

Implementação da Amostragem de Gibbs para a Estimação de Componentes de Co-variância e Parâmetros Genéticos em Dados de Campo de Bovinos Nelore





*Empresa Brasileira de Pesquisa Agropecuária
Centro de Pesquisa Agropecuária dos Cerrados
Ministério da Agricultura, Pecuária e Abastecimento*

ISSN 1517-5111

Dezembro, 2001

Documentos 37

Implementação da Amostragem de Gibbs para a Estimação de Componentes de Co-variância e Parâmetros Genéticos em Dados de Campo de Bovinos Nelore

Cláudio de Ulhôa Magnabosco
Carina Ubirajara de Faria
Arcadio de los Reyes Borjas
Raysildo Barbosa Lôbo
Roberto Daniel Sainz

Planaltina, DF
2001

Exemplares desta publicação podem ser adquiridos na:

Embrapa Cerrados

BR 020, Km 18, Rod. Brasília/Fortaleza

Caixa Postal 08223

CEP 73301-970 Planaltina - DF

Fone: (61) 388-9898

Fax: (61) 388-9879

<http://www.cpac.embrapa.br>

sac@cpac.embrapa.br

Comitê de Publicações

Presidente: *Ronaldo Pereira de Andrade*

Secretária-Executiva: *Nilda Maria da Cunha Sette*

Membros: *Maria Alice Bianchi, Leide Rovênia Miranda de Andrade, Carlos Roberto Spehar, José Luiz Fernandes Zoby*

Supervisão editorial: *Nilda Maria da Cunha Sette*

Revisão de texto: *Maria Helena Gonçalves Teixeira /
Jaime Arbués Carneiro*

Normalização bibliográfica: *Dauí Antunes*

Capa: *Chaile Cherne Soares Evangelista*

Editoração eletrônica: *Jussara Flores de Oliveira*

1ª edição

1ª impressão (2001): tiragem 300 exemplares

Todos os direitos reservados.

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

CIP-Brasil. Catalogação-na-publicação.
Embrapa Cerrados.

I345 Implementação da Amostragem de Gibbs para a estimação de componentes de co-variância e parâmetros genéticos em dados de campo de bovinos Nelore / Cláudio de Ulhôa Magnabosco ... [et al.]. – Planaltina, DF : Embrapa Cerrados, 2001.
50 p. – (Documentos / Embrapa Cerrados, ISSN 1517-5111; 37)

1. Bovino - genética. I. Magnabosco, Cláudio de Ulhôa. II. Série.

636.0821 - CDD 21

© Embrapa 2001

Autores

Cláudio de Ulhôa Magnabosco
Embrapa Cerrados/Arroz e Feijão
mclaudio@cpac.embrapa.br

Carina Ubirajara de Faria
Embrapa Arroz e Feijão
carina@cnpaf.embrapa.br

Arcadio de los Reyes Borjas
Universidade Federal de Goiás, Departamento de Produção Animal
da EV/UFG, Caixa Postal 131, Goiânia, GO
adreyesb@vet.ufg.br

Raysildo Barbosa Lôbo
Universidade de São Paulo, Departamento de Genética FMRP-USP,
Ribeirão Preto, SP
rayblobo@genbov.fmrp.usp.br

Roberto Daniel Sainz
University of California, Department of Animal Science, Shields Av,
Davis-EUA
rdsainz@ucdavis.edu

Agradecimentos

Ao Sr. Reinaldo Paulino pela compilação do MTGSAM em ambiente Conectiva Linux 5.0.

À Srª Ana Lúcia Delalibera de Faria pelas correções bibliográficas.

Sumário

Introdução	9
Material e Métodos	10
Descrição dos dados	10
Análise dos dados	11
Distribuições iniciais e posteriores	15
Algoritmo de Gibbs e MTGSAM	20
Implementação da Amostragem de Gibbs	23
Resultados e Discussão	25
Estimação dos componentes de co-variância utilizando o REML	25
Distribuições médias posteriores dos componentes de co-variância e parâmetros genéticos usando o GS	30
Densidades posteriores para os parâmetros genéticos	32
Conclusão	45
Referências Bibliográficas	46

Implementação da Amostragem de Gibbs para a Estimação de Componentes de Co-variância e Parâmetros Genéticos em Dados de Campo de Bovinos Nelore

*Cláudio de Ulhôa Magnabosco; Carina Ubirajara de Faria;
Arcadio de los Reyes Borjas; Raysildo Barbosa Lôbo;
Roberto Daniel Sainz*

Introdução

No desenvolvimento da pecuária de corte, exige-se a aplicação de métodos modernos de análise genética, buscando aumentar a produtividade do rebanho nacional. A raça Nelore, por apresentar grande efetivo populacional, permite a aplicação dessas técnicas, sendo necessário, para tanto, mais conhecimentos dos parâmetros genéticos que influenciam o crescimento dessa raça. O aperfeiçoamento de métodos para estimação dos componentes de co-variância, utilizando dados de campo, em características de interesse econômico, como as de crescimento, tem sido preocupação constante dos pesquisadores ao longo dos anos.

Na década de 80, foi possível a aplicação do Método da Máxima Verossimilhança Restrita (REML), em dados de campo. Ainda hoje, esse método é o mais utilizado para estimação de componentes de co-variância. Com o desenvolvimento de algoritmos mais eficientes, concomitantemente, com o aumento da capacidade de processamento dos computadores, tem sido possível o uso do REML em modelos mistos ([Boldman et al., 1993, 1995](#)). [Meyer \(1993\)](#) consolidou a aplicação do algoritmo denominado DFREML (*Derivative Free Restricted Maximum Likelihood* ou Máxima Verossimilhança Restrita por Processo Não-derivativo) desenvolvido por [Smith & Graser \(1986\)](#) e [Graser et al. \(1987\)](#) para estimação de componentes de variância e co-variância para modelos mistos.

Recentemente, métodos bayesianos vêm sendo utilizados como opção para solução de problemas relacionados à avaliação de mérito genético em populações animais. A aplicação de métodos de Markov Chain Monte Carlo - MCMC, dentre os quais se destaca a Amostragem de Gibbs (Gibbs Sampling - GS), pode ser utilizada como ferramenta, de forma a propiciar inferência bayesiana. GS é uma técnica de integração numérica por simulação, muito usual em situações nas quais a integração analítica completa é impossível. O GS é aplicável à estimação de componentes de variância que permite, por suas propriedades, a inferência bayesiana. GS é um método de utilização recente, sugerido por alguns autores para uso em contextos que envolvem o melhoramento animal ([Wang et al., 1993, 1994](#); [Jensen et al., 1994](#); [Sorensen et al., 1994](#); [Van Tassel & Van Vleck, 1996](#); [Lôbo et al., 1997](#); [Magnabosco et al., 1997, 2000](#)).

O algoritmo GS é usado para geração de amostras aleatórias via processo de reamostragem dessas variáveis aleatórias para um parâmetro θ , a partir do conjunto de dados \mathbf{y} , obtendo assim as distribuições posteriores dos parâmetros, sendo conhecidos os dados $\mathbf{p}(\theta | \mathbf{Y})$.

O objetivo deste estudo foi propor a aplicação e a utilização do método da Amostragem de Gibbs (GS) em problemas que envolvem estimativas de parâmetros genéticos, nas situações práticas do melhoramento genético animal, como é o caso das análises de dados de campo, referentes a bovinos da raça Nelore, que estudam a contribuição dos efeitos genéticos direto e maternal, bem como a correlação genética entre esses, para características de crescimento. O método da Máxima Verossimilhança Restrita por Processo Não-derivativo (DFREML) também foi utilizado para estimação dos componentes de co-variância e parâmetros genéticos.

Material e Métodos

Descrição dos dados

Para implementação do Método da Amostragem de Gibbs na estimação dos componentes de co-variância, foram utilizados dados de campo obtidos de rebanhos da raça Nelore, provenientes do Controle de Desenvolvimento Ponderal da Associação Brasileira de Criadores de Zebu - ABCZ em que foram analisados 29.245 registros de pesos de animais nascidos entre 1980 e 1993, oriundos de 24 rebanhos localizados nas Regiões Sudeste e Centro-Oeste do

Brasil onde esses animais foram criados exclusivamente em regime de pasto. Para as análises, foram consideradas três características de crescimento: peso aos 205 (P205), 365 (P365) e 550 (P550) dias de idade. As distribuições dos dados por característica estão apresentadas na Tabela 1.

Tabela 1. Número de animais (n), média, coeficiente de variação (CV), número de touros (NT) e número de vacas (NV) por característica, para as análises univariadas e bivariadas.

Característica	n	Média (kg)	CV (%)	NT	NV
P205	29.245	162,9	15,7	657	12.151
P365	18.897	216,3	17,1	584	9.562
P550	13.160	277,5	15,8	527	7.455

Análise dos dados

As análises, considerando cada característica separadamente (univariada), bem como as análises conjuntas de características aos pares (bivariadas), foram realizadas com base no modelo animal, descrito pela equação:

$$y = X\beta + Z_1a + Z_2m + Z_3p + e \quad (1)$$

No modelo linear utilizado, incluíram-se os efeitos fixos de grupos de contemporâneos (rebanho-ano-estação de nascimento) e idade da vaca ao parto (seis categorias). Nesse modelo também incluíram-se, como efeitos aleatórios, as contribuições dos efeitos genéticos aditivos direto e maternal e de ambiente permanente. Nesse modelo, **a** representa o vetor dos efeitos genéticos aditivos diretos, **m** o vetor das contribuições maternas (efeito genético aditivo maternal), **p** o vetor das contribuições de ambiente permanente e **e** o vetor de resíduos.

A estrutura das variâncias e co-variâncias para os efeitos aleatórios, utilizando uma única característica em modelo animal é:

Var

$$\begin{bmatrix} a \\ m \\ p \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & A\sigma_{am} & 0 & 0 \\ A\sigma_{am} & A\sigma_m^2 & 0 & 0 \\ 0 & 0 & I_d\sigma_p^2 & 0 \\ 0 & 0 & 0 & I_N\sigma_e^2 \end{bmatrix}$$

onde:

σ_a^2 = variância genética aditiva direta;

σ_m^2 = variância genética aditiva maternal;

σ_{am} = co-variância genética aditiva entre os efeitos direto e maternal;

A = matriz dos coeficientes de parentesco entre os animais;

σ_p^2 = variância de ambiente permanente;

σ_e^2 = variância residual;

I = matriz identidade de ordem apropriada;

d = número de vacas e

N = número de animais com dados.

Para estimação dos componentes de co-variância via Amostragem de Gibbs, β ($k \times 1$) é um vetor de efeitos "fixos" o qual, do ponto de vista bayesiano é, na realidade, um vetor de efeitos aleatórios, com distribuições iniciais não-informativas. Dessa forma, o número total de parâmetros de localização é: $r = k + 2r_1 + r_2$. A distribuição condicional do vetor de observações y , dados os parâmetros de localização e os de dispersão, pode ser apresentado como:

$$(y \mid \beta, a, m, p, \sigma^2) \sim \text{MVN} [X\beta + Z_1a + Z_2m + Z_3p, I_n\sigma^2] \quad (2)$$

onde σ^2 é a variância dos efeitos residuais. Para os outros efeitos aleatórios no modelo, assumiu-se que:

$$\begin{bmatrix} a \\ m \\ p \end{bmatrix} \sim N[0, \Sigma]$$

onde:

$$\Sigma = \begin{bmatrix} (G_0 \otimes A) & 0 \\ 0 & 0 \\ 0 & 0 & I\sigma_p^2 \end{bmatrix}$$

em que σ_p^2 é a variância dos efeitos de ambiente permanente, G_0 é uma matriz 2 x 2 das co-variâncias genéticas aditivas direta e maternal, A é a matriz de coeficientes de parentesco e \otimes é o operador produto de Kronecker. Para as análises bivariadas, os subscritos 1 e 2 referem-se às características analisadas, simultaneamente, e o modelo pode ser representado a seguir:

$$E \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1\beta \\ X_2\beta \end{bmatrix}$$

$$\text{Var} \begin{bmatrix} a_1 \\ m_1 \\ p_1 \\ e_1 \\ a_2 \\ m_2 \\ p_2 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma_{a_1}^2 & A\sigma_{a_1m_1} & 0 & 0 & A\sigma_{a_1a_2} & A\sigma_{a_1m_2} & 0 & 0 \\ A\sigma_{a_1m_1} & A\sigma_{m_1}^2 & 0 & 0 & A\sigma_{a_2m_1} & A\sigma_{m_1m_2} & 0 & 0 \\ 0 & 0 & I_d\sigma_{p_1}^2 & 0 & 0 & 0 & I_N\sigma_{p_1p_2} & 0 \\ 0 & 0 & 0 & I_N\sigma_{e_1}^2 & 0 & 0 & 0 & I_N\sigma_{e_1e_2} \\ A\sigma_{a_1a_2} & A\sigma_{a_2m_1} & 0 & 0 & A\sigma_{a_2}^2 & A\sigma_{a_2m_2} & 0 & 0 \\ A\sigma_{a_1m_2} & A\sigma_{m_1m_2} & 0 & 0 & A\sigma_{a_2m_2} & A\sigma_{m_2}^2 & 0 & 0 \\ 0 & 0 & I_N\sigma_{p_1p_2} & 0 & 0 & 0 & I_d\sigma_{p_2}^2 & 0 \\ 0 & 0 & 0 & I_N\sigma_{e_1e_2} & 0 & 0 & 0 & I_N\sigma_{e_2}^2 \end{bmatrix}$$

onde:

$\sigma_{a_1}^2$ = variância genética aditiva direta para a característica 1;

$\sigma_{m_1}^2$ = variância genética aditiva materna para a característica 1;

σ_{am} = co-variância genética aditiva entre os efeitos direto e maternal;

$\sigma_{a_2}^2$ = variância genética aditiva direta para a característica 2;

$\sigma_{m_2}^2$ = variância genética aditiva materna para a característica 2;

$\sigma_{a_1a_2}$ = co-variância genética aditiva entre os efeitos diretos para as características 1 e 2;

$\sigma_{m_1m_2}$ = co-variância genética aditiva entre os efeitos maternais para as características 1 e 2;

A = matriz de coeficientes de parentesco;

σ_p^2 = variância dos efeitos de ambiente permanente;

σ_e^2 = variância residual;

$\sigma_{p_1 p_2}$ = co-variância entre os efeitos permanentes para as características 1 e 2;

$\sigma_{e_1 e_2}$ = co-variância entre os efeitos residuais para as características 1 e 2.

O algoritmo DFREML ([Smith & Graser, 1986](#)) foi utilizado para estimar os componentes de co-variância, bem como as herdabilidades direta e maternal, empregando o Método da Máxima Verossimilhança Restrita com Processo Não-derivativo para Características Múltiplas (MTDFREML) descrito por [Boldman et al. \(1995\)](#). Para avaliar a eficiência desse método, os procedimentos para estimação o REML seguiram três estratégias, usando:

- (1) a média das estimativas para cada componente de co-variância, obtidos da literatura, como valores iniciais nas análises, considerando uma única característica e duas delas simultaneamente;
- (2) as estimativas das médias posteriores de cada componente de variância, resultantes do procedimento de amostragem de Gibbs, obtidos do *software* MTGSAM, como valores iniciais (análise de uma única característica); e,
- (3) os valores iniciais aleatórios para cada componente de co-variância (análise de uma única característica).

O critério de convergência adotado foi que a variância dos valores da função $(-2\log L)$ do algoritmo simplex fosse menor que 10^{-9} .

Para as análises que consideraram o método GS para estimação dos componentes de co-variância, o *software* MTGSAM desenvolvido por [Van Tassel & Van Vleck \(1996\)](#) foi adaptado e adotado para calcular as distribuições médias posteriores por meio de amostras e estimativas posteriores dos componentes de co-variância, geradas pela cadeia de Gibbs.

Os procedimentos para estimação, empregando-se o método da Amostragem de Gibbs (para características univariadas), seguiram três estratégias as quais utilizaram:

- (1) a média das estimativas para cada componente de co-variância, obtidos da literatura, como valores iniciais;
- (2) as estimativas dos componentes de co-variância como iniciais, obtidos do MTDFREML; e,
- (3) os valores iniciais arbitrários para cada componente de co-variância.

Distribuições iniciais e posteriores

A distribuição inicial cuja variância tende ao infinito para os efeitos “fixos” significa a suposição de que não há nenhum conhecimento inicial ou a priori para esses efeitos, sendo $p(\beta) \propto \text{constante}$. Supõe-se que os efeitos aleatórios tenham distribuição normal multivariada, como na equação [2] e, especialmente, para os efeitos genéticos é necessário que uma estrutura de co-variância entre esses efeitos genéticos correspondentes à matriz de parentesco seja considerada. Assumiu-se para os efeitos residuais uma distribuição normal. Usando esses pressupostos, as equações de modelo misto produzem soluções BLUE e BLUP para os efeitos fixos e aleatórios respectivamente se as variâncias são conhecidas ([Gianola & Fernando, 1986](#); [Gianola et al., 1990](#)). Os efeitos genéticos aditivos direto e maternal são assumidos como tendo co-variâncias diferentes de zero podendo ser estimadas para todas as características ou combinações de características. Não foram consideradas as possíveis co-variâncias entre os efeitos de ambiente permanente.

Distribuições de Wishart invertidas (IW) foram usadas como sendo as distribuições para cálculo dos componentes de co-variância, principalmente, por questões computacionais. As distribuições iniciais IW para os componentes de co-variância foram consideradas mutuamente independentes. A densidade de Wishart descreve a distribuição das somas de quadrados e produtos cruzados das variáveis aleatórias com distribuição normal. Em casos de modelos unicarateres, como o usado neste estudo, isto corresponde a uma Distribuição de Qui -Quadrado Invertida.

Seja T uma variável aleatória com distribuição IW. Sua função de densidade é dada por:

$$f(T | T_0, \nu) \propto |T|^{-1/2(\nu + m + 1)} \times \exp[\text{tr}(-1/2 T_0^{-1} T^{-1})] \quad (3)$$

com T_0 , matriz de parâmetro de escala de ordem $m \times m$, positiva definida e ν , o parâmetro de escala, associado ao grau de confiança e que dá a forma da distribuição dos componentes. No aplicativo MTGSAM, T_0 é calculada de tal forma que o valor esperado da distribuição inicial da matriz de co-variância seja igual ao valor fornecido pelo usuário.

Uma descrição muito bem detalhada da distribuição condicional completa e distribuição posterior conjunta, necessárias para implementação de GS em

modelos lineares mistos pode ser encontrada em [Wang et al. \(1993\)](#); [Jensen et al. \(1994\)](#); [Van Tassel et al. \(1995\)](#) e [Sorensen \(1996\)](#).

Neste estudo, as distribuições condicionais completas (DCC), para cada parâmetro foram obtidas da densidade posterior conjunta (DPC), ou seja, da densidade dos parâmetros considerando os dados e as informações iniciais. Assim a DPC pode ser escrita como sendo o produto de todas as distribuições iniciais e a função de verossimilhança ([Van Tassel & Van Vleck, 1996](#)). As DCC são necessárias para cada elemento ou conjunto de elementos do vetor de parâmetros a ser estimado. A DCC corresponde à distribuição de um parâmetro específico, sendo os outros valores dos parâmetros remanescentes, no modelo, considerados conhecidos ou fixados.

A DCC de cada parâmetro pode ser derivada da DPC, ao se ignorar os termos não-envolvidos no parâmetro de interesse. Considera-se, dessa forma, que os parâmetros, cujos valores são conhecidos, sejam considerados como constantes.

Fazendo uso das propriedades do núcleo de uma distribuição que é a parte da função que contém a variável aleatória de interesse e que permanece enquanto todas as constantes são descartadas, reorganizam-se os termos, permanecendo somente os parâmetros de interesse. Esse método de núcleo apresenta as vantagens de usar as informações contidas nos dados de maneira bastante eficiente, porém, apresenta a desvantagem de necessitar de informações extras e que têm de ser salvas a cada nova reamostragem. Outra desvantagem ([Jensen et al., 1994](#)) é que a densidade condicional completa das variâncias genéticas e dos componentes de variância não podem ser escritas, a menos que a co-variância entre os parâmetros genéticos seja zero. Entretanto, o uso eficiente da informação, em cada amostra individual, não deve ser considerado de tanta importância na Amostragem de Gibbs, já que é fácil obter maior número de amostras.

As DCC são fundamentais na descrição da implementação do método GS no MTGSAM, no entanto, a descrição a seguir, adaptada de [Van Tassel & Van Vleck \(1996\)](#), é geral visando a considerar análises multicarateres. Apesar de, neste estudo, as análises que empregam o método GS terem sido unicarater, as considerações são mantidas na forma empregada no *software* MTGSAM. Dessa maneira, para análises multicarateres a matriz Σ pode ser apresentada como:

$$\Sigma = \begin{bmatrix} (\mathbf{G} \otimes \mathbf{A}) & \mathbf{0} \\ \mathbf{0} & \bigoplus_{i=1}^{\gamma} [\mathbf{D}_i \otimes \mathbf{I}_{n_i}] \end{bmatrix}$$

em que,

\otimes e \oplus são os operadores de produto direto e da soma de matrizes direta, respectivamente. Seja $\Sigma_0 = \mathbf{G} \otimes \mathbf{A}$ e $\Sigma_i = \mathbf{D}_i \otimes \mathbf{I}_{n_i}$; dessa forma, Σ pode também ser escrita como $\Sigma = \bigoplus_{i=1}^{\gamma}$.

Um mesmo animal pode apresentar várias observações, então os efeitos residuais são assumidos como sendo não correlacionados entre os diferentes registros do mesmo animal, embora os efeitos residuais, em um mesmo registro, possam ser correlacionados. O vetor \mathbf{y} é ordenado por característica dentro de registro e animal, dessa forma \mathbf{R} é uma matriz bloco-diagonal.

$$\mathbf{R} = \bigoplus_{j=1}^{\rho} \mathbf{R}_j^*$$

em que \mathbf{R}_j^* , é uma matriz de co-variâncias ($r_j \times r_j$) residuais para as características medidas no registro j , q é o número de observações, e r_j , o número de características observadas para o registro j . Note que para um animal com todas as características medidas $\mathbf{R} = \bigoplus_{i=1}^{\rho} \mathbf{R}_i^*$, em que \mathbf{R}_i é a matriz de co-variâncias ($t_i \times t_i$) residuais para o grupo i dos efeitos residuais, t_i é o número de características no grupo residual i e ρ é o número de grupos dos efeitos residuais.

A matriz \mathbf{G} é a conhecida matriz ($d_0 \times d_0$) que descreve as co-variâncias genéticas para um animal em que d_0 é o número de características mais o número de características com efeitos genéticos correlacionados.

Seja, $\mathbf{u}' = [\mathbf{u}_0' \mathbf{u}_1' \dots \mathbf{u}_\gamma']$, em que \mathbf{u}_0 representa os efeitos genéticos aditivos, \mathbf{u}_1 os efeitos aleatórios não correlacionados para o grupo i , para $i > 0$, e γ o número de grupos dos efeitos aleatórios não-correlacionados. O vetor \mathbf{u}_0 é

ordenado por animal dentro de característica e estas ordenadas com os efeitos genéticos diretos de maneira que cada uma esteja associada aos efeitos genéticos correlacionados para as características que os possuam.

O vetor u_{γ} corresponde aos efeitos aleatórios que não são correlacionados com os efeitos genéticos e outros efeitos aleatórios (i.e. efeito de ambiente permanente). Os efeitos são ordenados por nível dentro de características que incluam aquele efeito. Dessa maneira, D_i é a matriz ($d_i \times d_i$) de co-variâncias entre os efeitos aleatórios entre características no grupo i para um animal, onde d_{ia} é o número de efeitos aleatórios não correlacionados no grupo i , ou seja, o número de características representadas no grupo i . O número de níveis em cada um dos d_{ia} efeitos não correlacionados no grupo i é representado por N_i .

É importante lembrar que as matrizes D e R dependem fundamentalmente da combinação de características para as quais o animal dispõe de dados.

Com base nas DCC, podem ser obtidas as distribuições condicionais para os efeitos fixos, genéticos de ambiente permanente e residuais. A distribuição condicional dos efeitos fixos e aleatórios, $s' = [\beta' u']$, foi obtida primeiramente, sendo escrita depois do uso do núcleo de uma densidade normal da seguinte forma:

$$s \mid G, D_1, D_2, \dots, D_{\gamma}, R, y \sim N(s, C^{-1}) \quad (4)$$

em que,

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{11} \end{bmatrix} e,$$

C^{-1} é a inversa da matriz de coeficientes das equações de modelos mistos.

Utilizando em [4] uma distribuição condicional normal multivariada, a DCC para um escalar ou subvetor de elementos s_1 de s , onde $s' = [s_1' s_2']$ é descrita como:

$$s_1 \mid s_2, G, D_1, D_2, \dots, D_{\gamma}, R, y \sim N(C_{11}^{-1} (r_1 - C_{12} s_2), C_{11}^{-1}) \quad (5)$$

Aplicando-se as propriedades de uma distribuição condicional normal para a DCC da equação [4], obtém-se a DCC para os efeitos genéticos:

$$s_1 | s_{-1}, \mathbf{G}, \mathbf{D}_1 \mathbf{D}_2 \dots \mathbf{D}_\gamma, \mathbf{R}, \mathbf{y} \sim N(S_1, (P_1 \mathbf{C} P_1')^{-1}) \quad (6)$$

em que $S_1 = (P_1 \mathbf{C} P_1')^{-1} (P_1 \mathbf{r} - P_1 \mathbf{C} P_1' \mathbf{S}_{-1})$; sendo P_1 uma matriz de permutação.

Para se especificar a distribuição condicional completa dos efeitos residuais para observações realizadas (e_o) e dos resíduos para observações perdidas (e_m), assume-se que a partição do vetor de resíduos seja:

$$\mathbf{e}_{ij} = \begin{bmatrix} \mathbf{e}_m \\ \mathbf{e}_o \end{bmatrix}$$

em que, \mathbf{e}_m é o subvetor dos efeitos residuais de observações perdidas e \mathbf{e}_o é o subvetor dos resíduos para características com observações. Para o grupo I dos efeitos residuais, os resíduos são distribuídos:

$$\mathbf{e}_i | \mathbf{R}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{R}_i)$$

Seja

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{mm} & \mathbf{R}_{mo} \\ \mathbf{R}_{om} & \mathbf{R}_{oo} \end{bmatrix}$$

A distribuição condicional completa dos efeitos residuais para observações realizadas (e_o) e dos resíduos para observações perdidas (e_m), pode ser apresentada como uma distribuição condicional normal, na mesma forma de [6], escrita como:

$$\mathbf{e}_m | \mathbf{e}_o, \mathbf{s}, \mathbf{G}, \mathbf{D}_1 \mathbf{D}_2 \dots \mathbf{D}_\gamma, \mathbf{R}, \mathbf{y} \sim \mathbf{N}(\mathbf{R}_{mo} \mathbf{R}_{oo}^{-1} \mathbf{e}_o, \mathbf{R}_{mm} - \mathbf{R}_{mo}^{-1} \mathbf{R}_{om}) \quad (7)$$

em que \mathbf{R}_{oo} , \mathbf{R}_{mm} , \mathbf{R}_{mo} e \mathbf{R}_{om} são as submatrizes da matriz de co-variância residual \mathbf{R} , referentes às observações realizadas e perdidas.

Para as matrizes de co-variância utilizadas, as distribuições condicionais completas foram obtidas das derivadas da distribuição posterior conjunta ([Van Tassel & Van Vleck, 1996](#)). Utilizando-se de propriedade do núcleo de uma densidade de IW, as DCC para as matrizes de co-variância genéticas, efeitos aleatórios não-correlacionados e dos efeitos residuais serão respectivamente:

$$\mathbf{G} \mid \mathbf{s}, \mathbf{G}_0, v_g, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_v, \mathbf{R}, \mathbf{y} \sim \text{IW} \left(\left(v_g^* \mathbf{G}_0 + \mathbf{S}_0 \right)^{-1}, n + v_g \right), \quad (8)$$

$$\mathbf{D}_i \mid \mathbf{s}, \mathbf{D}_{0_i}, v_{d_i}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_{i-1}, \mathbf{D}_{i+1}, \dots, \mathbf{D}_v, \mathbf{R}, \mathbf{y} \sim \text{IW} \left(\left(v_{d_i}^* \mathbf{D}_{0_i} + \mathbf{S}_i \right)^{-1}, n_i + v_{d_i} \right) \quad (9)$$

$$\mathbf{R}_i \mid \mathbf{e}, \mathbf{R}_{0_i}, v_{r_i}, \mathbf{s}, \mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_v, \mathbf{R}_{i-1}, \mathbf{R}_{i+1}, \dots, \mathbf{R}_v, \mathbf{y} \sim \text{IW} \left(\left(v_{r_i}^* \mathbf{R}_{0_i} + \mathbf{Q}_i \right)^{-1}, q_i + v_{r_i} \right) \quad (10)$$

Sejam $\mathbf{G}_0(v_g)$, $\mathbf{D}_{0_i}(v_{d_i})$, $\mathbf{eR}_{0_{ij}}(v_{r_i})$ as matrizes de escala (forma) dos efeitos genéticos, do i -ésimo grupo de efeito aleatório não correlacionado e do j -ésimo grupo de co-variâncias residuais, respectivamente, e v_g, v_{d_i}, v_{r_i} são seus respectivos graus de confiança em que:

$$v_g^* = v_g - \mathbf{d}_0 - 1, \quad v_{d_i}^* = v_{d_i} - \mathbf{d}_i - 1 \quad \text{e} \quad v_{r_i}^* = v_{r_i} - \mathbf{t}_i - 1$$

No MTGSAM, a matriz de escala é calculada de tal forma que o valor esperado da distribuição *a priori* da matriz de co-variância seja igual ao valor fornecido pelo usuário, e.g., no caso da matriz de co-variância genética, se o valor médio fornecido é \mathbf{G}^* , tem-se que $\mathbf{G}_0^{-1} = v_g^* \mathbf{G}^*$.

Outro aspecto importante nas análises com o MTGSAM é a necessidade de se especificar os parâmetros que determinam a forma das distribuições iniciais assumidas para cada parâmetro estudado. Neste estudo, análises foram efetuadas assumindo três formas de distribuições iniciais: distribuição *flat* ou não-informativa (não refletindo o conhecimento prévio do parâmetro), simétrica (permite alguma informação do parâmetro) e aguda (*sharp*) ou informativa (refletindo conhecimento do parâmetro). Os parâmetros de definição da forma (v) da distribuição inicial foram 0, 5 e 9, respectivamente, conforme requerido pelo MTGSAM.

Algoritmo de Gibbs e MTGSAM

O software MTGSAM ("*Multiple Trait Gibbs Sampling in Animal Models*"), desenvolvido por [Van Tassel et al. \(1995\)](#) e [Van Tassel & Van Vleck \(1996\)](#) é um conjunto de programas em Fortran para a estimação de componentes de co-variância em modelos animais usando algoritmo de Amostragem de Gibbs

(GS). Sua utilização é muito similar ao *software* MTDFREML amplamente utilizado no melhoramento animal, o que é explicado pelo fato de ambos, MTGSAM e MTDFREML, compartilharem vários programas e sub-rotinas. Para facilitar o entendimento do algoritmo GS e como ser implementado, [Uribe \(1996\)](#) apresentou uma maneira simplificada de seu funcionamento. Considere o modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

em que $\boldsymbol{\beta}$ é o vetor dos efeitos fixos que é associado por meio da matriz de incidência \mathbf{X} com as observações em \mathbf{y} ; \mathbf{u} é o vetor dos efeitos aleatórios, associado a \mathbf{y} por meio da matriz de incidência \mathbf{Z} ; e \mathbf{e} o vetor dos efeitos residuais (aleatórios). Efetuar os seguintes passos:

1. Fornecer os valores iniciais dos parâmetros de dispersão e localização do modelo com base em um sistema de equações de quadrados mínimos ordinário de um conjunto de dados qualquer.
2. Gerar inicialmente os efeitos fixos.
3. Gerar os efeitos aleatórios.
4. Gerar a soma de quadrados do resíduo e variância residual. Assume-se que a distribuição inicial definida para a variância residual seja uma distribuição de qui-quadrado invertida.
5. Gerar a variância para os efeitos aleatórios. Igualmente, ao passo quatro, assume-se que a distribuição inicial definida para a variância para os efeitos aleatórios seja uma distribuição de qui-quadrado invertida.
6. Calcular a nova relação de variância.

Uma vez que os parâmetros tenham sido criados, o procedimento iterativo é repetido tantas vezes quantas forem necessárias, criando uma cadeia de amostras aleatórias da distribuição posterior marginal. As estimativas finais dos parâmetros serão a média ou a moda das amostras consideradas.

Neste estudo, a aplicação do procedimento GS é mais complexa que a descrita acima. Considerando o modelo animal completo, utilizado nos dados de campo, essa aplicação deve apresentar os seguintes passos:

1. Fornecer todos os valores para os componentes de co-variância. Esses valores correspondem às médias das distribuições iniciais quando se usam valores iniciais informativos para as co-variâncias.
2. Utilizar os dados e os componentes de co-variância iniciais fornecidos para montar as equações de modelo misto ([Henderson, 1973](#)) e calcular as soluções dos efeitos fixos (b) e aleatórios (g, d), pelo método iterativo de Gauss-Seidel. Devem ser especificados o número de iterações e o critério de convergência adotado.
3. Oferecer novas soluções para os efeitos fixos, baseado na equação [5].
4. Indicar novas soluções para os valores genéticos aditivos e maternos, e para os efeitos aleatórios não correlacionados bloqueados, baseado na equação [6].
5. Fornecer novas soluções para os efeitos aleatórios não-correlacionados não-bloqueados, baseado na equação [5].
6. Calcular os efeitos residuais para características com observações (e_o) e produzir resíduos perdidos (e_m) com base na equação [7].
7. Calcular as formas quadráticas S_o , para os efeitos genéticos, com base em $u_{0_i}'' A^{-1} u_{0_j}$.
8. Produzir uma matriz G baseada na distribuição Whishart invertida da equação [8].
9. Calcular as formas quadráticas S_i para os efeitos aleatórios não-correlacionados, com base em $u_{j_j}'' u_{i_k}$.
10. Produzir cada D_i baseado na distribuição Whishart invertida da equação [9].
11. Calcular as formas quadráticas Q_i para os efeitos residuais com base em $e_{j_j}'' e_{i_k}$.
12. Produzir cada R_i baseado na distribuição Whishart invertida da equação [10].
13. Proceder à repetição dos passos três ao doze, tantas vezes quanto o tamanho de Gibbs estabelecido em um processo de reamostragem. Dessa maneira, a execução dos passos três a doze formarão um ciclo de Gibbs.

Neste trabalho, o MTGSAM foi usado para calcular as médias posteriores dos componentes de co-variância obtidas das amostras da cadeia de Gibbs, bem como para a estimação de correlações e herdabilidades.

As estimativas das médias posteriores para os componentes de co-variância foram baseadas no valor esperado das distribuições IW para todas as variáveis aleatórias ([Van Tassel & Van Vleck, 1996](#)). Dessa maneira, podem-se obter das equações [8], [9] e [10] os valores esperados para \mathbf{G} , \mathbf{D}_i e \mathbf{R}_i , considerando qualquer número de ciclos de amostragem de Gibbs conforme calculado:

$$E(\mathbf{G} \mid \mathbf{S}_0, \mathbf{G}_0, n_g) = \frac{v_g^* \mathbf{G}_0 + \mathbf{S}_0}{(n + v_g - d_0 - 1)}$$

$$E(\mathbf{D}_i \mid \mathbf{S}, \mathbf{D}_{oi}, n_{di}) = \frac{v_{d_i}^* \mathbf{D}_{oi} + \mathbf{S}_i}{(n + v_{d_i} - d_i - 1)}$$

$$E(\mathbf{R}_i \mid \mathbf{Q}, \mathbf{R}_{oi}, n_{ri}) = \frac{v_{r_i}^* \mathbf{R}_{oi} + \mathbf{Q}_i}{(q_i + v_{r_i} - t_i - 1)}$$

Tem-se, então, que a média de um componente de co-variância é calculada como a média dos valores esperados em toda a cadeia, depois do período de descarte das amostras iniciais. As estimativas das médias posteriores para as herdabilidades e correlações são calculadas como as médias das funções calculadas, usando os componentes de co-variância em cada ciclo de GS depois do período de descarte das amostras iniciais.

Implementação da Amostragem de Gibbs

A implementação do algoritmo de Gibbs requer alguns cuidados especiais com a definição das análises, como o critério de convergência, período de descarte amostral, intervalo amostral e tamanho de cadeia amostral.

Critério de convergência

O MTGSAM utiliza o método Iterativo de Gauss Seidel (IGS) sobre as equações de Modelos Mistos para obter os valores iniciais dos efeitos fixos e aleatórios para iniciar a amostragem de Gibbs. O método IGS determina a solução de um sistema linear por aproximações, usando operações aritméticas a partir de dado vetor de aproximação inicial ([Lopes et al., 1993](#)). O MTGSAM utiliza valores iniciais iguais a zero para todos os efeitos fixos e aleatórios. As iterações são

executadas até que um número máximo de iterações estabelecido ou o critério de convergência seja alcançado. O critério de convergência usado foi:

$$\frac{\sum_{i=1}^n |s_i^j - s_i^{j-1}|^2}{\sum_{i=1}^n |s_i^j|^2} < c$$

em que n é o número de equações, s_i^j é a solução para a equação i na iteração j , e c é um valor estabelecido pelo usuário. Neste estudo, o valor do critério de convergência usado no método IGS foi de $c = 10^{-12}$. O número de ciclos alcançados até a convergência variou de 40 a 80.

Período de descarte amostral

Em termos práticos, o período de descarte amostral é o número de ciclos que deve ser descartado antes que as amostras produzidas pelo GS possam ser consideradas amostras das distribuições posteriores, ou seja, a convergência ou distribuição tenha ocorrido. Programas como o GIBBSIT ([Raftery & Lewis, 1994](#)) podem ser usados para determinação do período ideal de descarte para cada conjunto de dados, porém, neste estudo, esse período foi avaliado subjetivamente, sendo descartadas amostras geradas dos primeiros 200 ciclos de GS. Análises adicionais foram feitas utilizando um período de descarte de 2000 ciclos.

Intervalo amostral

A estimativa de parâmetros com vistas a uma distribuição requer uma amostra de observações independentes. Amostras consecutivas, obtidas de uma cadeia GS, podem ser altamente correlacionadas, dessa forma, para se obterem amostras independentes ou não correlacionadas é necessário que se eliminem resultados de ciclos intermediários da cadeia GS. O intervalo de retirada das amostras é denominado intervalo de utilização amostral e deve ser suficiente para que as amostras utilizadas não sejam correlacionadas. Para se determinar o intervalo de utilização amostral, podem ser usados programas como o GIBBSIT ([Raftery & Lewis, 1994](#)) e GIBBANAL ([Van Kaam, 1997](#)). Trabalhos anteriores ([Van Tassel et al., 1995](#)) sugeriram que, quando foram utilizadas as amostras obtidas do vigésimo ciclo e a cada 20 ciclos, depois do período de descarte, a correlação entre as amostras consecutivas foi menor que 0,05. Neste estudo, o intervalo amostral foi de 30 ciclos e 200 ciclos.

Tamanho da cadeia de Gibbs

O algoritmo de GS produz uma cadeia de amostras de Gibbs de tantos ciclos quantos forem especificados. Esse aspecto é muito importante porque está intimamente relacionado com a convergência das distribuições posteriores dos parâmetros. Se o número de ciclos definido pelo usuário for adequado, as médias posteriores das amostras serão estimativas válidas das distribuições posteriores dos parâmetros. Para se verificar a acurácia dessas estimativas, pode-se utilizar o erro de Monte Carlo ([Sorensen et al., 1994](#)). Esse erro é devido ao número de amostras usado da cadeia de Gibbs, sendo, portanto, inversamente proporcional ao tamanho da cadeia.

As análises foram conduzidas utilizando um tamanho de cadeia amostral de 5000 ciclos, com um período de descarte amostral de 200 ciclos e intervalo amostral de 30 ciclos, obtendo nessas condições, 160 amostras das médias posteriores dos componentes de co-variância.

Em relação ao tamanho da cadeia amostral, neste estudo, verificaram-se três tamanhos de cadeia: 5000, 10.000 e 15.000 ciclos. Em outra análise na qual se aplicou o esquema de cadeia longa, também foram avaliados tamanhos de cadeia amostral de 100.000 e 200.000 ciclos, com um período de descarte amostral de 2000 ciclos e um total de 490 e 990 amostras dos parâmetros do modelo linear, respectivamente, retiradas de um intervalo amostral de 200 ciclos.

Resultados e Discussão

Estimação dos componentes de co-variância utilizando o REML

As estimativas dos componentes de co-variância para os efeitos direto, maternal e de ambiente permanente obtidas das análises univariadas e bivariadas estão apresentadas na [Tabela 2](#), bem como as herdabilidades direta, maternal e total para os pesos aos 205, 365 e 550 dias de idade. Os valores obtidos das análises univariadas para herdabilidade direta foram inferiores aos alcançados em análises bivariadas, para pesos aos 365 e 550 dias de idade. Entretanto, para herdabilidade maternal não foi observada praticamente nenhuma alteração dos valores obtidos pelas duas análises para todas as características.

Tabela 2. Estimativas dos componentes de co-variância e parâmetros genéticos para pesos aos 205 (P205), 365 (P365) e 550 dias (P550) com base nas análises univariadas (*) e bivariadas (+) obtidas do método REML.

Característica	σ_a^2	σ_m^2	σ_{am}	σ_{pe}^2	σ_E^2	σ_p^2	h_d^2	h_m^2	h_t^2	rg_{am}	c^2
P205 *	63,10	44,75	4,01	41,92	257,66	411,46	0,15	0,11	0,22	0,08	0,10
+ P365	66,95	45,00	3,98	43,30	265,52	424,76	0,16	0,11	0,22	0,07	0,10
+ P550	63,53	45,62	6,25	39,76	261,96	417,12	0,15	0,11	0,23	0,12	0,10
P365 *	77,90	46,51	4,68	52,41	494,44	675,90	0,12	0,07	0,16	0,08	0,08
+ P205	106,68	52,86	11,01	44,87	500,93	716,86	0,15	0,07	0,21	0,15	0,06
+ P550	214,07	5,11	-0,33	51,38	480,30	750,53	0,29	0,01	0,29	-0,01	0,07
P550 *	165,90	61,70	8,47	0,20	680,54	916,66	0,18	0,07	0,23	0,08	0,00
+ P205	232,82	60,01	6,90	7,60	659,65	966,65	0,24	0,06	0,28	0,06	0,00
+ P365	411,35	0,22	-0,10	7,83	606,70	1026,00	0,40	0,00	0,39	-0,01	0,00

^a σ_a^2 : variância genética aditiva direta; σ_m^2 : variância genética aditiva materna; σ_{am} : co-variância genética entre os efeitos genéticos aditivo direto e maternos; σ_{pe}^2 : variância de ambiente permanente; σ_E^2 : variância residual; σ_p^2 : variância fenotípica; h_d^2 : herdabilidade direta; h_m^2 : herdabilidade materna; h_t^2 : herdabilidade total; rg_{am} : correlação genética entre os efeitos genéticos aditivos diretos e maternos; c^2 : contribuição de ambiente permanente.

Peso aos 205 dias

Com referência ao peso aos 205 dias, as estimativas para todos os efeitos incluídos no modelo foram similares para as análises univariadas e bivariadas, indicando que a contribuição dos pesos pós-desmame em uma análise bivariada praticamente não alterou as estimativas para o peso aos 205 dias.

As estimativas de herdabilidade direta e maternal foram inferiores aos valores encontrados na literatura ([Eler et al., 1992, 1995](#); [Reyes et al., 1994, 1995](#); [Lôbo et al., 1995](#)). Em geral, as estimativas das correlações genéticas entre os efeitos diretos e maternais foram positivas, porém de baixa magnitude. Um programa de seleção que considera somente os efeitos diretos para peso não resultará necessariamente no aumento da habilidade materna. Resultados semelhantes foram descritos por [Trovo et al. \(1996\)](#), outros autores ([Koots et al., 1994](#); [Eler et al., 1995](#); [Lôbo et al., 1995](#); [Reyes et al., 1995](#); [Magnabosco et al., 1996](#); [Robinson, 1996](#)) encontraram uma magnitude negativa baixa para as estimativas de correlação genética entre os efeitos diretos e maternais. A contribuição do efeito de ambiente permanente representou 10%, indicando a necessidade da inclusão desse efeito no modelo nas análises para características de peso até aos 205 dias.

Pesos pós-desmames

Os componentes de variância para efeito direto dos pesos pós-desmame (P365 e P550) tiveram magnitude maior quando obtidos das análises bivariadas com o peso aos 205 dias. Esse fato pode ser explicado porque nas análises bivariadas foi levado em conta o descarte que ocorre, geralmente, depois da desmama dos bezerros. Dessa maneira, essas maiores magnitudes dos componentes de variância para os efeitos diretos constituem indicativo mais correto da variabilidade genética presente nos pesos pós-desmame.

Adicionalmente, as estimativas dos componentes de co-variância para P365 e P550 foram distintas quando obtidas de análises univariada e bivariada, considerando os efeitos diretos e maternais. Os resultados ([Tabela 2](#)) confirmam a necessidade de se considerar todas as relações existentes entre os pesos estudados. Caso contrário, podem-se subestimar os parâmetros utilizados na definição de critérios de seleção para peso na raça Nelore depois do desmame.

As correlações genéticas entre os efeitos diretos para P205, P365 e P550 foram uniformemente altos, variando de 0,87 a 0,98. Igualmente, as correlações genéticas entre os efeitos maternos para P205, P365 e P550 variaram de 0,57 a 0,96. Em contraste, as correlações genéticas entre os efeitos diretos e maternos para essas características foram menores, variando de 0,03 a 0,23. Alguma inconsistência foi observada nas estimativas das correlações genéticas entre os efeitos diretos e maternos, nas análises bivariadas para os pesos P365 e P550 dias simultaneamente com o P205, observando-se, inclusive, mudança de sinal. Isso pode ser explicado pela estrutura de dados que não permitiu análise com os efeitos maternos em análises bivariadas envolvendo P550, ou ainda, devido ao descarte de animais ocorrido depois da desmama e, posteriormente, depois de um ano, sendo o número de observações bastante reduzido, prejudicando a distribuição dos dados. Um estudo completo relativo à estrutura dos dados e aos modelos impróprios foram discutidos por [Meyer \(1994\)](#).

A magnitude dos componentes de variância para o efeito aditivo maternal ([Tabela 2](#)), bem como os valores observados para as correlações genéticas entre os efeitos diretos e maternos, obtidos das análises bivariadas para os pesos aos 205, 365 e 550 dias não justificaram a inclusão do efeito maternal para análises dos pesos aos 365 e 550 dias, porém, esse efeito deve ser considerado para P205. Esses resultados corroboram os relatos de [Meyer \(1994\)](#); [Eler et al. \(1995\)](#); [Reyes et al. \(1995\)](#); [Magnabosco et al. \(1996\)](#); [Robinson \(1996\)](#).

Valores iniciais usando REML

O objetivo dessas análises univariadas suplementares foi verificar se havia modificação das estimativas obtidas do método REML, em razão da alteração dos valores iniciais das estimativas fornecidas. Com diferentes valores iniciais, isto é, informativos, não informativos e obtidos do GS, verificaram-se as estimativas obtidas do mesmo critério de convergência descrito anteriormente. Outra abordagem foi tentar comparar essas diferentes situações nas quais se utilizou o método REML, com as estimativas alcançadas pelo método GS. As estimativas dos componentes de co-variância e dos parâmetros genéticos para pesos aos 205 (P205), 365 (P365) e 550 dias (P550), com base nas análises univariadas obtidas do método REML, são apresentadas na [Tabela 3](#).

Tabela 3. Estimativas de componentes de co-variância e parâmetros genéticos^a para pesos aos 205 (P205), 365 (P365) e 550 dias (P550), com base nas análises univariadas obtidas do método REML, considerando diferentes estimativas iniciais.

Característica	σ_a^2	σ_m^2	σ_{am}	σ_{pe}^2	σ_E^2	σ_p^2	h_d^2	h_m^2	h_T^2	rg_{am}	c^2
P205 _I	63,10	44,75	4,01	41,92	257,66	411,46	0,15	0,11	0,22	0,08	0,10
P205 _{NI}	63,17	44,81	3,88	41,90	257,68	411,45	0,15	0,11	0,22	0,07	0,10
P205 _{GS}	63,11	44,81	3,89	41,90	257,65	411,45	0,15	0,11	0,22	0,07	0,10
P365 _I	77,90	46,51	4,68	52,41	494,44	675,90	0,12	0,07	0,16	0,08	0,08
P365 _{NI}	77,75	46,61	4,70	52,34	494,51	675,91	0,12	0,07	0,16	0,08	0,08
P365 _{GS}	77,83	46,60	4,70	52,35	494,44	675,89	0,12	0,07	0,16	0,08	0,07
P550 _I	165,90	61,70	8,47	0,20	680,54	916,66	0,18	0,07	0,23	0,08	0,00
P550 _{NI}	165,33	60,91	8,90	0,00	681,16	916,33	0,18	0,07	0,23	0,09	0,00
P550 _{GS}	165,70	61,69	8,47	0,00	680,65	916,51	0,18	0,07	0,23	0,08	0,00

^a σ_a^2 : variância genética aditiva direta; σ_m^2 : variância genética aditiva materna; σ_{am} : co-variância genética entre os efeitos genéticos diretos e maternos; σ_{pe}^2 : variância de ambiente permanente; σ_E^2 : variância residual; σ_p^2 : variância fenotípica; h_d^2 : herdabilidade direta; h_m^2 : herdabilidade materna; h_T^2 : herdabilidade total; rg_{am} : correlação genética entre os efeitos diretos e maternos; c^2 : contribuição do ambiente permanente.

I: Estimativas obtidas da literatura pelo método REML; NI: Valores iniciais obtidos de forma não sistemática; GS: Valores iniciais obtidos das estimativas das médias posteriores de cada componente de co-variância, resultantes do método de amostragem de Gibbs.

Para P205, as estimativas dos componentes de variância não foram alteradas quando se utilizaram valores iniciais não-informativos nas análises do MTDFREML. No entanto, o tempo de processamento foi cerca de 30% maior, fato normal, pois a convergência, partindo desse tipo de valores iniciais, é mais difícil de ser alcançada. Nas três situações o MTDFREML convergiu para os mesmos resultados. Quando foram utilizados, para cada componente de co-variância, valores iniciais arbitrários, observou-se que para o uso do *software* MTDFREML havia necessidade de que durante o processo iterativo se alterasse a estrutura de co-variância mais de uma vez para obter a convergência do sistema. Nesse caso, o processamento foi 200% maior do que quando os valores iniciais usados eram estimativas do GS ou valores encontrados na literatura. Para os efeitos diretos e maternais, os valores iniciais devem refletir alguma informação (pelo menos o sinal) sobre o verdadeiro valor do parâmetro estudado.

Distribuições médias posteriores dos componentes de co-variância e parâmetros genéticos usando o GS

Nas análises univariadas, utilizando-se do método GS, gastou-se cerca de quatro vezes mais tempo de processamento do que com o DFREML. Com o GS, produz-se distribuição posterior marginal completa, a partir da qual são obtidas as estimativas de parâmetros genéticos, valores genéticos etc.

A convergência da distribuição para o algoritmo GS é mais difícil de ser alcançada comparativamente aos algoritmos de verossimilhança. O problema é que, no método GS, a convergência é de uma distribuição e não somente de um valor único. Para verificar a convergência, uma alternativa é produzir amostras ou estimativas da distribuição dos parâmetros de várias cadeias e compará-las. A partir do momento em que essas amostras ou médias posteriores dessas estimativas forem aproximadamente iguais, a convergência é alcançada.

Na [Tabela 4](#), são demonstradas as estimativas de cada parâmetro, na qual se observou que não houve alterações quando foram usadas diferentes formas de distribuição inicial para as análises. As estimativas de herdabilidades diretas e maternais, bem como as correlações genéticas para P205, obtidas das médias posteriores dos componentes de co-variância com diferentes valores iniciais, foram similares a outras estimativas encontradas para a raça Nelore, também utilizando o modelo animal ([Eler et al., 1992, 1995](#); [Reyes et al., 1994, 1995](#); [Lôbo et al., 1995](#)).

Tabela 4. Estimativas de médias posteriores dos componentes de co-variância e parâmetros genéticos^a para pesos aos 205 (P205), 365 (P365) e 550 dias (P550) das análises univariadas usando a amostragem de Gibbs, com base nos valores iniciais REML e considerando três formas de distribuições iniciais^b.

Característica	σ_a^2	σ_m^2	σ_{am}	σ_{pe}^2	σ_E^2	σ_p^2	h_d^2	h_m^2	h_T^2	rg_{am}	c^2
P205 _{NI}	98,42	70,37	-18,71	36,47	238,85	425,41	0,23	0,16	0,25	-0,20	0,08
P205 _S	112,62	64,50	-21,22	39,02	232,00	426,90	0,26	0,15	0,26	-0,23	0,09
P205 _I	111,09	63,87	-20,34	39,14	232,77	426,51	0,26	0,15	0,26	-0,23	0,09
P365 _{NI}	203,77	99,52	-58,32	46,86	423,91	715,73	0,28	0,13	0,23	-0,37	0,06
P365 _S	195,73	96,29	-53,73	46,57	428,34	713,21	0,27	0,13	0,23	-0,36	0,06
P365 _I	190,69	94,13	-50,81	47,08	430,86	711,94	0,26	0,13	0,23	-0,35	0,06
P550 _{NI}	365,35	166,90	-113,37	5,61	557,31	981,79	0,37	0,17	0,28	-0,40	0,00
P550 _S	352,25	162,24	-105,10	2,57	565,55	977,52	0,36	0,17	0,28	-0,39	0,00
P550 _I	344,19	158,67	-99,57	1,96	570,12	975,37	0,35	0,16	0,28	-0,37	0,00

^a σ_a^2 : variância genética aditiva; σ_m^2 : variância genética materna; σ_{am} : co-variância genética entre os efeitos genéticos diretos e maternais; σ_{pe}^2 : variância de ambiente permanente; σ_E^2 : variância residual; σ_p^2 : variância fenotípica; h_d^2 : herdabilidade direta; h_m^2 : herdabilidade materna; h_T^2 : herdabilidade total; rg_{am} : correlação genética entre os efeitos genéticos diretos e maternais; c^2 : contribuição de ambiente permanente.

^b NI: distribuição não informativa; S: distribuição simétrica; I: distribuição informativa.

Essas observações também são verdadeiras para os pesos pós-desmame em que não foram observadas diferenças nas estimativas obtidas de diferentes distribuições iniciais. Pelos resultados apresentados na [Tabela 4](#), observou-se que eles não foram influenciados pela forma de distribuição inicial dos parâmetros. Nessa consideração, todos os valores iniciais foram obtidos das análises REML e refletiam algum conhecimento sobre os parâmetros.

Observou-se ainda que as estimativas de herdabilidades direta e maternal obtidas do GS foram maiores do que aquelas obtidas do REML, entretanto, os resultados do GS, neste estudo, foram similares para as estimativas do REML encontradas na literatura ([Lôbo et al., 1994](#); [Eler et al., 1995](#); [Reyes et al., 1995](#); [Robinson, 1996](#)). O parâmetro que apresentou maior diferença em relação a estimativas obtidas do REML foi a correlação genética dos efeitos direto e maternal que apresentaram mudanças para valores negativos.

Os resultados da utilização dos dados simulados e com o mesmo método sugeriram que, quando as distribuições são relativamente simétricas (em populações simuladas e com dados balanceados), as estimativas posteriores de GS e as de REML são similares ([Van Tassel & Van Vleck, 1996](#)). Por sua vez, [Van Tassel & Pollak \(1994\)](#), estudando um conjunto de dados de campo relativamente grande, observaram que os resultados obtidos para as médias posteriores estimadas por GS e as estimativas REML foram distintas.

Analisando dados de campo da raça Nelore e utilizando os métodos GS e REML, [Lôbo et al. \(1997\)](#) observaram que as diferenças das estimativas das covariâncias e das herdabilidades direta e maternal obtidas dos dois métodos foram grandes. As estimativas obtidas do REML, em muitos parâmetros, aparecem nos extremos da distribuição posterior obtida do método bayesiano.

Densidades posteriores para os parâmetros genéticos

As densidades posteriores para os parâmetros genéticos de acordo com o grau de conhecimento ou de confiança relativos à forma das distribuições iniciais são mostradas nas [Figuras 1 a 9](#). A forma de distribuição inicial, aparentemente, não influenciou as estimativas dos parâmetros genéticos, mas são necessários mais estudos para confirmar esses resultados.

Para a interpretação desses resultados, é necessário considerar-se que, em uma análise bayesiana, a confirmação ou não de evidência a favor ou contra uma

hipótese é a probabilidade posterior. Neste estudo, levou-se em conta que as densidades posteriores por si só já seriam bastante informativas, propiciando acompanhamento visual dos possíveis valores de um parâmetro e, dessa maneira, permitindo, se necessário, o cálculo de probabilidade e intervalos de confiança dessas estimativas ([Wang et al., 1993](#); [Jensen et al., 1994](#); [Sorensen et al., 1994](#); [Van Tassel et al., 1995](#); [Van Tassel & Van Vleck, 1996](#)).

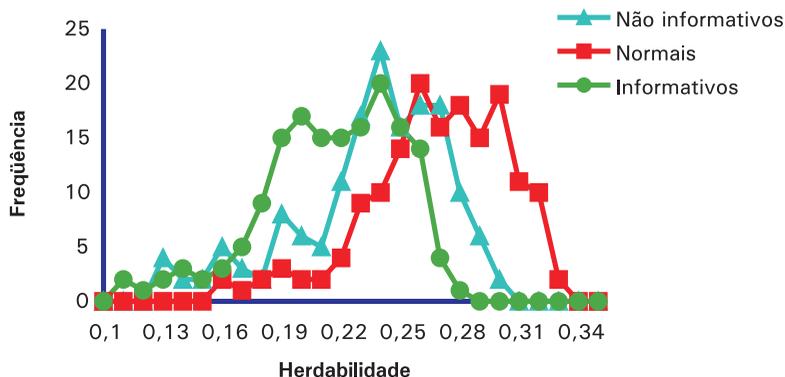


Figura 1. Distribuição das estimativas de herdabilidade direta obtidas da densidade posterior para peso aos 205 dias de acordo com diferentes parâmetros de forma assumidos.

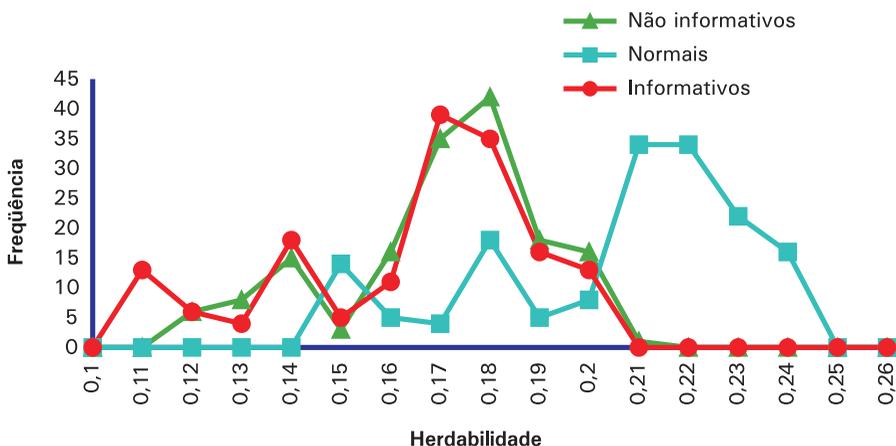


Figura 2. Distribuição das estimativas de herdabilidade maternal obtidas da densidade posterior para peso aos 205 dias de acordo com diferentes parâmetros de forma assumidos.

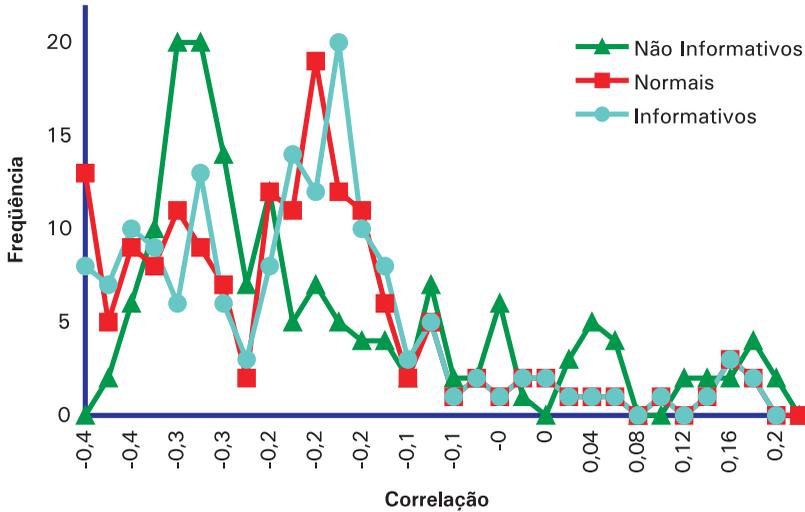


Figura 3. Distribuição das estimativas de correlação direta maternal obtidas da densidade posterior para peso aos 205 dias de acordo com diferentes parâmetros de forma assumidos.

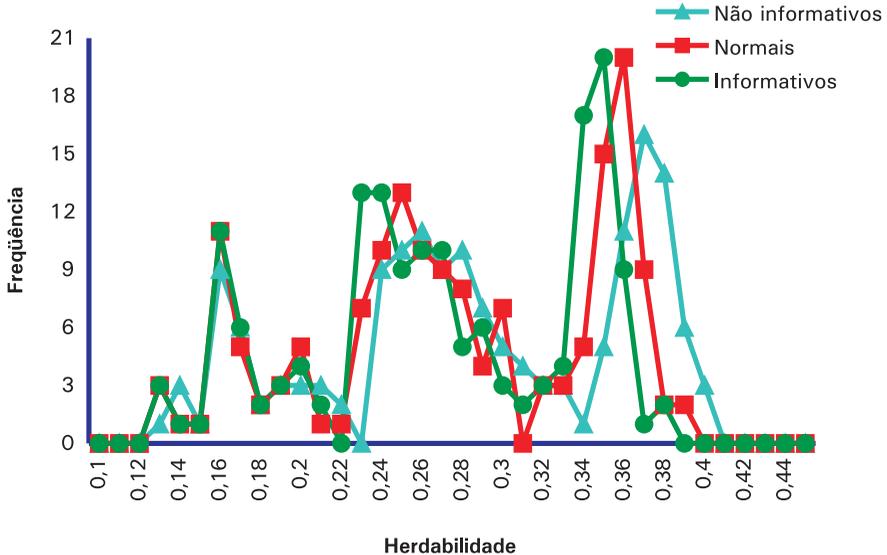


Figura 4. Distribuição das estimativas de herdabilidade direta obtidas da densidade posterior para peso aos 365 dias de acordo com diferentes parâmetros de forma assumidos.

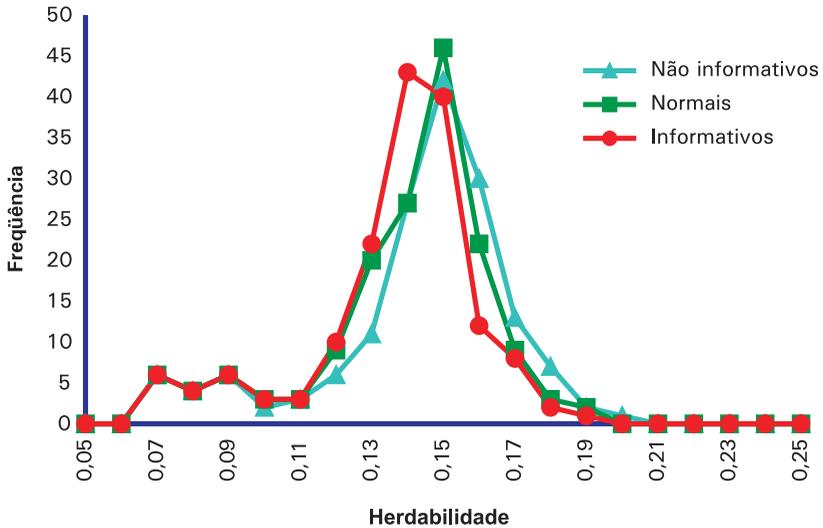


Figura 5. Distribuição das estimativas de herdabilidade maternal obtidas da densidade posterior para peso aos 365 dias de acordo com diferentes parâmetros de forma assumidos.

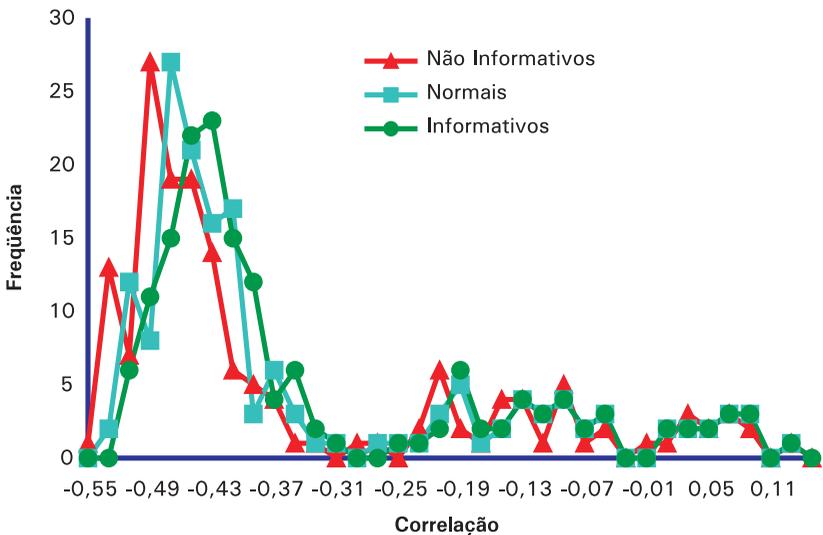


Figura 6. Distribuição das estimativas de correlação direta maternal obtidas da densidade posterior para peso aos 365 dias de acordo com diferentes parâmetros de forma assumidos.

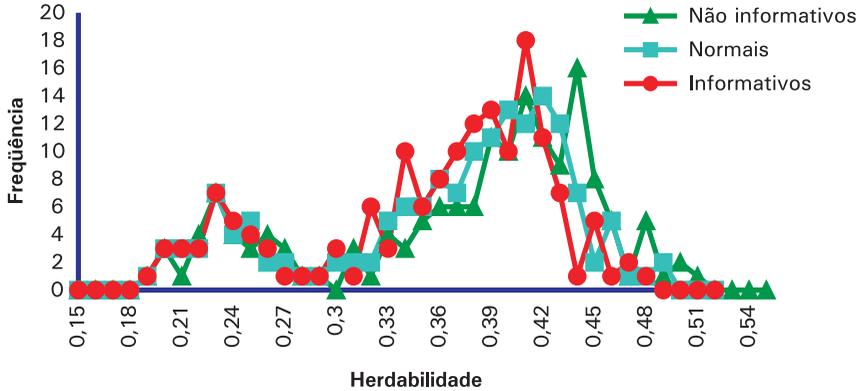


Figura 7. Distribuição das estimativas de herdabilidade direta obtidas da densidade posterior para peso aos 550 dias de acordo com diferentes parâmetros de forma assumidos.

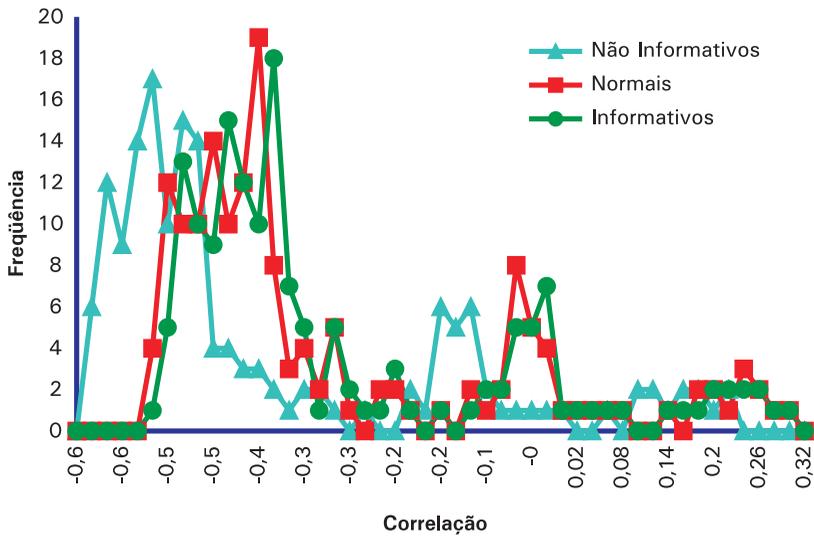


Figura 8. Distribuição das estimativas de herdabilidade maternal obtidas da densidade posterior para peso aos 550 dias de acordo com diferentes parâmetros de forma assumidos.

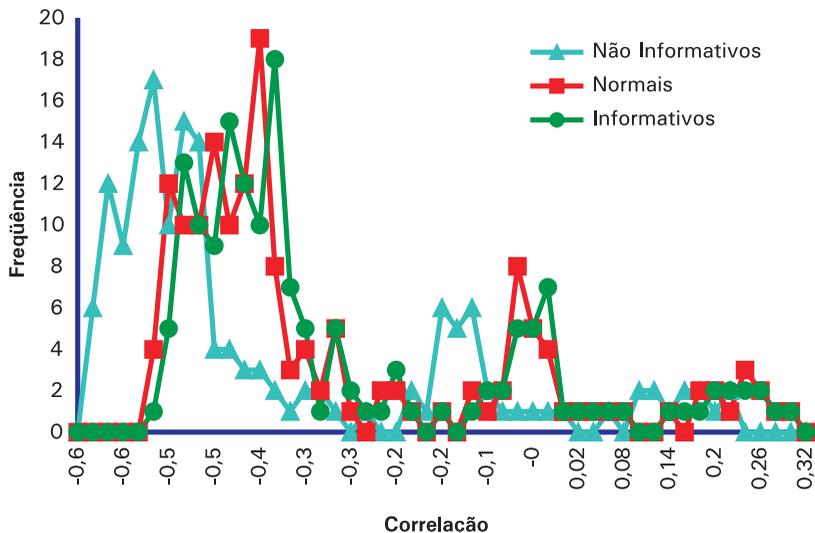


Figura 9. Distribuição das estimativas de correlação direta maternal obtidas da densidade posterior para peso aos 550 dias de acordo com diferentes parâmetros de forma assumidos.

Tamanho da cadeia de Gibbs

O objetivo, nesse caso, foi verificar como o tamanho da cadeia amostral influenciava as estimativas das médias posteriores dos componentes de covariância e dos parâmetros genéticos. As análises com GS foram feitas utilizando-se de um esquema tradicional de cadeia de Gibbs longa, em que o processo de reamostragem é contínuo tanto quanto forem os números de ciclos determinados. O processo alternativo de cadeia curta envolve a utilização de um pequeno número de ciclos, porém, recomeçando as análises várias vezes, com diferentes valores iniciais sendo fornecidos a cada amostra. A desvantagem de se usar o esquema de cadeia longa é que as amostras tendem a ser correlacionadas, todavia, isso pode ser minimizado, utilizando-se um aproveitamento amostral conservador. O termo conservador significa a adoção de cuidados especiais na especificação, principalmente, do período de descarte, antes de se aproveitar as amostras posteriores produzidas e do intervalo de utilização amostral, bem como a adoção de valores muito acima do recomendado por programas específicos para tal fim, tais como GIBBSIT e GIBANAL.

Primeiramente, optou-se por um período de uma cadeia de tamanho igual a 5000 ciclos, com os primeiros 200 ciclos descartados e frequência de utilização das amostras de 30 ciclos. A média de tempo de processamento para essas análises foi de aproximadamente 24 horas, utilizando um Pentium 166 com 128 Mb. Foram efetuadas, ainda, análises considerando tamanhos de cadeia de 10.000 e 15.000 ciclos cujos resultados são apresentados na [Tabela 5](#). O tempo de processamento para cadeias de 10.000 e 15.000 ciclos foi em torno de 40 e 60 horas, respectivamente, dependendo da característica estudada.

Na [Tabela 5](#), incluem-se, ainda, as estimativas das médias posteriores dos componentes de co-variância e dos parâmetros genéticos para peso aos 205 e 365 dias, utilizando diferentes tamanhos de cadeia amostral. O peso aos 205 dias foi escolhido por ser uma característica de grande importância econômica, servindo tanto para avaliar o potencial genético de crescimento do animal quanto para a produção de leite das vacas, tendo sido extensamente utilizado como critério de seleção na fazenda. Outro aspecto importante associado à característica P205 dias é que até essa fase todos animais “controlados” no Controle de Desenvolvimento Ponderal são mantidos no rebanho. Somente a partir da desmama, é que ocorrem a seleção e a venda de animais. Além disso, nessa fase, observa-se, também a correlação de mediana a alta com pesos às idades posteriores ([Mercadante, 1995](#)).

O peso aos 365 dias pós-desmame foi o escolhido, por ser influenciado pelo efeito direto e, provavelmente, pelos efeitos maternos efetivamente presentes no peso aos 205 dias. Conforme pode ser observado na [Tabela 5](#), os componentes de variância, bem como os parâmetros genéticos não sofreram grandes mudanças quando estimados com base em diferentes tamanhos de cadeias, e a proporcionalidade entre os componentes estudados foi mantida. Observa-se, porém, tendência de incremento na magnitude desses valores, sobretudo, da correlação genética entre os efeitos direto e maternal.

Considerando que as três cadeias estudadas tiveram origem em um mesmo valor inicial, com base nesses resultados, sugerem-se que talvez a convergência ainda não tenha sido alcançada com comprimentos menores de cadeia. No intuito de contribuir para essa discussão, foram efetuadas análises considerando 20.000 ciclos de tamanho de cadeia, não sendo verificada alteração para as estimativas de herdabilidade direta e maternal, embora a correlação entre os efeitos genéticos direto e maternal tenham apresentado ligeira diminuição.

Tabela 5. Estimativas^a de médias posteriores dos componentes de co-variância e parâmetros genéticos^b para os pesos aos 205 (P205) e 365 dias (P365) das análises univariadas, usando a amostragem de Gibbs em que consideram diferentes tamanhos de cadeia amostral^c.

Característica	σ_a^2	σ_m^2	σ_{am}	σ_{pe}^2	σ_E^2	σ_P^2	h_d^2	h_m^2	h_T^2	rg_{am}	c^2
P205 _{TC1}	112,28	70,01	-25,19	37,61	232,24	426,96	0,26	0,16	0,26	-0,28	0,09
P205 _{TC2}	113,25	78,20	-28,99	34,06	231,83	428,35	0,26	0,18	0,26	-0,30	0,08
P205 _{TC3}	116,99	78,59	-31,95	35,00	229,81	429,32	0,27	0,18	0,25	-0,32	0,08
P365 _{TC1}	218,21	99,50	-65,82	48,79	416,63	717,32	0,30	0,14	0,24	-0,44	0,07
P365 _{TC2}	217,24	113,53	-74,10	42,88	418,13	717,66	0,30	0,16	0,23	-0,47	0,06
P365 _{TC3}	226,08	121,01	-82,35	42,46	413,22	720,43	0,31	0,17	0,23	-0,49	0,06

^a Considerando distribuição inicial informativa e valores iniciais obtidos da literatura.

^b σ_a^2 : variância genética aditiva; σ_m^2 : variância genética materna; σ_{am} : co-variância genética entre os efeitos genéticos direto e maternal; σ_{pe}^2 : variância de ambiente permanente; σ_E^2 : variância residual; σ_P^2 : variância fenotípica; h_d^2 : herdabilidade direta; h_m^2 : herdabilidade materna; h_T^2 : herdabilidade total; rg_{am} : correlação genética entre os efeitos genéticos diretos e maternais; c^2 : contribuição de ambiente permanente.

^c Os tamanhos de cadeia amostral foram de TC1 = 5000, TC2 = 10000 e TC3 = 15000 ciclos, respectivamente.

Em outra análise, também avaliando diferentes tamanhos de cadeia de Gibbs, o método GS foi usado para calcular as médias posteriores dos componentes de co-variância para peso aos 205 dias, em bovinos da raça Nelore, obtidas das amostras da cadeia de Gibbs, bem como para a estimação de correlações e de herdabilidades direta e maternal (Tabela 6). Para essas análises, a Amostragem de Gibbs foi efetuada utilizando o esquema de cadeia longa, usando um Pentium III 750 com 640 Mb de RAM. Depois de um período de descarte amostral de 2000 ciclos, o total de 490 e 990 amostras dos parâmetros do modelo linear foi retirado com intervalo amostral de 200 ciclos, com o tamanho de cadeia amostral de 100.000 e 200.000 ciclos respectivamente.

Tabela 6. Estimativas de médias posteriores dos componentes de co-variância e parâmetros genéticos para peso aos 205 dias (P205), usando a Amostragem de Gibbs, com base nas distribuições iniciais não informativas ($v = 0$) e segundo dois tamanhos de cadeia longa pelo método Iterativo de Gauss Seidel (IGS).

Tamanho de cadeia	σ_a^2	σ_m^2	σ_{am}	σ_{pe}^2	σ_E^2	σ_P^2	h_d^2	h_m^2	h_T^2	rg_{am}	c^2
100.000	120,31	79,65	-32,24	35,36	227,77	430,85	0,27	0,18	0,26	-0,32	0,082
200.000	120,20	80,56	-32,93	35,10	227,85	430,79	0,27	0,18	0,26	-0,33	0,081

σ_a^2 : variância genética aditiva; σ_m^2 : variância genética materna; σ_{am} : co-variância genética entre os efeitos genéticos direto e maternal; σ_{pe}^2 : variância de ambiente permanente; σ_E^2 : variância residual; σ_P^2 : variância fenotípica; h_d^2 : herdabilidade direta; h_m^2 : herdabilidade materna; h_T^2 : herdabilidade total; rg_{am} : correlação genética entre os efeitos genéticos direto e maternal; c^2 : contribuição de ambiente permanente.

Nas [Figuras 10 e 11](#), apresentam-se as estimativas das densidades posteriores para os componentes de co-variância e, na [Figura 12](#), os resultados das estimativas de parâmetros genéticos, obtidos de diferentes tamanhos de cadeia amostral sempre utilizando o esquema de cadeia longa ([Jensen et al., 1994](#)) e considerando uma forma de distribuição inicial não informativa ($v = 0$). Verificou-se que o procedimento de adotar-se diferentes tamanhos de cadeia pouco ou nada influenciou o comportamento das distribuições marginais posteriores dos parâmetros genéticos em tamanhos de cadeia de 100.000 e 200.000 ciclos.

Com base nos resultados, pode-se observar que os parâmetros genéticos apresentaram densidades marginais posteriores muito mais estáveis com tamanho de cadeia longa, aproximando de uma distribuição normal. [Garcia-Cortés et al. \(1998\)](#) observaram que a convergência amostral dificilmente é alcançada com tamanhos de cadeia relativamente pequenos, o que pode sugerir tamanhos de cadeia de até 1.000.000 ciclos.

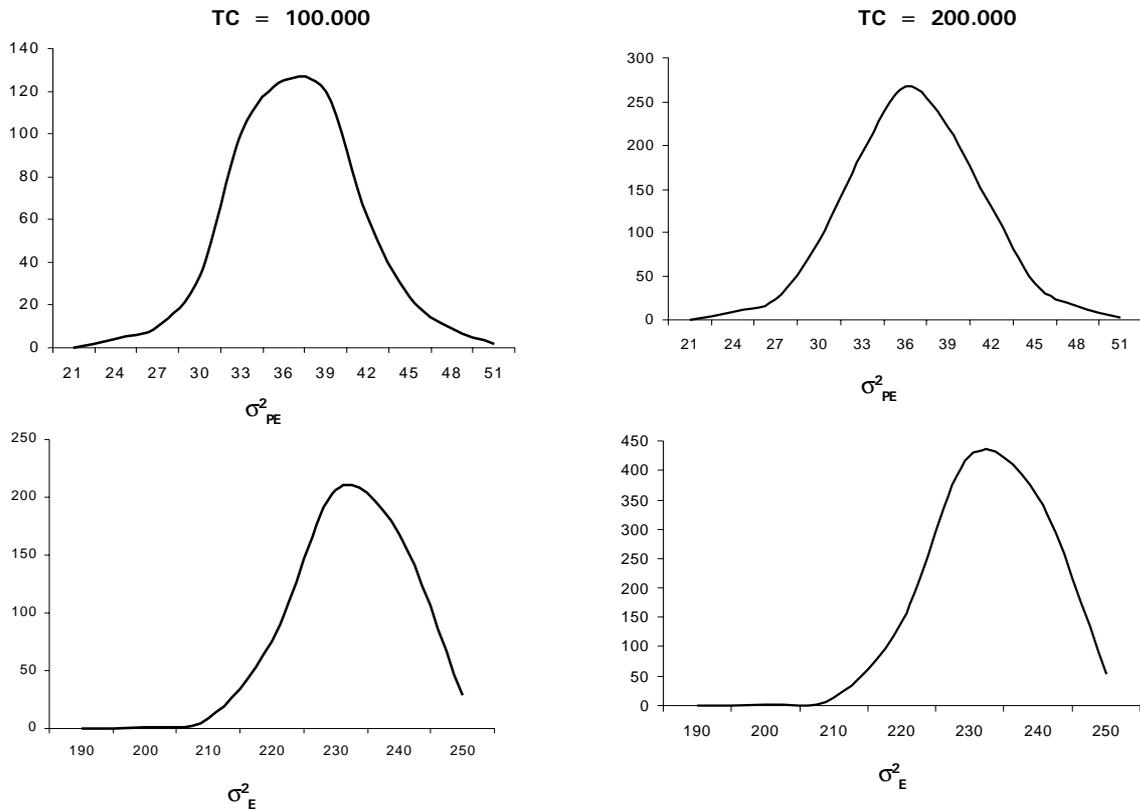


Figura 10. Densidades posteriores das variâncias permanente e residual, estudadas para peso aos 205 dias com esquema de cadeia longa (TC=100.000 e TC=200.000) e considerando uma $\nu = 0$, representada, respectivamente, nas partes superior e inferior da figura.

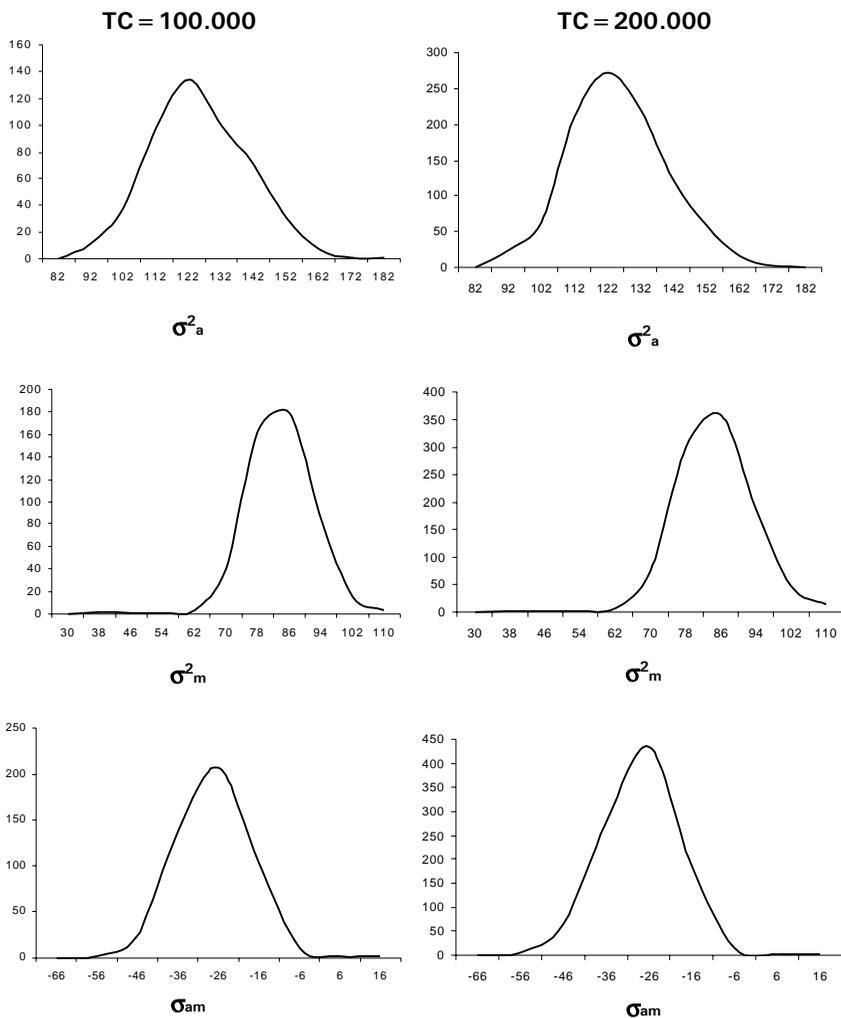


Figura 11. Densidades posteriores da variância aditiva direta, maternal e co-variância aditiva direta maternal, estudadas para peso aos 205 dias com esquema de cadeia longa (TC = 100.000 e TC = 200.000) e considerando uma $v = 0$, representada, respectivamente nas partes superior, central e inferior da figura.

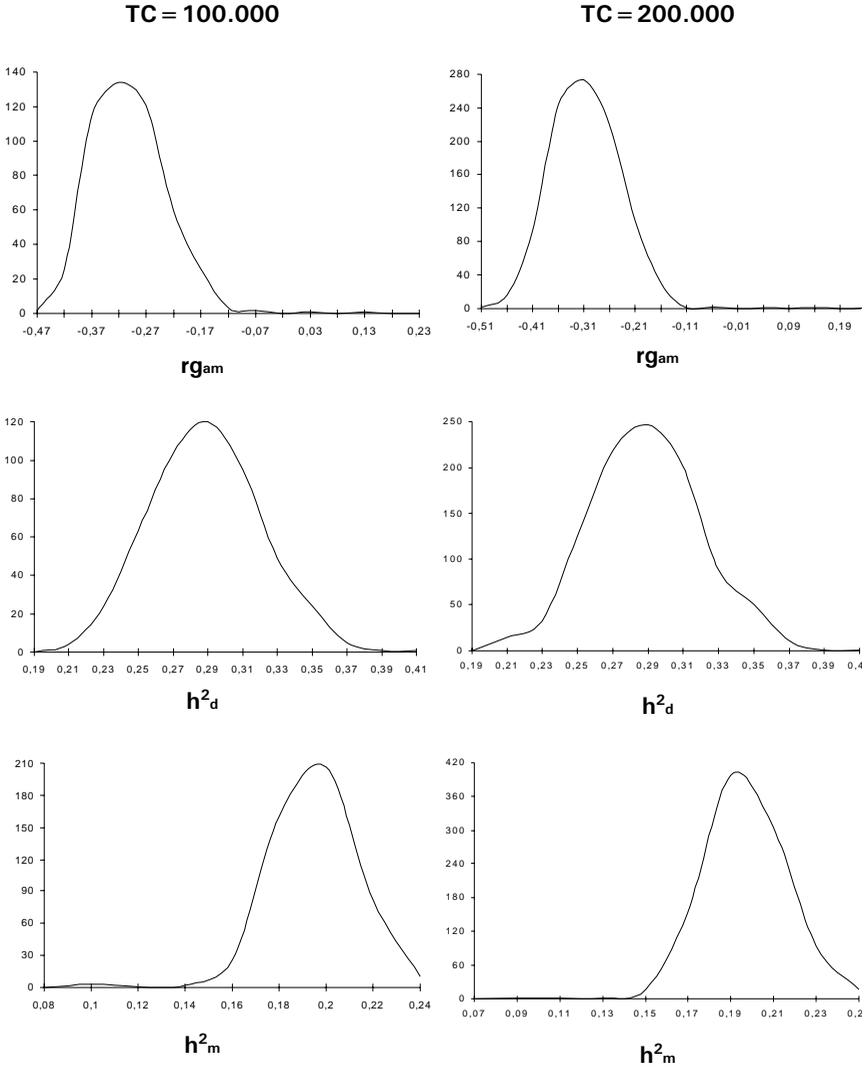


Figura 12. Densidades posteriores dos parâmetros estudados para peso aos 205 dias com esquema de cadeia longa (TC=100.000 e TC=200.000) e considerando uma $\nu = 0$, representada, respectivamente nas partes superior, central e inferior da figura.

Valores iniciais usando o GS

As estimativas das médias posteriores obtidas das análises univariadas dos componentes de co-variância e dos parâmetros genéticos para os pesos aos 205, 365 e 550 dias de idade foram obtidas de três conjuntos distintos de valores iniciais, com o parâmetro de forma igual a 5, ou seja, assumindo uma distribuição de forma simétrica. Os critérios adotados para a determinação dos três conjuntos de valores iniciais utilizados foram: (1) média de estimativas disponíveis na literatura; (2) estimativas obtidas de análises REML; (3) valores obtidos de forma não sistemática. Essas análises foram efetuadas para tentar propiciar melhor conhecimento de como a qualidade de valores iniciais de um parâmetro pode influenciar nas estimativas das médias posteriores obtidas do método GS (Tabela 7).

Tabela 7. Estimativas de médias posteriores dos componentes de co-variância e dos parâmetros genéticos^a para pesos aos 205 (P205), 365 (P365) e 550 dias (P550), com base nas análises univariadas, usando a amostragem de Gibbs em que se consideram as diferentes estimativas iniciais^b.

Carac- terís- tica	σ_a^2	σ_m^2	σ_{am}	σ_{pe}^2	σ_E^2	σ_P^2	h_d^2	h_m^2	h_T^2	rg_{am}	c^2
P205 ₋₁	115,82	109,40	-25,49	0,13	236,03	435,75	0,27	0,25	0,19	-0,22	0,01
P205 ₋₂	112,62	64,50	-21,22	39,02	232,00	426,90	0,26	0,15	0,26	-0,23	0,09
P205 ₋₃	111,36	41,23	-5,22	44,98	233,08	425,38	0,26	0,09	0,29	-0,03	0,11
P365 ₋₁	222,01	110,57	-70,62	43,32	414,84	720,14	0,31	0,15	0,30	-0,45	0,06
P365 ₋₂	195,73	96,29	-53,73	46,57	428,34	713,21	0,27	0,13	0,23	-0,36	0,06
P365 ₋₃	201,34	50,86	-22,48	56,91	426,16	712,79	0,28	0,07	0,27	-0,16	0,08
P550 ₋₁	382,09	168,65	-122,98	10,43	547,13	985,32	0,38	0,17	0,28	-0,46	0,01
P550 ₋₂	352,25	162,24	-105,10	2,57	565,55	977,52	0,36	0,17	0,28	-0,39	0,00
P550 ₋₃	337,22	39,46	-8,32	11,01	588,31	967,72	0,35	0,04	0,35	0,02	0,01

^a σ_a^2 : variância genética aditiva; σ_m^2 : variância genética materna; σ_{am} : co-variância genética entre os efeitos genéticos diretos e maternos; σ_{pe}^2 : variância de ambiente permanente; σ_E^2 : variância residual; σ_P^2 : variância fenotípica; h_d^2 : herdabilidade direta; h_m^2 : herdabilidade materna; h_T^2 : herdabilidade total; rg_{am} : correlação genética entre os efeitos genéticos direto e maternal; c^2 : contribuição de ambiente permanente;

^b 1: média de estimativas disponíveis na literatura; 2: estimativas obtidas de análises REML; 3: valores obtidos de forma não sistemática.

Como mostrado na [Tabela 7](#), as estimativas obtidas das situações 1 e 2 foram similares, principalmente, para P365 e P550. Para P205, observa-se maior contribuição devida aos efeitos maternos e conseqüente diminuição da magnitude dos efeitos do ambiente permanente na situação 1. Por sua vez, na situação 3, foram verificadas, para os três pesos, estimativas de herdabilidade maternal inferiores às da situação 2. Porém, conforme discutido anteriormente, recomenda-se a utilização de tamanhos de cadeia maiores quando se utilizam valores iniciais obtidos de forma não sistemática que poderiam levar a estimativas mais similares entre as três situações, pois o maior número de ciclos levaria à convergência da cadeia. Dessa forma, foi observado que o método GS permite a utilização de qualquer tipo de valores iniciais, lembrando que o uso desse método requer maior número de ciclos quando se utilizam de valores iniciais colhidos de forma não sistemática. Esses resultados indicam que o método GS não foi sensível à mudança radical dos valores iniciais.

Conclusão

O método GS pode ser usado em problemas práticos na estimação dos componentes de co-variância utilizando modelo animal que envolve dados de campo. Verificou-se que o método pode ser adequado para estimar as médias posteriores para todos os componentes de variância do modelo, por permitir a obtenção das distribuições posteriores dos componentes de co-variância e, conseqüentemente, dos parâmetros genéticos de interesse. No entanto, o tempo de processamento de GS foi maior que o requerido pelo método REML.

As estimativas dos parâmetros genéticos, obtidas deste estudo para a raça Nelore foram, de maneira geral, similares a boa parte das estimativas encontradas na literatura para a mesma raça quando se usou a teoria BLUP em modelo animal. Os valores obtidos das análises univariadas para herdabilidade direta foram inferiores aos das análises bivariadas para todas as características estudadas. Na estruturação dos dados, aparentemente, observaram-se problemas para a utilização de um modelo completo, considerando efeitos diretos, maternos e permanentes para os pesos pós-desmame, utilizando o método REML, principalmente, em análises bivariadas. Nessa estrutura de dados, a inclusão dos efeitos permanentes pode complicar ainda mais o confundimento entre os efeitos direto e maternal. Esse fato pode ser devido à dificuldade da partição de efeitos com base no mesmo critério de classificação, como é o caso em que os efeitos maternos e permanentes são obtidos da vaca.

Com base nos valores obtidos para as estimativas das correlações genéticas entre os efeitos direto e maternal pelo método GS, verifica-se a necessidade de se considerar a seleção conjunta para os dois componentes. Isso pode ser aplicado às estimativas obtidas do método REML já que, apesar de serem positivas, sua magnitude é muito baixa, não se permitindo afirmar que a seleção, que considera somente o efeito direto, contemple também o maternal.

Referências Bibliográficas

- BOLDMAN, K.G.; KRIESE, L.A.; VAN VLECK, L.D.; KACHMAN, S.D.A. **A Manual for Use of MTDFREML** : A set of programs to obtain estimatives of variance and covariances. USDA. ARS., 1993.
- BOLDMAN, K.G.; KRIESE, L.A.; VAN VLECK, L.D.; VAN TASSELL C.P.; KACHMAN, S.D. **A manual for use of MTDFREML**: A set of programs to obtain estimates of variance and covariances [DRAFT]. U.S.A., 1995.
- ELER, J.P.; FERRAZ, J.B.S.; LÔBO, R.B.; JOSAHKIAN, L.A. Genetic antagonism between growth and maternal ability in Nelore cattle. **Journal of Animal Science**, Champaign, v.70, p.138, 1992.
- ELER, J.P.; FERRAZ, J.B.S.; SILVA, P.R. Parâmetros genéticos de pesos de bovinos da raça Nelore. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 32., 1995, Brasília. **Anais...** Brasília: SBZ, 1995. p. 709-710.
- GARCIA-CORTÉS, L.A; RICO, M.; GROENEVELD, E. Using coupling with the Gibbs Sampler to asses convergence in animal Models. **Journal of Animal Science**, Champaign, v.76, p 441-447, 1998.
- GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, Champaign, v. 63, p.217, 1986.
- GIANOLA, D.; IM, S.; FERNANDO, R.L.; FOULLEY, J.L. Mixed models methodology and the Box-Cox theory of transformations: a Bayesian approach. In: **Advances in statistical methods for genetic improvement of Livestock** GIANOLA, D.; HAMMOND, K.(Ed.), New York: Springer - Verlay, 1990. p. 15-40.
- GRASER, H.U.; SMITH, S.P.; TIER, B. A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. **Journal of Animal Science**, Champaign, v.64, p.1362-1370, 1987.

HENDERSON, C.R. Sire evaluation and genetic trends. In: Proc. of the Anim. Breed. and Genet. Symp. in Honor of Dr. Jay L. Lush. Amer. Soc. Anim. Sci. and Amer. **Dairy Sci. Assoc.**, Champaign, IL. 1973.

JENSEN, J.; WANG, C.S.; SORENSEN, D.A.; GIANOLA, D. Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. **Acta Agricultural Scandinavica**, Stockholm, v.44, p.193, 1994.

KOOTS, K.R.; GIBSON, J.P.; WILTON, J.W. Analysis of published genetic parameters estimates for beef production traits. 2 - Phenotypic and genetics correlations. **Animal Breeding Abstracts**, Farnham Royal, v.62, p.825-853, 1994.

LÔBO, R.B.; REYES, A. de los; FERRAZ, J.B.S.; BEZERRA, L.A.F.; MERCADANTE, M.E.Z.; DUARTE, F.A.M. Bivariate animal model analysis of growth weights and scrotal circumference of Nelore cattle in Brazil. In: WORLD CONGRESS ON GENETIC APPLIED TO LIVESTOCK PRODUCTION, 5., 1994, Ontario. **Proceedings...** Ontario: University of Guelph, 1994. P. 199-201.

LÔBO, R.B.; REYES, A. de los; BEZERRA, L.A.F.; PIMENTA, C.M.S.J.; SILVA, J.A.II de V.; TAVARES, H de F. Parâmetros fenotípicos e genéticos de pesos e perímetro escrotal às idades-padrão em animais de raça Nelore. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 32., 1995, Brasília. **Anais...** Brasília: SBZ, 1995. p.625-627.

LÔBO, R.B.; OLIVEIRA, H.N.O.; BEZERRA, L.A.F.; MAGNABOSCO, C.D.U.; REYES, A. de los; RUSSO, A.R.V. Estimativa de componentes de (co)variância e herdabilidade para o peso real aos 120 dias de idade na raça Nelore usando análise bayesiana. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 34., 1997, Juiz de Fora. **Anais...** Juiz de Fora: SBZ, 1997. 1 CD - Rom.

LOPES, P.S.; MARTINS, E.N.; SILVA, M.A. **Métodos de resolução de sistemas de equações lineares**. Viçosa: UFV, 1993. 51p.

MAGNABOSCO, C.D.U.; FAMULA, T.R.; LÔBO, R.B.; REYES, A. de los; OJALA, M. Estimativas de parâmetros genéticos e de ambiente de características de crescimento em bovinos da raça Nelore. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 33., 1996, Fortaleza. **Anais...** Fortaleza: SBZ, 1996. p.142-144.

MAGNABOSCO, C.D.U.; LÔBO, R.B.; REYES, A. de los; MARTINS, E.N.; FAMULA, T. Inferência bayesiana em parâmetros genéticos para peso aos 205 dias de idade em bovinos da raça Nelore. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 34., 1997, Juiz de Fora. **Anais...** Juiz de Fora: SBZ, 1997. 1 CD - Rom.

MAGNABOSCO, C.D.U, LÔBO, R.B, FAMULA, T.R. Bayesian inference for genetic parameter estimation on growth traits for Nelore cattle in Brazil, using the gibbs sampler. **Journal of Animal Breeding and Genetics**, Innsbruck, v.117, p. 169-188, 2000.

MERCADANTE, M.E.Z. **Estudo das relações genético-quantitativas entre características de reprodução, crescimento e produção em fêmeas da raça Nelore**. 1995. 90 p. Dissertação (Mestrado) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto.

MEYER, K. DFREML: User Notes. Version 2.1. Edinburgh: University of Edinburgh. 1993.

MEYER, K. Estimates of direct and maternal correlations among growth traits in Australian Beef Cattle. **Livestock Production Science**, Amsterdam, v.38, p.91-105, 1994.

RAFTERY, A.E.; LEWIS, S.M. Gibbsit version 2.0. StaLib file/general/gibbsit. Disponível em: <<http://lib.stat.cmu.edu/general/gibbsit>>. Acesso em: 1994.

REYES, A. de los; LÔBO, R.B.; OLIVEIRA, H.N.; BEZERRA, S.P.F.; MESSAGE-JÚNIOR, A.; GESTAL, R.L.P. Estimativa de efeitos genéticos direto e materno para pesos até 365 dias, usando modelo animal bicaracter em rebanhos Nelore. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 31., 1994, Maringá. **Anais...** Maringá: SBZ, 1994. p.167.

REYES, A. de los; LÔBO, R.B.; OLIVEIRA, H.N.; TONHATI, H.; MARTINS-FILHO, R.; BEZERRA, L.A.F. Estimación de (co)varianzas y DEP's por modelo animal bicarácter para pesos y perímetro escrotal de ganado Nelore en Brasil. In: CONGRESO ARGENTINO DE PRODUCCIÓN ANIMAL, 14., 27 nov.-1 dez. 1995. v. 19, p.926-930.

ROBINSON, D.L. Estimation and interpretation of direct and maternal genetic parameters for weights of Australian Angus cattle. **Livestock Production Science**, Amsterdam, v.45, p.1-11, 1996.

SMITH, S.P.; GRASER, H.U. Estimating variance components in a class of mixed models by restricted maximum likelihood. **Journal of Dairy Science**, Champaign, v.69, p.1156-1165, 1986.

SORENSEN, D. Gibbs sampling in quantitative genetics. Denmark: Danish Institute of Animal, Department of Breeding and Genetics, 1996. (Intern Report, 82).

SORENSEN, D.A.; WANG, C.S.; JENSEN, J.; GIANOLA, D. Bayesian analysis of genetic change due to selection using Gibbs sampling. **Genetics Selection Evolution**, Paris, v.26, p.333-360, 1994.

TROVO, J.B.F.; RAZOOK, A.G.; LÔBO, R.B.; FIGUEIREDO, L.A.; BEZERRA, L.A.F. Avaliações genéticas em bovinos de corte utilizando estruturas de dados não-convencionais. In: CONGRESSO BRASILEIRO DAS RAÇAS ZEBUÍNAS, 2., 1996, Uberaba. **Anais...** Uberaba: ABCZ, 1996. p. 3-9.

URIBE, H. Random regressions in animal breeding. Campo Grande, 1996. **Workshop**, EMBRAPA, 17 a 26 /11/1996.

VAN KAAM. GIBANAL - Analyzing program for Markov Chain Monte Carlo sequences. Version 2.4., Netherlands, 1997.

VAN TASSELL, C.P.; POLLAK, E.J. Estimates of (co)variance components for weaning weight used in national beef cattle genetic evaluations. **Journal of Animal Science**. Supplement, n.1, Champaign, v.72, p.148, 1994.

VAN TASSELL, C.P.; CASELLA, G.; POLLAK, E.J. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. **Journal of Dairy Science**, Champaign, v.78, p.678, 1995.

VAN TASSELL, C.P.; VAN VLECK, L.D. Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. **Journal of Animal Science**, Champaign, v.74, p.2586-97, 1996.

WANG, C.S.; RUTLEDGE, J.J.; GIANOLA, D. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. **Genetics Selection Evolution**, Paris, v.25, p.41, 1993.

WANG, C.S.; RUTLEDGE, J.J.; GIANOLA, D. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. **Genetics Selection Evolution**, Paris, v.26, p.91, 1994.

Implementation of the Gibbs Sampling for variance components and genetic parameter estimation on data field for Nelore cattle

Abstract – This data set consisted of over 29245 field records of Nelore cattle born between 1980 and 1993, raised on pasture, from herds in Southeastern and Midwestern Brazil. Three growth traits were included in the analyses: 205 (W205), 365 (W365) and 550 days (W550) weight. The analyses were conducted using single-trait and multiple-trait animal models. (Co) variance components were estimated by Restricted Maximum Likelihood (REML) using a derivative-free algorithm (DFREML) for multiple traits (MTDFREML). Bayesian inference was obtained by a multiple trait Gibbs Sampling algorithm (GS) for (co) variance component inference in animal models (MTGSAM). Three different sets of prior distributions for the (co)variance components were used: flat, symmetric and sharp. The results suggested that shape of the prior distributions did not affect the estimates of (co) variance components. Different chain lengths were also appraised, using the long chain scheme. Increasing chain lengths above 100,000 rounds had little or no influence on the posterior marginal distributions of the genetic parameters. GS produced similar but slightly higher estimates of genetic parameters than REML, however, the greater robustness of GS makes it the method of choice for many applications.

Index terms: gibbs sampling, variance components, genetic.