

Modelo hierárquico bayesiano aplicado na avaliação genética de curvas de crescimento de bovinos de corte

[Bayesian hierarchical model applied to genetic evaluation of beef cattle growth curves]

N.A.M. Silva¹, R.R. Lima², F.F. Silva³, J.A. Muniz²

¹Escola de Veterinária - UFMG
Caixa Postal 567

30123-970 – Belo Horizonte, MG

²Departamento de Ciências Exatas - UFLA – Lavras, MG

³Departamento de Estatística - UFV – Viçosa, MG

RESUMO

A estimação dos componentes de (co)variância dos parâmetros de modelos de crescimento pode ser feita por vários métodos. A metodologia bayesiana se apresenta como uma forma alternativa de estimação. Foi realizado um estudo, por meio de dados simulados e de dados reais de animais Nelore, para a estimação dos componentes de (co)variância dos parâmetros do modelo de crescimento de Von Bertalanffy, por meio da metodologia hierárquica bayesiana. Com base nos componentes estimados, foram encontradas as herdabilidades para cada parâmetro do modelo e as correlações genéticas e ambientais entre esses parâmetros. As distribuições marginais a *posteriori* dos parâmetros a , \mathbf{R} , $\boldsymbol{\mu}$, \mathbf{u} , \mathbf{G} e σ_i^2 foram obtidas por meio do algoritmo Gibbs Sampler e as dos parâmetros b e k por meio do algoritmo Metropolis-Hastings. A metodologia se mostrou eficiente, proporcionando estimativas para os parâmetros próximas aos valores simulados. Os parâmetros a e k dos dados reais apresentaram valores de herdabilidades compatíveis com a realidade, indicando que esses parâmetros poderiam ser usados para fins de seleção.

Palavras-chave: gado Nelore, componentes de (co)variância, metodologia hierárquica bayesiana, simulação

ABSTRACT

The estimation of the (co)variance components for the parameters of the growth models can be evaluated by many methods. The Bayesian approach is an alternative method of the estimation. A study was performed using simulated and real data from Nelore cattle for estimation of the (co)variance components for the parameters of Von Bertalanffy growth curve, using a bayesian hierarchical model. From the estimated components, the heritabilities for each parameter and genetic and environmental correlations between these parameters were determined. The samples of posterior marginal distributions for the parameters a , \mathbf{R} , $\boldsymbol{\mu}$, \mathbf{u} , \mathbf{G} , and σ_i^2 were obtained by using Gibbs Sampler algorithm and for the parameters b e k by using the Metropolis-Hastings algorithm. The efficiency of the bayesian inference methodology was verified since estimated parameters were quite close to the simulated ones. The parameters a and k from real data showed heritabilities compatible with the reality indicating they could be used in selection programs.

Keywords: Nelore cattle, (co)variance components, bayesian hierarchical methodology, simulation

INTRODUÇÃO

O aumento do peso vivo ao longo da vida dos animais é um fenômeno complexo e dependente dos efeitos ambientais e genéticos que atuam ao longo do tempo. De acordo com Tedeschi et al.

(2000), o conhecimento e o controle do crescimento dos animais é fato importante tanto para criadores como para pesquisadores, pois o seu domínio permite que o manejo nutricional e programas de seleção dos animais possam ser conduzidos eficientemente.

Uma das formas de se avaliar o crescimento animal é por meio de curvas de crescimento, obtendo-se as estimativas dos parâmetros de modelos não lineares. Silva et al. (2001) destacam o modelo não linear de Von Bertalanffy como adequado para a descrição do crescimento de bovinos. De acordo com esses autores, esse modelo apresenta formulação matemática relativamente simples, inclusive com a definição de um ponto de inflexão, cujo valor é fixo em, aproximadamente, 30% do peso assintótico.

Nesse modelo, assim como em outros modelos de crescimento, o parâmetro a representa a estimativa de peso à maturidade (peso adulto); o parâmetro k , a taxa de maturidade (indica a velocidade de crescimento) e o parâmetro b não possui interpretação biológica. Por meio das estimativas desses parâmetros, é possível obter informações sobre as fases de crescimento dos animais (Guedes et al., 2005, Silva et al., 2008a).

Além dos parâmetros, que são estimados diretamente do modelo e que apresentam interpretação biológica, podem ser obtidos também os componentes do (co)variância relativos a esses parâmetros e, com base nas estimativas destes componentes, são obtidas as estimativas de parâmetros genéticos (herdabilidades e correlações).

Normalmente, essas estimativas são obtidas por meio de uma metodologia frequentista, que considera duas etapas distintas: a primeira corresponde ao ajuste individual dos modelos não lineares; e a segunda corresponde à caracterização das estimativas obtidas na primeira etapa, como variáveis dependentes em um modelo animal, contendo efeitos fixos (ambientais) e aleatórios (genéticos), de forma a se obter os componentes de (co)variância.

Um fato importante decorrente da aplicação dessa metodologia é que os fatores genéticos e ambientais não são considerados no processo de estimação individual, ou seja, as estimativas dos parâmetros não são corrigidas simultaneamente para estes fatores. Assim, uma possível solução para este problema é a utilização da inferência bayesiana considerando um modelo hierárquico em vários estágios. Essa metodologia contempla a estimação simultânea dos parâmetros e dos fatores genéticos e ambientais que atuam sobre eles, havendo, assim, uma correção instantânea para esses fatores, devido às atualizações realizadas no processo iterativo (Guedes et al., 2005; Silva et al., 2008b). Neste contexto, todos os parâmetros do modelo, sejam

fixos ou aleatórios, são considerados como variáveis aleatórias (Muniz et al., 2007).

Na teoria bayesiana, informações a *priori* sobre os parâmetros a serem estimados são utilizadas em associação com os dados amostrais (função de verossimilhança), gerando, assim, uma distribuição conjunta a *posteriori*, de forma que: *posteriori* \propto verossimilhança \times *priori* (Silva et al., 2005).

Varona et al. (1999) compararam a eficiência deste método em relação à análise realizada em dois estágios e mostraram que, com esta última, ocorreram perdas de informações, que refletiram em estimativas viesadas de parâmetros genéticos, por não considerar diretamente efeitos fixos e aleatórios no processo de obtenção das curvas.

O objetivo deste trabalho foi utilizar a inferência bayesiana para estimar os componentes de (co)variância dos parâmetros do modelo de Von Bertalanffy e com base nesses componentes obter as herdabilidades e correlações genéticas e ambientais para cada um destes parâmetros. Essa estimação foi feita considerando-se dados simulados e dados observados de animais da raça Nelore.

A simulação dos dados foi conduzida por meio de populações geradas em duas etapas distintas, segundo procedimentos apresentados por Varona et al. (1999).

Na primeira, os parâmetros da função de Von Bertalanffy e os valores genéticos foram gerados por meio de distribuições normais multivariadas :

MATERIAL E MÉTODOS

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{k} \end{pmatrix} \sim N \begin{pmatrix} \mu_a + \mathbf{Z}\mathbf{u}_a \\ \mu_b + \mathbf{Z}\mathbf{u}_b, \mathbf{R} \otimes \mathbf{I} \\ \mu_k + \mathbf{Z}\mathbf{u}_k \end{pmatrix} e \begin{pmatrix} \mathbf{u}_a \\ \mathbf{u}_b \\ \mathbf{u}_k \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0, \mathbf{G} \otimes \mathbf{A} \\ 0 \end{pmatrix}$$

em que: \mathbf{a} , \mathbf{b} , e \mathbf{k} são os vetores, $N \times 1$, de parâmetros da função de Von Bertalanffy; μ_a , μ_b e μ_k são as médias fenotípicas dos parâmetros a , b e k , respectivamente; \mathbf{Z} é a matriz de incidência dos efeitos aleatórios, com dimensão $N \times q$, sendo N o número total de animais e q o número de animais com informações no *pedigree*; \mathbf{u}_a , \mathbf{u}_b e \mathbf{u}_k são os vetores, $N \times 1$, de valores genéticos de cada parâmetro; \mathbf{R} é a matriz 3×3 de (co)variância residual; \mathbf{I} é uma matriz identidade $N \times N$; \mathbf{G} é matriz 3×3 , de (co)variância genética e \mathbf{A} é a matriz de parentesco entre os animais.

Na segunda etapa, os dados de peso de cada animal (y_{ij}) foram simulados considerando-se uma distribuição normal, sendo que $y_{ij} | a_i, b_i, k_i, \sigma_e^2 \sim N\left\{a_i \times [1 - b_i \exp(-k_i t_{ij})]^3, \sigma_e^2\right\}$ em que a_i , b_i e k_i são parâmetros da curva de crescimento gerados na etapa anterior, para o animal i , t_{ij} refere-se à idade em que o peso y_{ij} foi observado e σ_e^2 é a variância do erro, sendo $i = 1, 2, \dots, N$ e $j = 1, 2, \dots, n$, em que n é o número de medidas por animal.

A população necessária à formação da matriz **A** foi definida considerando-se três gerações sucessivas, não sobrepostas, tal que cada uma delas continha 15 machos e 135 fêmeas, perfazendo um total de 150 animais por geração e, conseqüentemente, 450 animais na população.

As médias μ_a , μ_b e μ_k foram de 700, 0,4 e 0,003 para os parâmetros a , b e k , respectivamente. As variâncias genéticas e ambientais para cada parâmetro foram consideradas iguais, assumindo os seguintes valores: $\sigma_g^2(a) = \sigma_r^2(a) = 5000$, $\sigma_g^2(b) = \sigma_r^2(b) = 3,25 \times 10^{-4}$ e $\sigma_g^2(k) = \sigma_r^2(k) = 1,25 \times 10^{-7}$, sendo estes os mesmos valores utilizados por Varona et al. (1999). Assim, a herdabilidade foi igual a 0,5 para cada parâmetro. As correlações genéticas e residuais foram assumidas nulas. Foram geradas 15 pesagens por animal, com σ_e^2 igual a 100kg². O processo de simulação foi repetido 18 vezes.

Os dados observados e utilizados neste trabalho são oriundos de machos e fêmeas da raça Nelore tradicional, nascidos entre os anos de 2001 a 2004 em dois rebanhos da região de Teófilo Otoni, MG. Estes dados fazem parte dos arquivos de registros da Associação Brasileira de Criadores de Zebu, com sede em Uberaba, MG e se referem a pesagens aproximadamente trimestrais, no período do nascimento aos 500 dias de idade. Para que o ajuste das funções de crescimento não fosse comprometido por irregularidades dos pesos observados, animais com menos de sete pesagens foram excluídos do estudo. Não houve a separação destes animais por grupos contemporâneos, pela falta de informações no arquivo de dados. Assim, animais nascidos em diferentes épocas e

condições de manejo foram todos considerados do mesmo grupo.

Um modelo hierárquico em três estágios foi utilizado para a estimação dos componentes de (co)variância. O primeiro estágio é representado pela modelagem individual da forma da curva de crescimento de cada animal, sendo este descrito pelo modelo de crescimento de Von Bertalanffy, que é definido por: $y_{ij} = a_i [1 - b_i \exp(-k_i t_{ij})]^3 + e_{ij}$, em que y_{ij} é o peso do indivíduo i tomado no tempo j ; ($i = 1, 2, \dots, N$; $j = 1, 2, \dots, n$); a_i , b_i e k_i são os parâmetros específicos de cada indivíduo i ; t_{ij} é o tempo correspondente à observação y_{ij} e e_{ij} é o erro aleatório, independentemente e normalmente distribuído, com média 0 e variância σ_e^2 .

O segundo estágio da hierarquia descreve como os parâmetros variam entre indivíduos. Cada parâmetro do modelo é determinado por um efeito fixo μ e por um efeito genético u , além de um efeito aleatório não controlado denotado por ε . Considerando essas definições, pode-se utilizar um modelo linear misto para representar a estrutura desses efeitos, que, na forma matricial, é dado por: $\mathbf{p} = \mathbf{X}\boldsymbol{\mu} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$, em que \mathbf{p} é um vetor de parâmetros $3N \times 1$, sendo $\mathbf{p} = [a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n, k_1, k_2, \dots, k_n]$; \mathbf{X} é uma matriz, $3N \times 3N$, de incidência dos efeitos fixos; $\boldsymbol{\mu}$ é um vetor, $3N \times 1$, de efeitos fixos; \mathbf{Z} é uma matriz, $3N \times 3q$, de incidência dos efeitos genéticos aditivos; \mathbf{u} é um vetor, $3q \times 1$, de efeitos genéticos aditivos, tal que $\mathbf{u} \sim N(0, \mathbf{G} \otimes \mathbf{A})$ e $\boldsymbol{\varepsilon}$ é um vetor $3N \times 1$ de erros aleatórios associados ao modelo, tal que $\boldsymbol{\varepsilon} \sim N(0, \mathbf{R} \otimes \mathbf{I})$;

O terceiro estágio introduzido na estrutura hierárquica consiste na descrição das distribuições *a priori* acerca de todos os parâmetros considerados no primeiro e segundo estágios. Para $\boldsymbol{\mu}$, \mathbf{u} e \mathbf{p} foram atribuídas distribuições normais multivariadas, para as matrizes \mathbf{R} e \mathbf{G} distribuições de Wishart invertida e para σ_e^2 distribuição uniforme (Chang et al., 2001; Blasco et al., 2003).

A função de verossimilhança é dada por:

$$f(y_i | p) = (\sigma_e^2)^{-\sum_{i=1}^N n/2} \exp \left[-\frac{1}{2\sigma_e^2} (y_i - a_i h_i)' (y_i - a_i h_i) \right], \quad (1)$$

em que $h_i = [1 - b_i \exp(-k_i t_{ij})]^3$.

Fazendo-se o produto da verossimilhança com as distribuições a priori, tem-se a distribuição conjunta a posteriori que é representada da seguinte forma:

$$f(\mathbf{p}, \boldsymbol{\mu}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \sigma_e^2 | \mathbf{y}_i) \propto f(\mathbf{y}_i | \mathbf{p}) f(\mathbf{p} | \boldsymbol{\mu}, \mathbf{u}, \mathbf{R}) f(\boldsymbol{\mu}) f(\mathbf{u} | \mathbf{G}) f(\mathbf{G}) f(\mathbf{R}) f(\sigma_e^2),$$

em que \mathbf{y}_i denota o vetor de dados observados para cada animal i .

A integração da distribuição conjunta a posteriori para a obtenção das marginais não tem solução analítica, sendo muito difícil e trabalhosa de ser resolvida por métodos numéricos. Portanto, faz-se necessário o uso de algoritmos especializados, que utilizam os métodos denominados Markov Chain - Monte Carlo (MCMC). Para a utilização desses algoritmos, é necessário que se obtenha, a partir da distribuição conjunta a posteriori, um conjunto

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

As distribuições condicionais completas a posteriori utilizadas para cada parâmetro, estão descritas a seguir. As matrizes de (co)variância ambiental (\mathbf{R}) e genética (\mathbf{G}) seguem distribuições de Wishart invertida (Sorensen, 1996), ou seja, $\mathbf{G} \sim IW_3 \left\{ (\mathbf{U}' \mathbf{A}^{-1} \mathbf{U} + n_G \mathbf{S}_G^2)^{-1}, n_G + q \right\}^e$

$\mathbf{R} \sim IW_3 \left\{ (\mathbf{E}' \mathbf{I}_N \mathbf{E} + n_R \mathbf{S}_R^2)^{-1}, n_R + N \right\}$, em que

$$\mathbf{U}' \mathbf{A}^{-1} \mathbf{U} = \begin{bmatrix} \mathbf{u}_a' \\ \mathbf{u}_b' \\ \mathbf{u}_c' \end{bmatrix} \mathbf{A}^{-1} \begin{bmatrix} \mathbf{u}_a & \mathbf{u}_b & \mathbf{u}_c \end{bmatrix} \quad \text{e} \quad \mathbf{E}' \mathbf{I}^{-1} \mathbf{E} = \begin{bmatrix} \mathbf{e}_a' \\ \mathbf{e}_b' \\ \mathbf{e}_c' \end{bmatrix} \mathbf{I}_N \begin{bmatrix} \mathbf{e}_a & \mathbf{e}_b & \mathbf{e}_c \end{bmatrix}.$$

Seguindo as definições de Chang et al. (2003), o parâmetro a e os vetores $\boldsymbol{\mu}$ e \mathbf{u} são normalmente distribuídos, ou seja, $a_i | b_i, k_i, \boldsymbol{\mu}, \mathbf{u}, \mathbf{R}, \mathbf{G}, \mathbf{y} \sim N \left[\hat{a}_i, (h_i' \sigma_e^{-2} h_i + r^{aa})^{-1} \right]$

em que h_i está definido em (1) e $\hat{a}_i = \frac{h_i' \sigma_e^{-2} y_i + r^{aa} m_{a,i} - r^{ab} (b_i - m_{b,i}) - r^{ak} (k_i - m_{k,i})}{h_i' \sigma_e^{-2} h_i + r^{aa}}$,

$$\boldsymbol{\mu} | \mathbf{u}, \mathbf{R}, \mathbf{G}, \mathbf{p} \sim N \left[(\mathbf{X}' \mathbf{R}^{-1} \otimes \mathbf{I} \mathbf{X} + \mathbf{V}^{-1})^{-1} \mathbf{X}' \mathbf{R}^{-1} \otimes \mathbf{I} (\mathbf{p} - \mathbf{Z} \mathbf{u}), (\mathbf{X}' \mathbf{R}^{-1} \otimes \mathbf{I} \mathbf{X} + \mathbf{V}^{-1})^{-1} \right]$$

$$\mathbf{u} | \boldsymbol{\mu}, \mathbf{R}, \mathbf{G}, \mathbf{p} \sim N \left[(\mathbf{Z}' \mathbf{R}^{-1} \otimes \mathbf{I} \mathbf{Z} + \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})^{-1} \mathbf{Z}' \mathbf{R}^{-1} \otimes \mathbf{I} (\mathbf{p} - \mathbf{X} \boldsymbol{\mu}), (\mathbf{Z}' \mathbf{R}^{-1} \otimes \mathbf{I} \mathbf{Z} + \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})^{-1} \right].$$

A distribuição da variância do erro (σ_e^2) foi identificada como sendo uma gama inversa, isto é,

$$\sigma_e^2 \sim GI \left[N/2 - 1, \frac{1}{2} \sum_i \sum_j \{y_{ij} - a_i [1 - b_i \exp(-k_i t_{ij})]^3\}^2 \right].$$

Para os parâmetros b e k , as distribuições obtidas não são representadas por distribuições conhecidas, uma vez que

$$f(b_i | a_i, k_i, \boldsymbol{\mu}, \mathbf{u}, \mathbf{R}, \mathbf{G}, \mathbf{y}) \propto$$

$$\exp \left\{ -\frac{1}{2\sigma_e^2} (y_i - a_i h_i)' (y_i - a_i h_i) - \frac{1}{2r^{bb}} \left[r^{bb} b_i - (r^{bb} m_{b,i} - r^{ab} (a_i - m_{a,i})) - r^{bk} (k_i - m_{k,i}) \right]^2 \right\} \quad (2)$$

$f(k_i | a_i, b_i, \mu, u, R, G, y) \propto$

$$\exp \left\{ -\frac{1}{2\sigma_e^2} (y_i - a_i h_i)' (y_i - a_i h_i) - \frac{1}{2r^{kk}} \left[r^{kk} k_i - (r^{bk} m_{k,i} - r^{ak} (a_i - m_{a,i})) - r^{bk} (b_i - m_{b,i}) \right]^2 \right\}, \quad (3)$$

em que h_i está definido em (1), $[m_{a,i}, m_{b,i}, m_{k,i}]' = X_i \mu + Z_i u$ e $r^{bb}, r^{kk}, r^{ab}, r^{ak}$ e r^{bk} são elementos da

inversa da matriz R , dada por $r = R^{-1} = \begin{bmatrix} r^{aa} & r^{ab} & r^{ac} \\ r^{ab} & r^{bb} & r^{bc} \\ r^{ac} & r^{bc} & r^{cc} \end{bmatrix}$.

Assim, para os parâmetros a , G , R , μ , u e σ_e^2 , foi utilizado o algoritmo Gibbs Sampler, pois apresentaram distribuições condicionais conhecidas. Para os parâmetros b e k , foi utilizado o algoritmo Metropolis-Hastings, pois não apresentaram distribuições condicionais conhecidas. Para a implementação deste, foram utilizadas cadeias de passeios aleatórios, conforme descrito por Lima (2005), para garantir que os valores simulados para b e k se mantivessem restritos aos seus intervalos paramétricos, isto é, b e $k \in [0,1]$.

O valor candidato a b_{i+1} , b^* , é gerado a partir de b_i , utilizando-se

$$\text{logit}(b^*) = \text{logit}(b_i) + w_i, \quad (4)$$

em que $\text{logit}(b) = \ln \left(\frac{b}{1-b} \right)$ e w_i é um valor simulado da variável aleatória W , tal que $W \sim N(0, \sigma_b^2)$, em que σ_b^2 é um hiperparâmetro. Essa transformação de variável na geração de valores candidatos de b é necessária, pois $b \in [0,1]$ e $w \in [-\infty, \infty]$. Processo semelhante foi utilizado para o parâmetro k , pois este apresenta o mesmo espaço paramétrico que b . Neste caso, apenas utilizou-se o

$$q(b_i, b^*) = \frac{1}{b^*(1-b^*)} \times \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{1}{2\sigma^2} \left[\ln \left(\frac{b^*}{1-b^*} \right) - \mu_{b^*} \right]^2 \right\} I_{(0,1)}(b^*),$$

em que $\mu_{b^*} = \ln \left(\frac{b_i}{1-b_i} \right)$. Vale lembrar que $q(b^*, b_i)$ é definido da mesma maneira, bastando substituir b^* por b_i e vice-versa.

hiperparâmetro σ_k^2 , ao invés de σ_b^2 . Assim, a

partir da equação (4), tem-se $b^* = \left(1 + \frac{1-b_i}{b_i e^{w_i}} \right)^{-1}$

e, de forma similar, $k^* = \left(1 + \frac{1-k_j}{k_j e^{w_j}} \right)^{-1}$.

Uma vez gerados b^* e k^* , deve-se verificar se estes serão aceitos ou não. De acordo com o algoritmo de Metropolis-Hastings, o valor proposto, b^* , será aceito com probabilidade $\delta(b_i, b^*)$, dada por

$$\delta(b_i, b^*) = \min \left\{ 1, \frac{f(b^* | a_i, k_i, \mu, u, R, G, y) q(b^*, b_i)}{f(b_i | a_i, k_i, \mu, u, R, G, y) q(b_i, b^*)} \right\} \\ = \min \left\{ 1, \frac{f(b^* | a_i, k_i, \mu, u, R, G, y) b^*(1-b_i)}{f(b_i | a_i, k_i, \mu, u, R, G, y) b_i(1-b^*)} \right\},$$

em que $q(b_i, b^*)$ é a função de transição, que é obtida utilizando-se o conceito de transformação de variáveis aleatórias, ou seja,

Assim, os valores gerados b^* e k^* são os valores candidatos a serem aceitos ou não pelo algoritmo Metropolis-Hastings e as variâncias σ_b^2 e σ_k^2 interferem diretamente na porcentagem de aceitação dos valores propostos, b^* e k^* , respectivamente. Portanto, os valores de σ_b^2 e σ_k^2 foram definidos de tal maneira que a taxa de aceitação dos valores candidatos permanecessem entre 17 e 45%, conforme recomendação de Blasco et al. (2003).

Os algoritmos Gibbs Sampler e Metropolis-Hastings foram implementados no *software* R (A language ..., 2007). Foram executadas 51000 iterações, respeitando, respectivamente, um *burn-in* e *thin* de 1000 e 10 iterações. Assim, obteve-se ao final uma cadeia de 5000 iterações. Tratando-se de processos iterativos, a constatação de sua convergência é imprescindível para a validade do estudo. Portanto, para verificar a convergência, utilizaram-se os critérios de Geweke (1992), e de Raftery e Lewis (1992), disponíveis no pacote Bayesian Output Analysis (BOA) do aplicativo R.

As distribuições marginais para as herdabilidades (h^2), correlações ambientais (r_r) e correlações genéticas (r_g) foram obtidas da combinação dos componentes de (co)variância estimados, como indicado por Rosa e Padovani (2000).

RESULTADOS E DISCUSSÃO

Na Tab. 1, são apresentadas as estimativas (médias *a posteriori*) das variâncias genéticas e ambientais dos parâmetros a , b e k , das herdabilidades de cada parâmetro e das correlações genéticas e residuais entre os parâmetros. Os valores apresentados são as médias das 18 repetições do processo de simulação. De acordo com esses resultados, verifica-se que o processo de simulação foi eficiente. Todos os valores médios estimados para os componentes de variância ficaram próximos dos valores simulados. Todos esses resultados concordam com os de Varona et al. (1999), que também encontraram valores estimados muito próximos dos simulados, sendo até melhores do que os encontrados no presente estudo. Isso se deve, provavelmente, ao uso de

condicionais diferentes no processo de simulação, pois esses autores usaram normais univariadas e, neste trabalho, foram utilizadas normais multivariadas.

Em relação à constatação da convergência das cadeias, em geral os valores do escore Z foram todos menores que 1,96, conforme recomendação de Blasco et al. (2003), o que, segundo Geweke (1992), significa estacionariedade da cadeia gerada. Além disso, os valores do fator de dependência (FD) de Raftery e Lewis permaneceram praticamente todos menores que 5, o que, segundo Nogueira et al. (2004), significa que a convergência já foi atingida. Outro indicativo de qualidade de convergência são os valores encontrados para os Erros de Monte Carlo (EMC), que foram de baixa magnitude, o que, segundo Blasco et al. (2003), também asseguram convergência.

A Fig. 1 é relativa aos intervalos de maior densidade *a posteriori* (HPD) das herdabilidades dos parâmetros a e k e da correlação genética entre eles. Por essa figura pode-se afirmar que a metodologia utilizada foi eficiente, pois todos os valores paramétricos considerados na simulação ficaram incluídos no HPD. Para a herdabilidade do parâmetro k , o valor simulado ficou bem próximo ao limite inferior do HPD. Isso se deve provavelmente ao fato de k ser um parâmetro que exige mais complexidade para sua estimação, por ser um parâmetro de condicional não conhecida e necessitar do uso do algoritmo Metropolis-Hastings.

Em relação aos dados reais, observa-se, na Tab. 2, que a convergência foi alcançada, pois os valores de Z do teste de Geweke não foram superiores a 1,96, os FD foram menores que 5, além de a maioria dos valores de EMC serem próximos de zero, o que significa convergência, como foi obtido com os dados simulados. Portanto, nenhuma irregularidade foi observada no processo de amostragem, o que assegura a constatação da convergência requerida.

Para o parâmetro a , a estimativa de herdabilidade encontrada foi de 0,21, que é considerada moderada. Isso indica que esse parâmetro poderia ser utilizado para fins de seleção, semelhante ao resultado obtido por Elias (1998), que relatou valores de herdabilidade para o parâmetro a de animais Nelore que variaram de

0,04 a 0,69, para o modelo de crescimento de Von Bertalanffy.

Para o parâmetro k, a estimativa da herdabilidade encontrada foi de 0,14, que, apesar de não ser alta, também poderia ser utilizada para fins de seleção, concordando com Elias (1998) que relatou valores de herdabilidade variando de 0,12 a 0,44, mas diferindo de Forni (2007), que encontrou valores mais baixos para a herdabilidade de k (0,06).

Os valores de herdabilidade encontrados neste trabalho (herdabilidades baixas a moderadas) se devem ao valor alto da variância residual de cada parâmetro nesta população. Isso provavelmente ocorreu porque os animais foram considerados todos do mesmo grupo contemporâneo, logo não

houve separação ambiental entre esses animais no processo de estimação dos parâmetros, o que, provavelmente, fez com que a variância residual ficasse mais elevada. Assim, a consideração dos efeitos ambientais no processo de estimação poderia aumentar as estimativas das herdabilidades.

Segundo Elias (1998), observa-se uma grande variabilidade nas estimativas de herdabilidades dos parâmetros dos vários estudos de crescimento de bovinos. Segundo esse autor, isso é de se esperar, pois os parâmetros obtidos nesses estudos podem ser considerados extrapolações de dados de pesagens obtidas somente até os 24 meses de idade, quando os animais ainda não atingiram a idade adulta.

Tabela 1. Valores paramétricos utilizados nas simulações, médias e desvios-padrões das estimativas dos componentes de variância, herdabilidades, correlações genéticas e residuais das 18 repetições do processo de simulação

Variância e parâmetros genéticos	Parâmetro	Média	Desvio-padrão
$\sigma_g^2(a)^{(1)}$	5000	5684,93	243,53
$\sigma_g^2(b)^{(1)}$	$3,25 \times 10^{-4}$	$3,42 \times 10^{-4}$	$9,99 \times 10^{-6}$
$\sigma_g^2(k)^{(1)}$	$1,25 \times 10^{-7}$	$1,42 \times 10^{-7}$	$1,48 \times 10^{-8}$
$\sigma_r^2(a)^{(1)}$	5000	4528,52	224,32
$\sigma_r^2(b)^{(1)}$	$3,25 \times 10^{-4}$	$3,14 \times 10^{-4}$	$5,25 \times 10^{-7}$
$\sigma_r^2(k)^{(1)}$	$1,25 \times 10^{-7}$	$1,34 \times 10^{-7}$	$3,89 \times 10^{-10}$
$h_a^2^{(1)}$	0,5	0,55	0,002
$h_b^2^{(1)}$	0,5	0,52	0,007
$h_k^2^{(1)}$	0,5	0,52	0,007
$r_g(a, b)^{(2)}$	0	0,017	0,025
$r_g(a, k)^{(3)}$	0	-0,006	0,02
$r_g(b, k)^{(4)}$	0	0,008	0,01
$r_r(a, b)^{(2)}$	0	0,08	0,09
$r_r(a, k)^{(3)}$	0	-0,15	0,18
$r_r(b, k)^{(4)}$	0	-0,02	0,05

⁽¹⁾variâncias genéticas, residuais e herdabilidades para os parâmetros a, b e k. ⁽²⁾correlação genética e residual entre os parâmetros a e b. ⁽³⁾ correlação genética e residual entre os parâmetros a e k. ⁽⁴⁾ correlação genética e residual entre os parâmetros b e k.

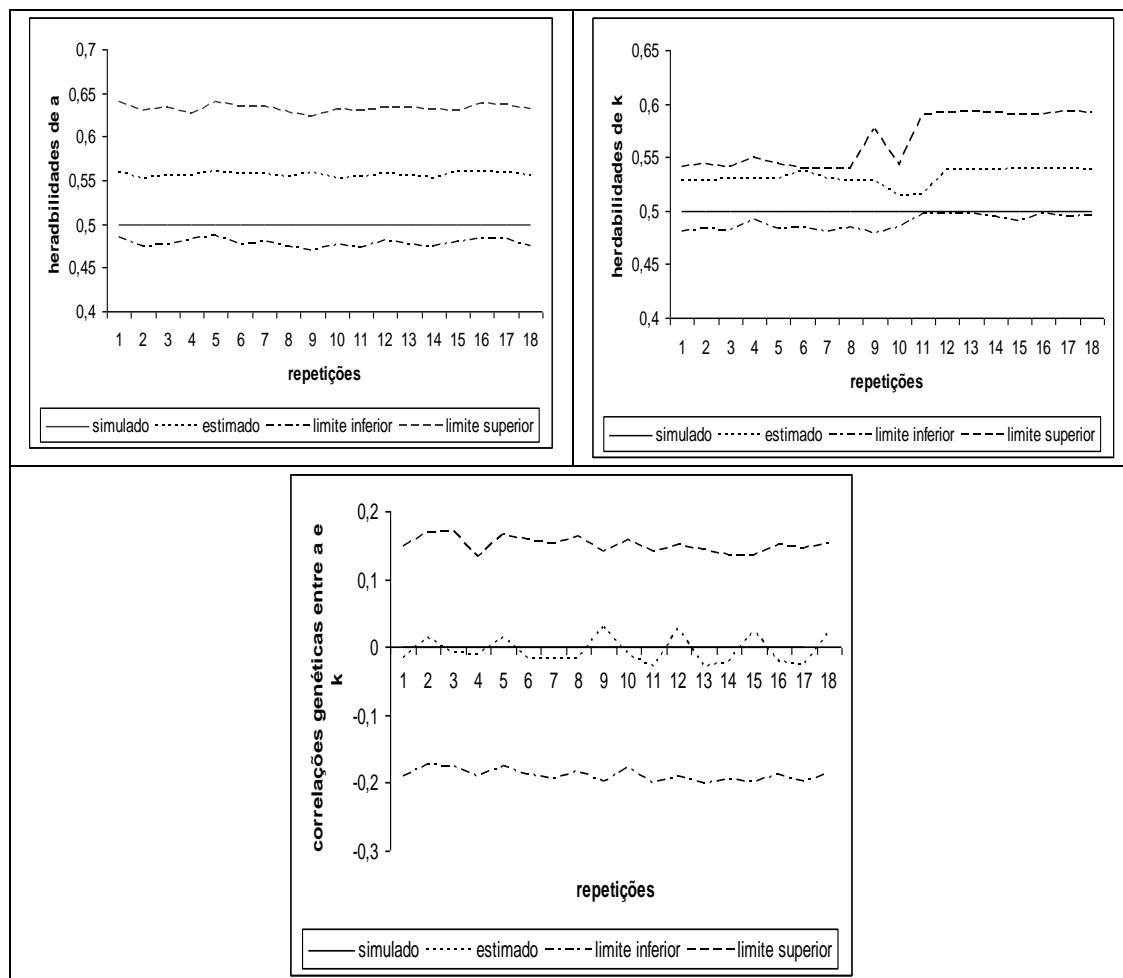


Figura 1. Valores paramétricos, estimados e limite inferior e superior do intervalo de maior densidade a *posteriori* para a herdabilidade dos parâmetros a e k e para a correlação genética entre eles.

Tabela 2. Média, desvio-padrão (DP), intervalo de maior densidade a *posteriori* (HPD), fator de dependência para o teste de Raftery e Lewis (FD), valor de z para o teste de Geweke (Z) e Erro de Monte Carlo (EMC) obtidos da densidade a *posteriori* das variâncias genéticas e ambientais, herdabilidades e correlação genética para os parâmetros a e k para os dados de Nelore

Variâncias e parâmetros genéticos	Média (DP)	HPD		FD	Z	EMC
		LI	LS			
$\sigma_g^2(a)^{(1)}$	2557,60(484,80)	1647,07	3513,54	1,0	1,1	7,65
$\sigma_r^2(a)^{(1)}$	9748,84(1044,15)	7700,08	11800,12	0,9	0,1	18,58
$\sigma_g^2(k)^{(1)}$	$1,92 \times 10^{-7}(7,18 \times 10^{-9})$	$1,78 \times 10^{-7}$	$1,99 \times 10^{-7}$	1,1	0,5	$1,0 \times 10^{-9}$
$\sigma_r^2(k)^{(1)}$	$1,15 \times 10^{-6}(2,66 \times 10^{-7})$	$8,00 \times 10^{-6}$	$1,66 \times 10^{-6}$	2,1	-0,6	$9,5 \times 10^{-9}$
$h_a^2^{(1)}$	0,21(0,03)	0,14	0,26	1,0	1,0	$4,3 \times 10^{-4}$
$h_k^2^{(1)}$	0,14(0,02)	0,09	0,19	3,3	0,8	$9,1 \times 10^{-4}$
$r_g(a, k)^{(2)}$	0,02(0,08)	-0,12	0,19	0,9	-1,3	$1,0 \times 10^{-3}$

⁽¹⁾variâncias genéticas, residuais e herdabilidades para os parâmetros a e k. ⁽²⁾correlação genética entre os parâmetros a e k.

A correlação genética encontrada entre os parâmetros a e k foi de 0,02, que é baixa e positiva e muito próxima de zero, diferindo da maioria dos trabalhos encontrados na literatura, em que essa correlação é normalmente alta e negativa. Segundo Santoro et al. (2005), que encontraram correlações genéticas positivas entre os parâmetros a e k para animais Nelore mocho, esse é um comportamento de difícil explicação. Provavelmente, se deve a problemas de amostragem.

CONCLUSÕES

A metodologia hierárquica bayesiana se mostrou uma alternativa viável de se estimar os componentes de (co)variância dos parâmetros do modelo de Von Bertalanffy, pois produziu boas estimativas do processo de simulação, próximas aos valores simulados e com valores paramétricos sempre dentro dos HPD. Além disso, produziu estimativas para os dados reais condizentes com a realidade. Os parâmetros a e k podem ser usados para fins de seleção em programas de melhoramento animal.

REFERÊNCIAS BIBLIOGRÁFICAS

A LANGUAGE and environment for statistical computing. Viena: R Foundation for Statistical Computing, 2007.

BLASCO, A.; PILES, M.; VARONA, L.A. Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. *Gen. Sel. Evol.*, v.35, p.21-41, 2003.

CHANG, Y.M.; REKAYA, R.; GIANOLA, D. et al. Genetic variation of lactation curves in dairy sheep: a Bayesian analysis of Wood's function. *Liv. Prod. Sci.*, v.71, p.241-251, 2001.

ELIAS, M.A. *Análise de curvas de crescimento de vacas das raças Nelore, Guzerá e Gir*. 1998. 128f. Dissertação (Mestrado) - Escola Superior de Agricultura de "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

FORNI, S. *Análise da curva de crescimento de bovinos da raça nelore utilizando funções não lineares em análises bayesianas* 2007. 75f. Tese (Doutorado) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal.

GUEDES, M.H.P.; MUNIZ, J.A.; SILVA, F.F. et al. Análise Bayesiana da curva de crescimento de cordeiros da raça Santa Inês. *Arq. Bras. Med. Vet. Zootec.*, v.57, p.415-417, 2005.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: BERNARDO, J.M.; BERGER, J.O.; DAWID, A.P. et al. (Eds). *Bayesian statistics*. Oxford: Oxford University, 1992. p.625-631.

LIMA, R.R. *Modelagem espaço-temporal para dados de incidência de doenças em plantas*, 2005. 149f. Tese (Doutorado) - Escola Superior de Agricultura de "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

MUNIZ, J.A.; SILVA, F.F., GUEDES, M.H.P. et al. Evaluation of lactation curve of low-yielding gir cows: a Bayesian approach *Acta Sci. Anim. Sci.*, v.29, p.79-83, 2007.

NOGUEIRA, D.A.; SÁFADI, T.; FERREIRA, D.F. Avaliação de critérios de convergência univariados para o método de Monte Carlo via Cadeias de Markov. *Rev. Bras. Est.*, v.65, p.59-88, 2004.

RAFTERY, A.E.; LEWIS, S. How many iterations in the Gibbs Sampler? In: BERNARDO, J.M.; BERGER, J.O.; DAWID, A.P. et al. (Eds). *Bayesian statistics*. Oxford: Oxford University, 1992. p.763-773.

ROSA, G.J.M.; PADOVANI, C.R. Modelos lineares mistos robustos utilizando-se a distribuição normal contaminada. *Rev. Mat. Est.* v.18, p.65-81, 2000.

SANTORO, R.S.; BARBOSA, S.B.P.; BRASIL, L.H.A. et al. Estimativas de parâmetros de curvas de crescimento de bovinos zebu, criados no estado de Pernambuco. *Rev. Bras. Zootec.*, v.34, p.2262-2279, 2005.

SILVA, F.F.; AQUINO, L.H.; OLIVEIRA, J.A. Influência de fatores genéticos e ambientais sobre as estimativas dos parâmetros das funções de crescimento em gado Nelore. *Cienc. Agrotecnol.*, v.25, p.1195-1205, 2001.

SILVA, F.F.; MUNIZ, J.A.; AQUINO, L.H. et al. Abordagem Bayesiana da curva de lactação de cabras Saanen de primeira e segunda ordens de parto. *Pesq. Agropec. Bras.*, v.40, p.27-33, 2005.

SILVA, F.F.; SÁFADI, T.; MUNIZ, J.A. et al. Comparação bayesiana de modelos de previsão para diferenças esperadas nas progênes no melhoramento genético do gado Nelore. *Pesq. Agropec. Bras.*, v.43, p.37-45, 2008a.

SILVA, N.A.M.; MUNIZ, J.A.; SILVA, F.F. et al. Aplicação do método bayesiano na estimação de curvas de crescimento de animais da raça Nelore. *Rev. Ceres*, v.54, p.192-199, 2008b.

TEDESCHI, L.O.; BOIN, C.; NARDON, R.F. et al. Estudo da curva de crescimento de animais da raça Guzerá e seus cruzamentos alimentados a pasto, com e sem suplementação. 1. análise e seleção das funções não lineares *Rev. Bras. Zootec.*, v.29, p.630-637, 2000.

VARONA, L.; MORENO, C.; GARCIA CORTÉS, L.A. et al. Two-step versus joint analysis of Von Bertalanffy function. *J. Anim. Breed.*, v.116, p.331-338, 1999.