitude dos intervalos de máxima densidade *a posteriori*, ou seja, os valores encontrados para o modelo logístico estão um pouco mais dispersos em relação aos valores pontuais da média e moda *a posteriori*.

Ainda nas Figuras, percebe-se que, com exceção do parâmetro *b*, todos os outros apresentam uma distribuição com tendência à simetria, fato já observado pela proximidade das estimativas. O mesmo não ocorre com o parâmetro *b* para o qual se observou uma distribuição assimétrica à esquerda.

4.5 Comparação dos modelos estudados

A escolha do melhor modelo foi feita com base no critério denominado Fator de Bayes, o qual já foi descrito na seção 2.3.4. Ao comparar os modelos utilizando o referido critério, o objetivo é verificar qual dos modelos se ajusta melhor aos dados de crescimento de tomates e deve, então, ser utilizado para prever o fenômeno.

Na Tabela 11 são apresentados os resultados das comparações dos dados reais. Nesta comparação, o modelo de Gompertz foi colocado no numerador e o modelo logístico, no denominador.

Para o estudo de simulação, não será utilizado o fator de Bayes como critério de comparação entre os modelos, uma vez que foram gerados valores para os diferentes modelos. Já para os dados reais, a comparação será feita por meio do referido critério, pois será verificado qual modelo melhor se ajusta àqueles dados de crescimento de diâmetro de tomates.

Tabela 11 Fator de Bayes (FB) para a comparação dos modelos de Gompertz e logístico no ajuste dos dados reais.

Dados reais	
Modelos	FB
MG/ML	732

De acordo com o resultado apresentado na Tabela 11 e seguindo a interpretação sugerida por Jeffreys (1961) já descrita anteriormente, percebe-se na comparação dos ajustes dos modelos aplicados aos dados reais, que o modelo logístico é mais plausível ou verossímil que o modelo de Gompertz, ou seja, ele se adequa melhor na descrição de curvas de crescimento de dados como estes, de frutos de tomateiro. Num estudo de métodos de comparação de modelos não-lineares em bananeiras, Maia et al. (2009) encontraram os melhores valores para os avaliadores da qualidade do ajuste para o modelo logístico. Um dos avaliadores adotados para medir a qualidade do ajuste foi o critério de informação bayesiano, o qual também apontou a superioridade do modelo logístico.

A função logística também foi selecionada para explicar o crescimento de girassol, num estudo realizado por Amabile, Guimarães e Farias Neto (2003), no qual o modelo logístico foi comparado ao logístico generalizado e a função Weibull. Num estudo sobre crescimento de árvores, Pires e Calegario (2007) utilizaram tanto modelos lineares quanto não-lineares e concluíram que entre os modelos não-lineares, o que obteve melhores resultados foi o modelo logístico. Resultado semelhante foi obtido por Terra, Muniz e Savian (2010) no qual o modelo logístico apresentou melhores avaliadores da qualidade do ajuste quando comparado ao modelo de Gompertz, no estudo de crescimento de frutos da tamareira-anã.

5 CONCLUSÕES

De acordo com os resultados obtidos neste estudo, pôde-se concluir que:

Tanto o modelo de Gompertz quanto o modelo logístico apresentaram valores para as estimativas dos parâmetros coerentes com a realidade dos frutos de tomateiro do tipo “Santa Clara”. Além disso, a metodologia bayesiana, utilizada na estimação dos parâmetros, mostrou-se eficiente no estudo de crescimento de tomates, o que foi comprovado por meio dos processos de simulação avaliados no presente trabalho.

Na comparação dos modelos, apesar de ambos apresentarem bom ajuste aos dados, o modelo logístico foi considerado como o melhor e mais adequado para explicar o comportamento destes dados.

REFERÊNCIAS

- AMABILE, R. F.; GUIMARÃES, D. P.; FARIAS NETO, A. L. de. Análise de crescimento de girassol em Latossolo com diferentes níveis de saturação por bases no Cerrado. **Pesquisa Agropecuária Brasileira**, Brasília, v. 38, n. 2, p. 219-224, fev. 2003.
- ANDRADE, P. C. de R. **Comparações múltiplas bayesianas em modelos normais homocedásticos e heterocedásticos**. 2008. 96 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2008.
- ANUÁRIO da agricultura brasileira. São Paulo: FNP Consultoria e Comércio, 2007. 495 p.
- BINO, R. J. et al. Pesquisa para o aprimoramento de tecnologia de sementes. **Scientia Agricola**, Piracicaba, v. 55, p. 19-26, 1998. Número especial.
- 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.
- CARVALHO, N. M.; NAKAGAWA, J. **Sementes: ciência, tecnologia e produção**. 4. ed. Jaboticabal: FUNEP, 2000. 588 p.
- CHIB, S.; GREENBERG, E. Understanding the metropolis-hasting algorithm. **The American Statistician**, Salt Lake, v. 49, n. 4, p. 327-345, Nov. 1995.
- CORRÊA, P. C.; ARAÚJO, E. F.; AFONSO JÚNIOR, P. C. Determinação dos parâmetros de secagem em camada delgada de sementes de milho doce (*Zea mays* L.). **Revista Brasileira de Milho e Sorgo**, Sete Lagoas, v. 2, n. 2, p. 110-119, abr./jun. 2003.

CORSATO, C. E.; SCARPARE FILHO, J. A.; VERDIAL, M. F. Fenologia do caquizeiro Rama Forte em clima tropical. **Bragantia**, Campinas, v. 64, n. 3, p. 223-229, 2005.

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

DUARTE, F. A. M. **Estudo da curva de crescimento de animais da raça “Nelore” (*Bos taurus indicus*) através de cinco modelos estocásticos**. 1975. 284 f. Tese (Livre Docência em Genética e Matemática Aplicada à Biologia) - Universidade de São Paulo, Ribeirão Preto, 1975.

FIDELES FILHO, J. et al. Monitoramento de área foliar e biomassa do feijoeiro usando índice de vegetação por diferença normalizada. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 13, n. 2, p. 212-218, 2005.

FITZHUGH JUNIOR, H. A. Analysis of growth curves and strategies for altering their shapes. **Journal of Animal Science**, Champaign, v. 42, n. 4, p. 1036-1051, Apr. 1976.

GAMERMAN, D. **Markov chain Monte Carlo: stochastic simulation for bayesian inference**. London: Chapman & Hall, 1997. 245 p.

GELFAND, A. E. Model determination using sampling based methods. In: GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. (Ed.). **Markov chain Monte Carlo in practice**. London: Chapman & Hall, 1996. p. 145-162.

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. 457-511, May 1992.

GEWEKE, J. **Evaluating the accuracy of sampling-based approaches to calculating posterior moments**. New York: Oxford University, 1992. 631 p.

GOMES, F. R. C. et al. Influência do manejo do solo e da intensidade de raleio de frutas, no crescimento e qualidade de pêssegos, cvs. Cerrito e Chimarrita. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 27, n. 1, p. 60-63, jan./fev. 2005.

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, jul./set. 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. 381-388, mar./abr. 2004.

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.

JEFFREYS, H. **Theory of probability**. Oxford: Clarendon, 1961. 325 p.

KASS, R. E.; RAFTERY, A. E. Bayes factor. **Journal of the American Statistical Association**, Alexandria, v. 90, n. 430, p. 773-795, June 1995.

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.

LINHARES, R. R.; SILVA, H. D.; LUCAS, M. B. **Avaliação da qualidade do ajuste da curva de degradação de Fenitrothion em frutos de berinjela, por meio de modelos não-lineares e linearizáveis**. Uberlândia: UFU, 2005.

Disponível em:

<http://www.propp.ufu.br/revistaeletronica/edicao2005_2/a/avaliacao.pdf>.

Acesso em: 8 ago. 2008.

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.

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

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, nov./dez. 2008.

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

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

MELO, P. C. T. de. **Produção de sementes de tomate**. Piracicaba: USP/ESALQ, 2007. 12 p.

MENDES, P. N. **Curvas de crescimento difásicas de fêmeas Hereford com erros auto-regressivos e heterogeneidade de variâncias**. 2007. 98 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2007.

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.

_____. Modelo logístico difásico no estudo do crescimento de fêmeas da raça Hereford. **Ciência Rural**, Santa Maria, v. 38, n. 7, p. 1984-1990, out. 2008.

MENDES, P. R.; AQUINO, L. H.; MUNIZ, J. A. Estudo da curva de crescimento de fêmeas da raça Nelore. **Ciência e Agrotecnologia**, Lavras, v. 25, n. 4, p. 958-968, jul./ago. 2001.

MENDOZA, M. A. Bayesian analysis of the slope ratio bioassay. **Biometrics**, Alexandria, v. 46, n. 4, p. 1059-1069, Dec. 1990.

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

MOURA, M. A. de. **Crescimento pós-colheita de frutos do tomateiro cv. Santa Clara e do seu Mutante Firme**. 1999. 86 f. Dissertação (Mestrado em Agronomia) - Universidade Federal de Viçosa, Viçosa, MG, 1999.

MOURA, M. L. et al. Crescimento e desenvolvimento de frutos do tomateiro "Santa Clara" e do seu mutante natural "Firme". **Ciência e Agrotecnologia**, Lavras, v. 28, n. 6, p. 1284-1290, nov./dez. 2004.

MUNIZ, J. A. et al. Evaluation of lactation curve of low-yielding gir cows: a Bayesian approach. **Acta Scientiarum, Animal Sciences**, Maringá, v. 29, n. 1, p. 79-83, 2007.

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.

PAIXÃO, F. J. R. da et al. Estimativa da infiltração da água no solo através de modelos empíricos e funções não-lineares. **Revista de Biologia e Ciências da Terra**, Campina Grande, v. 5, n. 1, 2004. Disponível em: <<http://www.eduep.uepb.edu.br/rbct/sumarios/pdf/estimativa.pdf>>. Acesso em: 5 nov. 2010.

PAULINO, C. D.; TURKMAN, A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 446 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. et al. Comparação entre modelos para predição do nitrogênio mineralizado: uma abordagem bayesiana. **Ciência e Agrotecnologia**, Lavras, v. 33, p. 1792-1797, 2009. Edição especial.

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.

PIRES, L. M.; CALEGARIO, N. Ajuste de modelos estocásticos lineares e não-lineares para a descrição do perfil longitudinal de árvores. **Revista Árvore**, Viçosa, MG, v. 31, n. 5, p. 845-852, set./out. 2007.

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

RAFTERY, A. L.; LEWIS, S. How many iterations in the Gibbs sampler? In: BERNARDO, J. M. et al. (Ed.). **Bayesian statistics**. Oxford: Oxford University, 1992. p. 763-774.

REIS, R. L. dos et al. Abordagem bayesiana da sensibilidade de modelos para o coeficiente de endogamia. **Ciência Rural**, Santa Maria, v. 39, n. 6, p. 1752-1759, set. 2009.

REZENDE, D. M. L. C. et al. Ajuste de modelos de platô de resposta para a exigência de zinco em frangos de corte. **Ciência e Agrotecnologia**, Lavras, v. 31, n. 2, p. 468-478, mar./abr. 2007.

RODRIGUES, P. B.; MUNIZ, J. A.; PEREIRA, F. A. Estudo comparativo de curvas de crescimento em suínos. **Ciência e Prática**, Lavras, v. 16, n. 1, p. 151-157, jan./mar. 1992.

ROSA, G. J. M. **Análise Bayesiana de modelos lineares mistos robustos via Amostrador de Gibbs**. 1998. 57 f. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 1998.

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

_____. Comparison of two mathematical models in the description of in situ degradability. **Revista Brasileira de Veterinária**, Seropédica, v. 14, n. 2, p. 72-78, maio/ago. 2007a.

SAVIAN, T. V. et al. Degradability study of neutral detergent fiber of coast cross (*Cynodon dactylon* x *Cynodon nlemfunensis*). **Ciência e Agrotecnologia**, Lavras, v. 31, n. 4, p. 1184-1190, jul./ago. 2007b.

_____. Study of in situ degradability: heterogeneity of variances and correlated errors. **Scientia Agricola**, Piracicaba, v. 64, n. 5, p. 548-554, set./out. 2007c.

SEARLE, S. R. **Linear models**. New York: J. Wiley, 1971. 536 p.

SILVA, F. F.; AQUINO, L. H.; OLIVEIRA, A. I. G. Influência de fatores genéticos e ambientais sobre as estimativas dos parâmetros das funções de crescimento em gado nelore. **Ciência e Agrotecnologia**, Lavras, v. 25, n. 5, p. 1195-1205, set./out. 2001.

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

SILVA, F. F. et al. Comparação Bayesiana de modelos de previsão para diferenças esperadas nas progênes no melhoramento genético do gado Nelore. **Pesquisa Agropecuária Brasileira**, Brasília, v. 43, n. 1, p. 37-45, jan. 2008.

SILVA, N. A. M. et al. Aplicação do método Bayesiano na estimação de curva de crescimento em animais da raça Nelore. **Revista Ceres**, Viçosa, MG, v. 54, n. 2, p. 191-198, mar./abr. 2007.

_____. Estudo de parâmetros de crescimento de bezerros Nelore por meio de um modelo de regressão linear: uma abordagem Bayesiana. **Ciência Animal Brasileira**, Goiânia, v. 7, n. 1, p. 57-65, jan./mar. 2006.

_____. 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.

SIMÕES, N. R. **Estimadores de Bayes**: aplicação ao modelo de regressão linear simples. 1981. 138 f. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 1981.

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

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, jan./mar. 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, 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 JÚNIOR, P. A. et al. Estimativas do potencial total de água em sementes, usando modelos não-lineares. **Revista Brasileira de Sementes**, Brasília, v. 21, n. 1, p. 87-92, jan./mar. 1999.

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.