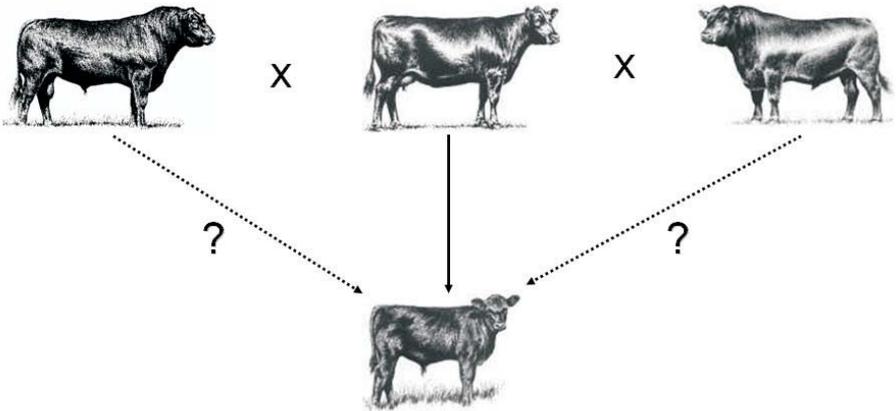


**Metodologia para estimação do mérito genético de animais com paternidade incerta sob inferência bayesiana**



**Qual é o pai verdadeiro?**

ISSN 1983-0467

Dezembro, 2008

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

# *Boletim de Pesquisa e Desenvolvimento 32*

## **Metodologia para estimação do mérito genético de animais com paternidade incerta sob inferência bayesiana**

*Fernando F. Cardoso*

*Luciana Shiotsuki*

*Robert J. Tempelman*

Embrapa Pecuária Sul  
Bagé, RS  
2008

Exemplares desta publicação podem ser adquiridos na:

**Embrapa Pecuária Sul**

BR 153, km 603 - Caixa Postal 242

96401-970 - Bagé, RS

Fone/Fax: (0XX53) 3242-8499

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

[sac@cppsul.embrapa.br](mailto:sac@cppsul.embrapa.br)

**Comitê Local de Publicações da Embrapa Pecuária Sul**

Presidente: Alexandre Varella

Secretário-Executivo: Ana Maria Sastre Sacco

Membros: Eduardo Salomoni, Eliane Mattos Monteiro, Eliara Freire Quincozes, Graciela Olivella Oliveira, João Batista Beltrão Marques, Magda Vieira Benavides, Naylor Bastiani Perez

Supervisor editorial: Comitê Local de Publicações - Embrapa Pecuária Sul

Revisor de Texto: Comitê Local de Publicações - Embrapa Pecuária Sul

Tratamento de ilustrações: Kellen Pohlmann

Editoração eletrônica: Kellen Pohlmann

**1ª edição (2007)**

**Todos os direitos reservados.**

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

**Dados internacionais de Catalogação na Publicação (CIP)  
Embrapa Pecuária Sul**

---

Cardoso, Fernando Flores

Metodologia para estimação do mérito genético de animais com paternidade incerta sob inferência bayesiana / Fernando Flores Cardoso, Luciana Shiotsuk, Robert J. Tempelman.\_ Bagé: Embrapa Pecuária Sul, 2008.

(Boletim de pesquisa e desenvolvimento / Embrapa Pecuária Sul, ISSN 1983-0467 ; 32)

Sistema requerido: Adobe Acrobat Reader

Modo de acesso:

<<http://www.cppsul.embrapa.br/unidade/publicacoes/list/193>>

Título da página Web (acesso em 30 dez. 2008)

1. Gado de corte. 2. Parâmetro genético. 3. Genética quantitativa. I. Shiotsuki, Luciana. II. Tempelman, Robert J. III. Título. IV. Série.

---

CDD 636.0821

© Embrapa, 2007

# Sumário

<b>Resumo</b> .....	9
<b>Abstract</b> .....	11
<b>1. Introdução</b> .....	13
<b>2. Metodologia</b> .....	15
2.1 Modelo animal reduzido com efeito materno.....	15
2.2 Modelo hierárquico bayesiano para paternidade incerta.....	17
2.3 Validação por Simulação.....	22
2.4 Análise de dados de desempenho de um rebanho Hereford.....	24
2.5 Inferência Bayesiana.....	25
2.5.1 Análises do estudo de simulação.....	25
2.5.2 Análises de dados do rebanho Hereford.....	26
2.6. Critérios para escolha do melhor modelo.....	28
<b>3. Resultados</b> .....	30
3.1 Inferência sobre dados simulados.....	30

3.2 Inferência nos dados do rebanho Hereford.....	34
3.2.1 Ganho de peso pós-desmama.....	34
3.2.2. Peso à desmama.....	37
<b>4. Discussão.....</b>	<b>42</b>
<b>5. Conclusões.....</b>	<b>44</b>
<b>Agradecimentos.....</b>	<b>44</b>
<b>Referências citadas.....</b>	<b>45</b>

# Lista de Figuras

<b>Fig. 1.</b> Descrição dos quatro estágios do modelo hierárquico bayesiano desenvolvidos por Cardoso e Tempelman (2003).....	21
<b>Fig. 2.</b> Descrição da população simulada por Cardoso e Tempelman (2003).....	24
<b>Fig. 3.</b> Quadrado médio do erro de predição (MSEP) dos efeitos aditivos e maternos de animais pais e não pais com incerteza de paternidade para as características 1 e 2 sobre os três modelos.....	32
<b>Fig. 4.</b> Correlação de ordem de efeitos aditivos e maternos de animais pais e não pais com paternidade incerta para característica 1 e 2 sobre três modelos.....	33
<b>Fig. 5.</b> Desvio padrão (SD) do efeito genético aditivo do ganho pós-desmama, em kg, de um rebanho Hereford brasileiro com incerteza de paternidade, obtidos pelo ANRM vs. HIER.....	37
<b>Fig. 6.</b> Desvio padrão (SD) do efeito genético aditivo do peso à desmama, em kg, de um rebanho Hereford brasileiro com incerteza de paternidade, obtidos pelo ANRM vs. HIER.....	41

# Lista de Tabelas

<b>Tabela 1.</b> Descrição dos conjuntos de dados do rebanho Hereford para o peso à desmama (PD) e o Ganho de peso pós-desmama (GPD).....	25
<b>Tabela 2.</b> Médias a posteriori das probabilidades dos pais verdadeiros ( $E(\pi_j^{(mc)}   y)$ ) e probabilidade do pai verdadeiro ser designado ao indivíduo $j$ ( $\text{Prob}(s_j' = s_j^{(mc)}   y)$ ) para características 1 e 2 pelo tamanhos de grupos de reprodutores múltiplos de animais com (pais) e sem progênie (não pais).....	31
<b>Tabela 3.</b> Deviance ( $\bar{D}$ ), penalização por número efetivo de parâmetros ( $p_D$ ) e critério de informação da deviance (DIC) para os modelos ANRM e HIER usados na análise do ganho de peso pós-desmama (GPD) e do peso à desmama (PD) de um rebanho Hereford brasileiro.....	34
<b>Tabela 4.</b> Mediana e desvio padrão (DP) a posteriori, percentis de 2.5% e 97.5% e tamanho efetivo da amostra (TEA) de componentes de variância (em $\text{kg}^2$ ) e parâmetros genéticos para ganho de peso pós-desmama em um rebanho Hereford brasileiro, obtidos pelos modelos de ANRM e HIER.....	35
<b>Tabela 5.</b> Mediana e desvio padrão (DP) a posteriori, percentis de 2.5% e 97.5% e tamanho efetivo da amostra (TEA) de componentes de variância (em $\text{kg}^2$ ) para peso à desmama em um rebanho Hereford brasileiro, obtidos usando o modelo hierárquico Bayesiano (HIER) e matriz de parentesco médio (ANRM).....	39

**Tabela 6.** Mediana e desvio padrão (DP) a posteriori, percentis de 2,5% e 97,5% e tamanho efetivo da amostra (TEA) de parâmetros genéticos para peso à desmama em rebanho Hereford brasileiro, obtidos usando o modelo hierárquico Bayesiano (HIER) e matriz de parentesco média (ANRM).....40

# Metodologia para estimação do mérito genético de animais com paternidade incerta sob inferência bayesiana

---

*Fernando F. Cardoso*<sup>1</sup>

*Luciana Shiotsuki*<sup>2</sup>

*Robert J. Tempelman*<sup>3</sup>

## Resumo

Os objetivos do presente trabalho foram apresentar um modelo animal hierárquico (HIER) sob enfoque bayesiano usando métodos Monte Carlo via cadeias de Markov para predizer o mérito genético de animais com paternidade incerta; usar o modelo proposto para estimar probabilidades a posteriori de cada possível pai, combinando informações a priori e do conjunto de dados; e comparar a performance do modelo proposto com modelo baseado na matriz de parentesco médio de Henderson (ANRM), que utiliza probabilidade igual para todos os touros candidatos, via simulação e na análise do peso à desmama (PD) e o ganho de peso pós-desmama (GPD) de animais da Raça Hereford. O estudo de simulação consistiu de dez conjuntos de dados caracterizados por 30% dos animais tendo paternidade incerta para cada uma de duas características, uma tendo herdabilidade moderada para efeito genético direto e efeito genético materno similar a PD e a outra com alta herdabilidade para efeito genético direto, similar a GPD. Inferências a posteriori sobre os componentes de variância foram similares entre os dois modelos no estudo de simulação e nos dados de animais Hereford.

---

<sup>1</sup>Médico Veterinário, PhD, Pesquisador da Embrapa Sul, Caixa postal 242, Bagé-RS, CEP 96401-970; fcardoso@cppsul.embrapa.br;

<sup>2</sup>Zootecnista, MS, Doutoranda em Genética e Melhoramento Animal/ UNESP; luushi@bol.com.br;

<sup>3</sup>Department of Animal Science, Michigan State University; tempelma@msu.edu;

Além disso, a correlação de ordem para a média a posteriori de efeitos genéticos entre os dois modelos excedeu 0,90. Entretanto, grandes diferenças na média a posteriori entre os dois modelos foram observadas para alguns animais com paternidade desconhecida, que também apresentaram maior desvio padrão à posteriori no modelo HIER comparado ao modelo ANRM, uma vez que o modelo HIER considera a incerteza nas atribuições de paternidade. Finalmente, critérios bayesianos de escolha de modelo favoreceram consistentemente o modelo HIER em relação ao ANRM nos estudos de simulação e com animais Hereford, indicando que HIER ajusta melhor os dados de animais com paternidade incerta e representa uma alternativa a ser considerada para a avaliação genética de populações que adotam acasalamentos com reprodutores múltiplos.

# Metodologia para estimação do mérito genético de animais com paternidade incerta sob inferência bayesiana

---

## Abstract

*The objectives of the current study were to propose a hierarchical animal model (HIER) via Bayesian Markov Chain Monte Carlo methods for prediction of genetic merit of animals with uncertain paternity; use the proposed model to estimate posterior probability of each possible sire, combining prior and data information; and compare the performance of the proposed model with a model based on the use of Henderson's average numerator relationship matrix (ANRM), that use equal probabilities for all candidate sire, via simulation study and analysis of weaning weight (WWT) and postweaning gain (PWG) of Hereford animals. The simulation study consisted of ten datasets characterized by 30% of animals having uncertain paternity for each of two traits, one having moderate heritabilities for direct and maternal genetic effects as WWT and another having a high heritability for direct genetic effects as PWG. Posterior inference on the variance components was very similar between the two models in the simulation study and Hereford data. Furthermore, rank correlations on posterior means for genetic effects between the two models exceeded 0.90. Nevertheless, large differences in posterior means between these two models were observed for animals with uncertain paternity, which also had larger posterior standard deviations of genetic effects using the HIER model compared to the ANRM model, since the HIER model accounts for the uncertainty on sire assignments. Bayesian model choice criteria consistently favored the HIER model over the ANRM model in both simulation and Hereford data*

*analysis studies, indicating that HIER better fits data of animals with uncertain paternity and represents an alternative to be considered for genetic evaluation of populations that adopt multiple sires mating.*

## 1. Introdução

A utilização de touros múltiplos é muito comum em sistemas de criação sob pastejo como os encontrados, por exemplo, na Argentina, Austrália, Brasil e algumas partes dos Estados Unidos. Esses sistemas adotam a exposição de um grupo de vacas a vários touros dentro de uma mesma estação de monta. Conseqüentemente, o pedigree desses rebanhos é incerto, afetando a acurácia de avaliações genéticas e a precisão de seleção.

Diversos modelos estatísticos têm sido propostos para avaliação de animais com incerteza de paternidade. Uma solução simples é a utilização de grupos genéticos, descrito por Westell et al. (1988), pelo qual, grupos de “pais imaginários” são designados aos animais que foram acasalados no mesmo piquete. Nos grupos genéticos, os grupos de pais imaginários são tipicamente definidos como o próprio grupo de reprodutores múltiplos que foram acasalados com um determinado grupo de vacas. Esta metodologia minimiza o erro de predição, especialmente quando a seleção ocorre com o uso de dados de desempenho (Cantet et al., 1993; Quaas, 1988). O uso de grupos genéticos para acasalamentos de reprodutor múltiplo, entretanto, é equivalente a assumir um número infinito de pais candidatos não endogâmicos, não aparentados e que, dentro de cada grupo, cada candidato tenha a mesma probabilidade de ser o pai correto do animal com incerteza de paternidade (Perez-Enciso; Fernando, 1992; Sullivan, 1995). No entanto, somente os touros candidatos usados dentro de cada grupo podem ser considerados candidatos, quando esta informação está disponível.

Essa estratégia é particularmente apropriada, trabalhando-se com a matriz de parentesco médio (ANRM) proposta por Henderson (1988). Este procedimento se baseia no conhecimento da probabilidade verdadeira de cada macho candidato ser o pai correto. A ANRM auxilia a correta especificação da matriz de (co)variâncias genéticas, quando essas probabilidades são presumidamente conhecidas, facilitando a melhor predição linear não-viesada (BLUP) do mérito genético. Um algoritmo simples e rápido para computar a inversa da ANRM está disponível e foi demonstrado por estudos de simulação (Perez-Enciso e Fernando, 1992; Sullivan, 1995) que se obtêm melhores respostas na seleção utilizando ANRM em comparação com grupos genéticos, quando informação sobre os possí-

veis pais foi registrada. Probabilidades a priori iguais podem ser assumidas para cada touro, entretanto, tipagem sanguínea, marcadores genéticos, fertilidade, período reprodutivo e gestação também podem ser usados para calcular as probabilidades mais distintamente (Foulley et al., 1987; Henderson, 1988).

Um procedimento empírico bayesiano inovador para inferir sobre a paternidade incerta foi proposto por Foulley et al. (1987). Essa metodologia combina dados e informações a priori para determinar probabilidades à posteriori dos pais candidatos para cada animal com paternidade incerta. Esse procedimento, entretanto, apresenta a limitação de restringir-se ao uso de modelo touro. Com o advento dos métodos Monte Carlo via Cadeias de Markov (MCMC) aplicados ao melhoramento animal (Wang et al., 1994), foi possível implementar o método de Foulley et al. (1987) para modelo animal permitindo uma avaliação estatística mais formal da precisão do mérito genético e das probabilidades das atribuições de paternidade (Cardoso; Tempelman, 2003).

Os objetivos do presente artigo são:

- 1) apresentar um modelo animal hierárquico sob enfoque bayesiano via MCMC para prever o mérito genético de animais com incerteza de paternidade;
- 2) usar o modelo proposto para estimar probabilidades a posteriori de cada possível pai, combinando informações a priori e do conjunto de dados;
- 3) Comparar a performance do modelo proposto com modelo baseado na matriz de parentesco médio proposta por Henderson (1988), que utiliza probabilidade inicial igual para todos os touros candidatos designados, via simulação e na análise o peso à desmama e o ganho de peso pós-demama de animais da Raça Hereford;

## 2. Metodologia

### 2.1 Modelo animal reduzido com efeito materno

Considere que  $\mathbf{y} = \{y_{ij}\}$  é o vetor  $n \times 1$  de registros de performance, para  $i=1,2,\dots,n$ ;  $j=1,2,\dots,q$ . Em que  $i$  identifica o registro e  $j$  o animal com a  $i$ -ésima observação. Sendo permitido que um animal  $j$  não tenha informação. Não obstante, a avaliação genética desse animal pode ser predita por parentes que tenham informações. No modelo animal reduzido (RAM) de Quaas e Pollak (1980),  $\mathbf{y}$  é separado em dois subconjuntos:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_p \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{X}_p \\ \mathbf{X}_t \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_{1p} \\ \mathbf{Z}_{1t} \mathbf{P}_t \end{bmatrix} \mathbf{a}_p + \begin{bmatrix} \mathbf{Z}_{2p} \\ \mathbf{Z}_{2t} \end{bmatrix} \mathbf{m}_p + \begin{bmatrix} \mathbf{e}_p \\ \mathbf{Z}_{1t} \boldsymbol{\gamma}_t + \mathbf{e}_t \end{bmatrix} \quad (1)$$

O primeiro,  $n_p \times 1$  subconjunto  $\mathbf{y}_p$  de  $\mathbf{y}$ , é observado em  $q_p$  animais, que aparecem com pais ou ancestrais de outros animais com registro. Na equação 1,  $\mathbf{y}_p$  é a função linear do vetor  $p \times 1$  dos efeitos fixos  $\boldsymbol{\beta}$ ; do vetor  $q_p \times 1$  de efeitos genéticos aditivo direto  $\mathbf{a}_p$ , e do vetor  $q_p \times 1$  de efeitos genéticos aditivos materno  $\mathbf{m}_p$ . Onde  $\mathbf{a}_p$  e  $\mathbf{m}_p$  correspondem aos efeitos dos pais. Neste caso,  $\mathbf{X}_p$ ,  $\mathbf{Z}_{1p}$ , e  $\mathbf{Z}_{2p}$ , são matrizes de incidência que conectam  $\boldsymbol{\beta}$ ,  $\mathbf{a}_p$  e  $\mathbf{m}_p$ , a  $\mathbf{y}_p$ , respectivamente. O subconjunto restante em  $\mathbf{y}_t$ ,  $n_t \times 1$ , são informações de animais que não são pais de nenhum outro animal no conjunto de dados. Do mesmo modo que  $\mathbf{y}_p$ ,  $\mathbf{y}_t$  é modelado em função de  $\boldsymbol{\beta}$  e  $\mathbf{m}_p$ , exceto que  $t$  é utilizado ao invés de  $p$  é usado com índice das matrizes de incidência como verificado na equação 1. Além disso,  $\mathbf{y}_t$  é modelado como uma função linear (através de  $\mathbf{Z}_{1t} \mathbf{P}_t$ ) de  $\mathbf{a}_p$ ;  $\mathbf{Z}_{1t}$  é matriz de incidência  $n_t \times q_p$ ;  $\mathbf{P}_t$  é a matriz  $q_p \times q_p$ , que conecta os efeitos genéticos dos  $q_t$  animais não-pais com os de seus pais. Isto é, em  $\mathbf{P}_t$ , a linha  $j$ , indexada  $j = q_p + 1, q_p + 2, \dots, q$ , conecta o efeito genético do animal não-pai  $j$  com aquele de seu pai  $s_j^*$  e mãe  $d_j^*$ , tal que  $(j, s_j^*)$  e  $(j, d_j^*)$  sejam elementos da matriz  $\mathbf{P}_t$  iguais a 0,5, relacionando o animal  $j$  com seus pais. O vetor de efeitos residuais é composto por  $\mathbf{e}_p$  e  $\mathbf{e}_t$ , para pais e não pais, respectivamente. Adicionalmente, para os animais não-pais com registros próprios, tem-se no vetor  $\boldsymbol{\gamma}_t$  os termos referentes à segregação mendeliana, os quais são relacionados com  $\mathbf{y}_t$  através de  $\mathbf{Z}_{1t}$ .

Assume-se que a matriz de (co)variâncias no vetor residual do modelo animal reduzido é:

$$\mathbf{R} = \text{var} \begin{bmatrix} \mathbf{e}_p \\ \mathbf{Z}_{1t} \boldsymbol{\gamma}_t + \mathbf{e}_t \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{q_p} \sigma_e^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{1t} \boldsymbol{\Omega}_{tt} \mathbf{Z}'_{1t} \sigma_a^2 + \mathbf{I}_{q_t} \sigma_e^2 \end{bmatrix} \quad (2)$$

onde  $\boldsymbol{\Omega}_{tt} = \text{diag} \{ \omega_j \}_{j=q_p+1}^q$  é a matriz diagonal  $q_t \times q_t$ , com o  $j^{\text{th}}$  elemento correspondendo a proporção da variância genética aditiva ( $\sigma_a^2$ ) do animal  $j$  que é devida à segregação mendeliana (Quaas, 1988);  $\sigma_e^2$  é a variância residual.

$$p \left( \begin{bmatrix} \mathbf{a}_p \\ \mathbf{m}_p \end{bmatrix} \mid \mathbf{G} \right) = N \left( \mathbf{0}, \mathbf{G} \otimes \mathbf{A}_{pp} \right) \quad (3)$$

onde  $\mathbf{G} = \begin{bmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{bmatrix}$  é a matriz de (co)variância para efeitos genéticos diretos e maternos;  $\sigma_m^2$  é a variância genética materna;  $\sigma_{am}$  é a covariância entre o efeito genético direto e o materno;  $\mathbf{A}_{pp}$  é a matriz de parentesco médio entre todos os pais, e  $\otimes$  é o produto direto de Kronecker. Por conveniência, uma distribuição a priori uniforme limitada ou normal  $p(\beta)$  pode ser presumida para  $\beta$ , uma densidade à priori de Wishart invertida  $p(\mathbf{G})$  especificada para  $\mathbf{G}$ , e uma densidade gama invertida  $p(\sigma_e^2)$  especificada para  $(\sigma_e^2)$ .

Adicionalmente, temos que:

$$\begin{bmatrix} \mathbf{a}_t \\ \mathbf{m}_t \end{bmatrix} = [\mathbf{I}_2 \otimes \mathbf{P}_t] \begin{bmatrix} \mathbf{a}_p \\ \mathbf{m}_p \end{bmatrix} + \begin{bmatrix} \boldsymbol{\gamma}_t \\ \boldsymbol{\delta}_t \end{bmatrix} \quad (4)$$

onde  $\mathbf{a}_t$  e  $\mathbf{m}_t$  são vetores  $q_t \times 1$  de efeitos genéticos aditivos e maternos, respectivamente, associado à animais sem progênie e  $\boldsymbol{\gamma}_t$  e  $\boldsymbol{\delta}_t$  são vetores referentes a segregação mendeliana, aditivos e maternos, respectivamente, tal que:

$$\begin{bmatrix} \boldsymbol{\gamma}_t \\ \boldsymbol{\delta}_t \end{bmatrix} | \mathbf{G} \sim N \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \mathbf{G} \otimes \boldsymbol{\Omega}_t \right) \quad (5)$$

## 2.2 Modelo hierárquico bayesiano para paternidade incerta

Em populações onde se utilizam acasalamentos com reprodutores múltiplos, há um numero de possíveis pais candidatos para cada animal. Isso se traduz na incerteza no valor de vários elementos de  $\mathbf{P}_t$  para animais sem progênie e vários elementos de  $\mathbf{A}_{pp}$  para os pais.

Esse modelo pode ser descrito em quatro estágios, o primeiro estagio considera a paternidade incerta dos animais não pais  $q_t$ , sendo  $j = q_p + 1, q_p + 2, \dots, q$ , associado com os  $n_t$  dados de  $\mathbf{y}_t$ .

Considerando que  $\mathbf{Z}_1 = \begin{bmatrix} \mathbf{Z}_{1p} \\ \mathbf{Z}_{1t} \mathbf{P}_t \end{bmatrix}$  então se um animal não pai  $j$  tem

incerteza de paternidade, isto se traduz em incerteza no elemento  $(j, s_j^*)$  de  $\mathbf{P}_t$ , ou equivalente o elemento  $s_j^*$  de  $\mathbf{z}'_{1ij}$  é desconhecido;  $\mathbf{z}'_{1ij}$  é a linha de  $\mathbf{Z}_1$  com a mesma posição de  $y_{ij}$  em  $\mathbf{y}$ .

Supondo que o animal  $j$  tenha  $v_j$  possíveis pais com identificação

$$\mathbf{s}_j = \left\{ s_j^{(1)}, s_j^{(2)}, \dots, s_j^{(v_j)} \right\}.$$

A distribuição condicional de  $y_{ij}$ , dado a atribuição de paternidade  $s_j^* = s_j^{(k)}$   $1 \leq k \leq v_j$ , para o animal  $j$  e todos os demais parâmetros é dada por:

$$y_{ij} | \boldsymbol{\beta}, \mathbf{a}_p, \mathbf{m}_p, s_j^* = s_j^{(k)}, \sigma_a^2, \sigma_e^2 \sim N \left( \mathbf{x}'_{ij} \boldsymbol{\beta} + 0.5a_{s_j^{(k)}} + 0.5a_{d_j} + \mathbf{z}'_{2ij} \mathbf{m}_p, \sigma_e^2 + \omega_j^{(k)} \sigma_a^2 \right)$$

$$i = n_p + 1, n_p + 2, \dots, n; j = q_p + 1, q_p + 2, \dots, q. \quad (6a)$$

Neste caso,  $\mathbf{x}'_{ij}$  e  $\mathbf{z}'_{2ij}$  são linhas de  $\mathbf{X}$  e  $\mathbf{Z}_2$  coincidentes com o endereço de  $y_{ij}$  em  $\mathbf{y}$ . Quando o animal  $j$  tem pais conhecidos,  $v_j = 1$ , tal que  $s_j^*$  não é aleatório.

Note que a condição de  $d_j^*$  conhecido (mãe identificada) é presumida na metodologia, entretanto  $s_j^* = s_j^{(v_j)}$  é modelado explicitamente uma vez que o valor de  $s_j^*$  pode ser desconhecido. Esta incerteza é refletida na equação (6a) no termo  $\omega_j^{(k)} = \omega_j \Big|_{s_j^* = s_j^{(k)}}$  em que  $\omega_j^{(k)}$  é a fração de variân-

cia genética atribuída à segregação mendeliana para o animal  $j$  que é a função de coeficiente de endogamia, dada a atribuição de paternidade  $s_j^* = s_j^{(k)}$ .

Caso seja considerada a possibilidade de se ter pelo menos um dos animais com progênie, de 1 a  $q_p$ , com paternidade incerta, alguns elementos de  $\mathbf{A}_{pp}$  também serão desconhecidos. Mesmo assim, a distribuição amostral de  $y_{ij}$ , dos animais com progênie,  $j = 1, 2, \dots, q_p$ , não é condicional às atribuições de paternidade:

$$y_{ij} \mid \boldsymbol{\beta}, \mathbf{a}_p, \mathbf{m}_p, \sigma_e^2 \sim N(\mathbf{x}'_{ij}\boldsymbol{\beta} + a_j + \mathbf{z}'_{2ij}\mathbf{m}_p, \sigma_e^2)$$

$$i = 1, 2, \dots, n_p; \quad j = 1, 2, \dots, q_p. \quad (6b)$$

O segundo estágio do modelo que inclui animais com incerteza de paternidade é a estrutura a priori de  $\mathbf{a}_p$  e  $\mathbf{m}_p$  na equação 3.

A decomposição de  $\mathbf{A}_{pp}^{-1}$  é apresentada por Henderson (1976) e Quaas (1988):

$\mathbf{A}_{pp}^{-1} = \mathbf{T}_p \boldsymbol{\Omega}_{pp}^{-1} \mathbf{T}'_p$  onde  $\mathbf{T}_p$  é uma matriz triangular inferior  $q_p \times q_p$  e

$\boldsymbol{\Omega}_{pp} = \text{diag} \left\{ \omega_j \right\}_{j=1}^{q_p}$  é a diagonal  $q_p \times q_p$  da matriz análoga a  $\boldsymbol{\Omega}_{tt}$ , mas com cada elemento correspondente a fração da  $\sigma_a^2$  devido a segregação mendeliana de cada animal com progênie  $j$ .

Todos os elementos da diagonal da  $\mathbf{T}_p$  são iguais a 1 com outros dois elementos por linha,  $(j, s_j^*)$  e  $(j, d_j^*)$ , sendo iguais a -0.5, se os pais correspondentes  $s_j^*$  e  $d_j^*$  do animal  $j$  forem conhecidos, para  $j = 1, 2, \dots, q_p$ . Conseqüentemente,  $|\mathbf{A}_{pp}^{-1}| = |\mathbf{T}_p| |\boldsymbol{\Omega}_{pp}^{-1}| |\mathbf{T}'_p| = |\boldsymbol{\Omega}_{pp}^{-1}|$  já que  $|\mathbf{T}_p| = 1$ . Dados

esses resultados, a densidade a priori conjunta de  $\mathbf{a}_p$  e  $\mathbf{m}_p$  condicional em

$A_{pp}$ , pode ser escrita da seguinte maneira:

$$P\left(\begin{matrix} \mathbf{a}_p \\ \mathbf{m}_p \end{matrix} \middle| \mathbf{G}, \mathbf{A}_{pp}\right) \propto |\mathbf{G}|^{\frac{p}{2}} |\boldsymbol{\Omega}_{pp}^{-1}| \exp\left(-0.5\left(\mathbf{a}'_p \mathbf{T}_p \boldsymbol{\Omega}_{pp}^{-1} \mathbf{T}'_p \mathbf{a}_p g^{11} + 2\mathbf{a}'_p \mathbf{T}_p \boldsymbol{\Omega}_{pp}^{-1} \mathbf{T}'_p \mathbf{m}_p g^{12} + \mathbf{m}'_p \mathbf{T}_p \boldsymbol{\Omega}_{pp}^{-1} \mathbf{T}'_p \mathbf{m}_p g^{22}\right)\right) \quad (7)$$

onde  $g_{ij}$  é o elemento  $(i,j)$  de  $\mathbf{G}^{-1}$  para  $i,j = 1,2$ . Considere  $\mathbf{t}'_j$  a  $j$ -ésima linha de  $\mathbf{T}_p$ ; Então, os termos da segregação mendeliana, aditiva e maternal, respectivamente, podem ser apresentados da seguinte forma:

$$\gamma_j = \mathbf{t}'_j \mathbf{a}_p = a_j - .5a_{s_j^*} - .5a_{d_j^*} \quad \text{e} \quad \delta_j = \mathbf{t}'_j \mathbf{m}_p = m_j - .5m_{s_j^*} - .5m_{d_j^*}, \text{ para}$$

$j = 1, \dots, q_p$ . Se os pais de  $j$ ,  $s_j^*$  e  $d_j^*$  são completamente desconhecidos, as suas contribuições correspondentes,  $a_{s_j^*}$  e  $a_{d_j^*}$  para  $\gamma_j$ ,  $m_{s_j^*}$  e  $m_{d_j^*}$  para

$\delta_j$  serão iguais a 0. Considerando que  $\mathbf{s}_p^* = \left\{s_j^*\right\}_{j=1}^{q_p}$  denota o vetor

aleatório de atribuições de paternidade para os animais com progênie e

$\mathbf{s}_p^{(k)} = \left\{s_j^{(k)}\right\}_{j=1}^{q_p}$  um valor possível ou realização de  $\mathbf{s}_p^*$  advinda do conjunto

$\mathbf{S}_p = \left\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \dots, \mathbf{s}_{q_p}\right\}$  e o  $j$ -ésimo elemento de  $\mathbf{s}_p^{(k)}$  é um dos  $v_j$  pais

possíveis escolhido de  $\mathbf{s}_j = \left\{s_j^{(1)}, s_j^{(2)}, \dots, s_j^{(v_j)}\right\}$  para  $j = 1, 2, \dots, q_p$ .

Podemos então reescrever (7), condicionando explicitamente nas atribuições de paternidade, como segue:

$$P\left(\begin{matrix} \mathbf{a}_p \\ \mathbf{m}_p \end{matrix} \middle| \mathbf{G}, \mathbf{s}_p^* = \mathbf{s}_p^{(k)}\right) \propto |\mathbf{G}|^{\frac{q_p}{2}} \prod_{j=1}^{q_p} \left( (\omega_j^{(k)})^{-1} \exp\left(-0.5 (\omega_j^{(k)})^{-1} \left( (\gamma_j^{(k)})^2 g^{11} + (\delta_j^{(k)})^2 g^{22} + 2\gamma_j^{(k)} \delta_j^{(k)} g^{12} \right) \right) \right) \quad (8)$$

Onde  $\delta_j^{(k)} = \delta_j \big|_{s_j^* = s_j^{(k)}}$  e  $\gamma_j^{(k)} = \gamma_j \big|_{s_j^* = s_j^{(k)}}$  indicam a dependência dos termos de segregação mendeliana na atribuição de paternidade  $s_j^* = s_j^{(k)}$ .

O terceiro estágio do modelo especifica as probabilidades a priori de cada um dos  $v_j$  touros ser o pai verdadeiro do animal  $j$ . Similarmente aos

animais com progênie, temos que:  $\mathbf{s}_t^* = \left\{s_j^*\right\}_{j=q_p+1}^q$  denota o vetor aleatório

de atribuições de paternidade para animais sem progênie e  $\mathbf{s}_t^{(k)} = \left\{s_j^{(k)}\right\}_{j=q_p+1}^q$

denota uma realização particular de  $\mathbf{s}_t^*$  do conjunto  $\mathbf{S}_t = \{s_{q_p+1}, s_{q_p+2}, \dots, s_{q_t}\}$ .

Para todos os  $q$  animais, tem-se que  $\mathbf{s}^{(k)} = \begin{bmatrix} \mathbf{s}_p^{(k)} \\ \mathbf{s}_t^{(k)} \end{bmatrix} = \{s_j^{(k)}\}_{j=1}^q$  é uma realização

de  $\mathbf{s}^* = \begin{bmatrix} \mathbf{s}_p^* \\ \mathbf{s}_t^* \end{bmatrix} = \{s_j^*\}_{j=1}^q$  do conjunto  $\mathbf{S} = \{\mathbf{S}_p, \mathbf{S}_t\}$ . A probabilidade que  $s_j^{(k)}$  seja o pai do animal  $j$  é definida por  $\pi_j^{(k)} = \text{Prob}(s_j = s_j^{(k)})$  para  $k=1, 2, \dots, v_j$

tal que  $\sum_{k=1}^{v_j} \pi_j^{(k)} = 1$ . Para animais com paternidade conhecida, existe somente um touro candidato  $s_j^* \equiv s_j^{(1)}$  tal que  $\pi_j^{(1)} = 1$ , e então constante.

O conjunto de probabilidades  $\boldsymbol{\pi}_j = \{\pi_j^{(1)}, \pi_j^{(2)}, \dots, \pi_j^{(v_j)}\}$  para os  $v_j$  pais candidatos para o animal  $j$ , pode ser atribuído usando informações externas (e.x. marcadores moleculares). O conjunto inteiro de probabilidades

des  $\boldsymbol{\pi} = \{\boldsymbol{\pi}_j\}_{j=q_b+1}^q$  raramente é conhecido com absoluta certeza, podendo ser considerado como quantidades aleatórias da distribuição de *Dirichlet*, no quarto estágio do modelo:

$$p(\boldsymbol{\pi}_j | \boldsymbol{\alpha}_j) \propto \prod_{k=1}^{v_j} (\pi_j^{(k)})^{\alpha_j^{(k)}} \quad (9)$$

onde  $\boldsymbol{\alpha}_j = \{\alpha_j^{(k)}\}_{k=1}^{v_j}$ ,  $\alpha_j^{(k)} > 0$  para  $k = 1, 2, \dots, v_j$  e  $\pi_j^{(v_j)} = 1 - \sum_{k=1}^{v_j-1} \pi_j^{(k)}$ .

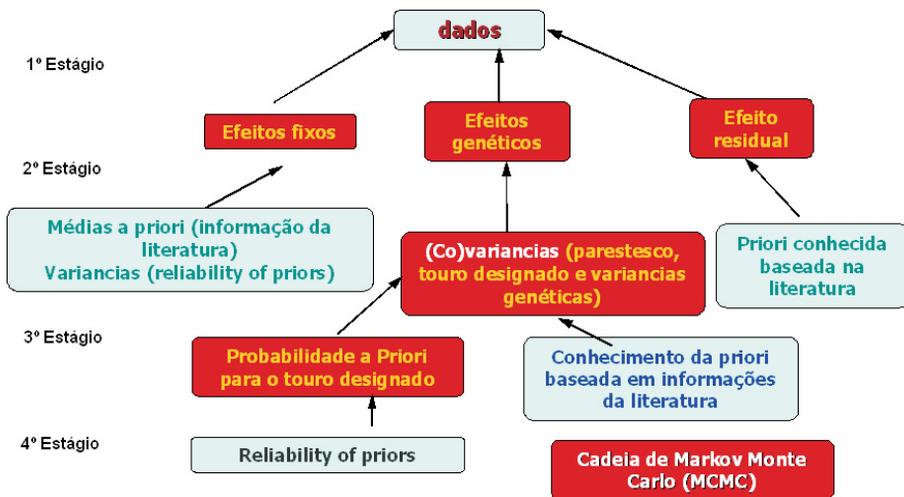
Especificações do conjunto de hiper-parâmetros  $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_j\}_{j=q_b+1}^q$  podem

ser baseadas na avaliação da confiabilidade de uma fonte de informação externa sobre as probabilidades a priori de cada touro designado. Do produto dos quatro estágios apresentados acima resulta a densidade a posteriori conjunta de todos os parâmetros desconhecidos do modelo.

$$\begin{aligned} & p(\boldsymbol{\beta}, \mathbf{a}_p, \mathbf{m}_p, \mathbf{s}^* = \mathbf{s}^{(k)}, \boldsymbol{\pi}, \mathbf{G}, \sigma_e^2 | \mathbf{y}) \propto \\ & \prod_{i=1}^{n_p} p(y_{ij} | \boldsymbol{\beta}, \mathbf{a}_p, \mathbf{m}_p, \sigma_e^2) \prod_{i=n_p+1}^n p(y_{ij} | \boldsymbol{\beta}, \mathbf{a}_p, \mathbf{m}_p, \mathbf{s}_i^* = \mathbf{s}_i^{(k)}, \sigma_a^2, \sigma_e^2) \\ & p(\mathbf{a}_p, \mathbf{m}_p | \mathbf{s}_p^* = \mathbf{s}_p^{(k)}, \mathbf{G}) p(\boldsymbol{\beta}) \text{Prob}(\mathbf{s}^* = \mathbf{s}^{(k)} | \boldsymbol{\pi}) p(\boldsymbol{\pi} | \boldsymbol{\alpha}) p(\mathbf{G}) p(\sigma_e^2) \end{aligned} \quad (10)$$

onde  $\text{Prob}(\mathbf{s}^* = \mathbf{s}^{(k)} | \boldsymbol{\pi}) = \prod_{j=q_b+1}^q \text{Prob}(s_j^* = s_j^{(k)} | \boldsymbol{\pi}_j) = \prod_{j=q_b+1}^q \prod_{k=1}^{v_j} (\pi_j^{(k)})^{I_j^{(k)}}$ , sendo que  $I_j^{(k)} = 1$  se  $s_j^* = s_j^{(k)}$  e  $I_j^{(k)} = 0$ , em caso contrário. Finalmente, temos que  $p(\boldsymbol{\pi} | \boldsymbol{\alpha}) = \prod_{j=q_b+1}^q p(\boldsymbol{\pi}_j | \boldsymbol{\alpha}_j) = \prod_{j=q_b+1}^q \prod_{k=1}^{v_j} (\pi_j^{(k)})^{\alpha_j^{(k)}}$

A Figura 1 descreve resumidamente os quatro estágios do HIER.



**Figura 1.** Descrição dos quatro estágios do modelo hierárquico bayesiano desenvolvidos por Cardoso e Tempelman (2003). Caixas de texto com fundo vermelho representam quantidades desconhecidas (parâmetros do modelo) e caixa com fundo azul turquesa representam oportunidades de incorporação de informação externa aos dados.

As distribuições condicionais completas (DCC) dos parâmetros desconhecidos da equação 10, necessárias para conduzir inferências MCMC e detalhes sobre estratégias de amostragem foram derivadas no apêndice do trabalho de Cardoso e Tempelman (2003) e uma exposição completa da implementação de MCMC no melhoramento animal é apresentada por Wang et al. (1994).

## 2.3 Validação por Simulação

Um estudo de simulação foi desenvolvido para comparar dois modelos que predizem o mérito genético quando alguns animais têm incerteza de paternidade. O primeiro é o modelo hierárquico proposto nesse trabalho, apresentado na seção 2.2, que infere sobre essa incerteza utilizando dados fenotípicos, o outro modelo, é baseado na utilização da matriz de parentesco médio (ANRM) proposta por Henderson.

Dois tipos de característica foram utilizadas e, para cada uma delas, dez conjuntos de dados foram gerados. A característica 1, possuía herdabilidade direta média ( $h^2_d = 0,30$ ), herdabilidade materna média ( $h^2_m = 0,20$ ) e correlação entre efeito direto e materno baixa e negativa ( $r_{am} = -0,20$ ), semelhante a peso à desmama (PD), enquanto que a característica 2, possuía alta herdabilidade direta ( $h^2_d = 0,50$ ) e não possuía  $h^2_m$ , como é o caso do ganho de peso pós-desmama (GPD). As variâncias residuais ( $\sigma_e^2$ ) foram de 60 e 50, respectivamente para as características 1 e 2.

Como observado na Figura 2, cada população incluía 80 touros, 400 matrizes (480 pais) e 2000 animais que não eram pais, todos descendentes de uma população base composta por 20 touros e 100 vacas. A partir desses animais bases, cinco gerações foram criadas. Quinze machos e 75 fêmeas foram selecionados ao acaso em cada geração para ser pais na próxima geração. Além disso, cinco touros e 25 fêmeas das gerações prévias foram retidos como reprodutores, tal que um total de 20 touros e 100 vacas foram usados em reprodução para cada geração. A população foi estruturada tendo sobreposição de gerações.

A probabilidade de um animal ter paternidade incerta foi de 0,30. Neste caso, ele foi indicado ao acaso para um de seis possíveis grupos de touros múltiplos em cada uma das cinco gerações. Esses grupos apresentaram seis tamanhos diferentes:  $v_j = 2,3,4,6,8$  ou 10 touros candidatos. Assim que o grupo foi escolhido, um macho de cada grupo foi selecionado para ser o pai verdadeiro com probabilidades iguais ( $1/v_j$ ) ou probabilidades diferentes em relação aos demais candidatos. Essa última hipótese teve a intenção de representar a situação de macho dominante, comum em gado de corte.

Os cinco touros selecionados para serem pais das próximas gerações possuíam somente progênes conhecidas. Adicionalmente, dez touros foram usados em grupos de reprodutores e tendo também progênie conhecida, enquanto os cinco touros remanescentes tinham somente progênie não identificada. Um grupo de três touros em cada população foi formado com touros tendo apenas progênes não identificadas, com o propósito de comparar a performance dos dois modelos, caso os touros tivessem apenas suas próprias informações de desempenho e o pedigree como fonte de informação para sua avaliação genética, além da progênie com paternidade incerta. Todos os outros grupos de acasalamento tiveram pelo menos, um touro que era conhecido como pai de outro animal. A intenção desse procedimento foi a de imitar as situações observadas em algumas fazendas que participam de programas de avaliação genética no Brasil, que selecionam seus próprios touros jovens para servir ao rebanho por monta natural (MN) e também, para coletar sêmen para utilizar na inseminação artificial (IA). Além disso, existe a importação de material genético externo, especialmente via IA. Nesse cenário, os touros podem ser divididos em três categorias diferentes: (1) touros tendo somente progênes conhecidas (ex: importação de touros para IA); (2) touros que possuíam tanto progênes conhecidas como desconhecidas, de tal modo que os melhores touros do rebanho foram utilizados para IA, acasalamento por MN com progênie conhecida ou também em grupos de reprodutores múltiplos em estação de monta; (3) touros tendo somente atribuições desconhecidas de progênie.

Somente um registro foi gerado por animal e para ambas as características, a média geral foi igual a 100 e o fator de efeito fixo com três níveis, tendo valores de 25, -25 e 0, foi atribuídos ao acaso para gerar os dados individuais.



observou-se 2.702 (57,4%) animais com certeza na paternidade, 700 com incerteza de paternidade e 1.301 (27,7%) animais da população base. Dentre os animais com incerteza de paternidade, 6,9% eram pais e 20,5% não tinham progênie. O número de touros candidatos nos grupos de reprodutores múltiplos, com seus respectivos número de progênies descrito entre parênteses por características (PD/GPD) foi 2 (57/23), 3 (2/2), 4 (234/167), 5(207/117), 6 (272/176), 10 (135/43), 12(16/9) e 17 (248/163).

**Tabela 1.** Descrição dos conjuntos de dados do rebanho Hereford para o peso à desmama (PD) e o Ganho de peso pós-desmama (GPD).

	<b>PD</b>	<b>GPD</b>
	<b>Média ± DP</b>	<b>Média ± DP</b>
<b>N</b>	5.339	3.402
<b>AGA</b>	6905	4703
<b>Valor observado, kg</b>	172,6 ± 35,6	110,7 ± 30,4
<b>Idade do animal, dias</b>	201,7 ± 25,5	422,3 ± 37,2
<b>Peso ao nascer, dias</b>	36,2 ± 5,9	NA
<b>Período de teste, dias</b>	NA	218,0 ± 32,5
<b>Idade da mãe, anos</b>	4,9 ± 2,6	5,2 ± 2,6
<b>Filhos por touro</b>	63,0 ± 83,9	42,0 ± 61,4

DP= desvio padrão; N= número total de indivíduos com registro para cada característica; AGA = animais geneticamente avaliados, incluindo animais da população base, sem arquivo de pedigree; NA = Não se aplica, covariável não considerada no modelo para a característica.

## 2.5 Inferência Bayesiana

### 2.5.1 Análises do estudo de simulação

As dez réplicas de cada uma das duas características foram analisadas usando três modelos diferentes:

- (1) HIER: modelo hierárquico Bayesiano considerando a incerteza nas atribuições de paternidade de animais filhos de reprodutores múltiplos, proposto na sessão 2.2.
- (2) ANRM: Modelo de efeitos lineares mistos baseados na matriz de parentesco médio proposta por Henderson (1988). As probabilidades são iguais e fixas, designadas a cada pai candidato do animal que tem para paternidade incerta.
- (3) TRUE: modelo misto utilizando o pedigree completo (sem paternidade incerta). Esse modelo foi incluído para servir como controle positivo para os outros dois modelos.

No HIER, para cada pai candidato de animais com paternidade desconhecida foram especificadas probabilidades a priori iguais à inversa do número de touros candidatos, através da distribuição a priori de Dirichlet (4º estágio do modelo). O hiper-parâmetro dessas densidades a priori foram iguais à inversa do número de touros candidatos dentro de cada grupo

$\alpha_j^{(k)} = \frac{1}{v_j}$ . Distribuições a priori difusas qui-quadrado invertidas foram

especificada para cada componente de variância, ex.  $\sigma^2 \sim (\nu \sigma_o^2) \chi_{\nu}^{-2}$ , onde  $\sigma_o^2$  era o valor verdadeiro para o respectivo componente de variância com  $\nu = 8$  grau de confiança no valor a priori. Adicionalmente, para análises do PD, uma densidade a priori Wishart invertida

$\mathbf{G} \sim IW_2(\nu \mathbf{G}_o, \nu)$  com  $E(\mathbf{G}) = \frac{\nu}{\nu - 3} \mathbf{G}_o$  foi atribuída a  $\mathbf{G}$ , onde  $\mathbf{G}_o$

representou o valor verdadeiro de  $\mathbf{G}$ , novamente com  $\nu = 8$ .

### 2.5.2 Análises de dados do rebanho Hereford

A característica GPD foi modelada como uma função linear de efeitos fixos (ex. efeitos com a priori subjetivas uniformes limitadas) e efeitos aleatórios (ex. efeitos com distribuição a priori normal multivariada). Nos efeitos fixos foram incluídos efeitos de sexo, idade da vaca (2-12 anos) linear e quadrático, idade do bezerro (linear e quadrático) no fim do período teste (305-658 dias), efeito do período do teste pós-desmama (111-453 dias) linear e quadrático. Os efeitos aleatórios considerados foram: o efeito genético aditivo (1.932 pais), com correlação determinada pela

matriz de parentesco e o efeito de grupos de contemporâneos (ano-estação-grupo de manejo; 237 níveis). Neste caso, foram utilizados somente os modelos HIER e ANRM, uma vez que não se conhecia o pedigree verdadeiro do rebanho Hereford, impossibilitando o uso de TRUE. Para ambos os modelos, foram computados cadeias de 70.000 ciclos, sendo os 4.000 primeiros ciclos descartados.

O modelo estatístico para PD incluiu efeitos fixos de sexo, efeito linear e quadrático para idade do bezerro a desmama (100-293 dias), efeito linear do peso ao nascer (16-60 Kg). As variações desses efeitos, embora amplas, estão dentro do padrão observado para dados de outros rebanhos criados em condições extensivas similares e avaliados pelo PROMEBO<sup>4</sup>. Efeitos aleatórios incluíram efeitos genéticos aditivos e maternos com distribuição normal multivariada conjunta, efeito de ambiente permanente da mãe (2.281 níveis) com variância  $\sigma_{pe}^2$  e efeitos de grupos de contemporâneos (ano-estação-grupo de manejo – 96 níveis) com variância  $\sigma_{cg}^2$ . A cadeia de MCMC foi de 180.000 e 120.000 ciclos foram rodados para o HIER e ANRM, respectivamente, sendo em ambos os casos retirados os 4.000 primeiro ciclos.

Foi surpreendente a pequena velocidade com que a cadeia de MCMC para PD, baseado no modelo animal reduzido (MAR), percorreu todo o espaço amostral, especificadamente para o HIER de Cardoso e Tempelman (2003). Portanto, adotou-se o modelo animal completo para HIER e ANRM em análises de PD, entretanto o MAR foi mantido para o HIER e ANRM nas análises de GPD. Os hiper-parâmetros das distribuições qui-quadrado escalonadas invertidas a priori para os componentes de variâncias escalares foram, para GPD,  $\sigma_o^2 = 200$  para  $\sigma_e^2$ ,  $\sigma_o^2 = 100$  para  $\sigma_a^2$  e  $\sigma_o^2 = 800$  para  $\sigma_{cg}^2$ , enquanto para PD,  $\sigma_o^2 = 200$  para  $\sigma_e^2$ ,  $\sigma_o^2 = 100$  para  $\sigma_{pe}^2$  e  $\sigma_o^2 = 600$  para  $\sigma_{cg}^2$ . Além disso, o prior da matriz de (co)variância genética aditiva para PD foi especificada como

$$\mathbf{G} \sim IW_2 \left( \nu \begin{bmatrix} 100 & -5 \\ -5 & 50 \end{bmatrix}, \nu \right)$$
. Como no estudo de simulação,  $\nu = 8$  para todos os componentes de (co)variâncias e uma distribuição Dirichlet a priori foi especificada para as probabilidade de atribuição de paternidade com os

<sup>4</sup>Programa de Melhoramento Genético de Bovinos de Carne da Associação Nacional de Criadores Herdbook Collares, Pelotas/RS;

mesmos hiper-parâmetros especificados. Para cada componente de variância, a seqüência monotônica inicial (Geyer, 1992) foi usada para determinar o tamanho efetivo da amostra (TEA), como uma estimativa do número de amostras independentes com conteúdo de informação correspondente ao das amostras dependentes que foram salvas (Sorensen et al., 1995). O comprimento de cadeia MCMC foi determinado tal que todos os parâmetros tivessem TEA de pelo menos 100.

A inferência baseada em MCMC foi implementada utilizando-se o programa INTERGEN (Cardoso, 2007). Médias, medianas, percentis entre 2,5 a 97,5% e desvio padrão a posteriori dos parâmetros foram obtidos pela densidade posteriori marginal. Medias a posteriori e desvio padrão de cada efeito genético animal individual foi também determinado para todos os modelos.

## 2.6. Critérios para escolha do melhor modelo

Tanto para dados simulados como os do rebanho de Hereford, para verificar o modelo de melhor ajuste foram utilizados o critério de informação da deviance (DIC) (Spiegelhalter et al., 2002) e pseudo fatores de Bayes (PBF) (Gelfand, 1996).

Para comparar dois modelos  $M_1$  e  $M_2$ , o correspondente PBF foi definido como:

$$PBF_{1,2} = \prod_{i=1}^n \frac{p(y_{ij} | \mathbf{y}_{(-ij)}, M_1)}{p(y_{ij} | \mathbf{y}_{(-ij)}, M_2)}$$

Onde  $p(y_{ij} | \mathbf{y}_{(-ij)}, M_r)$  é a ordenada preditiva condicional (CPO) para observação  $y_{ij}$ , destinada ser a densidade de validação cruzada, que sugere quais valores de  $y_{ij}$  são prováveis quando o modelo  $M_r$  é ajustado a todas as outras observações  $\mathbf{y}_{(-ij)}$  exceto  $y_{ij}$ . Uma aproximação de MCMC para o CPO do modelo  $M_r$  com parâmetros  $\theta$  é obtido pela média harmônica de  $G$  ciclos de MCMC:

$$p\left(y_{ij} \mid \mathbf{y}_{(-ij)}, M_r\right) \approx \frac{1}{\frac{1}{G} \sum_{l=1}^G p^{-1}\left(y_{ij} \mid \boldsymbol{\theta}^{(l)}, M_r\right)}$$

O DIC é composto por uma medida de ajuste global, a média a posteriori da deviance, e uma penalização para a complexidade do modelo. A deviance do modelo  $M_r$  segundo Spiegelhalter et al. (2002) é estimada por  $\bar{D}_r = \frac{1}{G} \sum_{l=1}^G -2 \log p\left(\mathbf{y} \mid \boldsymbol{\theta}^{(l)}, M_r\right)$ . A complexidade do modelo  $M_r$  é determina-

da pelo número efetivo de parâmetros  $p_{D(r)} = \bar{D}_r - D_r(\bar{\boldsymbol{\theta}})$ , onde  $D_r(\bar{\boldsymbol{\theta}}) = -2 \log p\left(\mathbf{y} \mid \bar{\boldsymbol{\theta}}, M_r\right)$  com  $\bar{\boldsymbol{\theta}}$  sendo a media a posteriori do  $\boldsymbol{\theta}$ , isto é,

$p_{D(r)}$  representa a diferença entre a média a posteriori da deviance e a deviance baseada na média a posteriori dos parâmetros sob o modelo  $M_r$ . O DIC para o modelo  $M_r$  é determinado por:

$$DIC_r = \bar{D}_r + p_{D(r)}$$

Menores valores de DIC são indicativos de melhor ajuste do modelo, enquanto que a relação de PBF menor que um para ANRM dividido pelo HIER favorece o HIER.

Além disso, no caso do estudo de simulação, os métodos estudados foram comparados pelo quadrado médio do erro de predição (MSEP), o viés médio de predição (MBIAS) e a correlação de ordem de Spearman entre estimativas e valores genéticos verdadeiros. O MSEP para cada modelo foi estimado como  $\sum_{h=1}^{10} \sum_{j=1}^q (\hat{u}_{hj} - u_{hj})^2 / q / 10$ , onde 10 é o

número de réplicas; q é o número total de animais com paternidade incerta por réplica;  $\hat{u}_{hj}$  ( $u_{hj}$ ) é a estimativa do (verdadeiro) efeito genético aditivo ou materno do animal j na repetição h. O MBIAS foi estimado como  $\sum_{h=1}^{10} \sum_{j=1}^q (\hat{u}_{hj} - u_{hj}) / q / 10$ . Variáveis descrevendo incerteza de

paternidade, especificamente,  $s_j^*$  e  $\pi_j^{(k)}$ , foram analisadas separadamente animais com e sem progênie, uma vez que os com progênie tem uma maior quantidade de informações para seu mérito genético comparado com os sem progênie. Comparações baseadas nas estimativas dos méritos genéticos sob os três diferentes modelos foram realizadas usando o

test  $t$ .

### 3. Resultados

#### 3.1 Inferência sobre dados simulados

Para os dados simulados, a variável indicativa  $s_j^*$  ou o parâmetro  $\pi_j^{(k)}$ , podem ser usados para inferir sobre a incerteza da atribuição de paternidade para o animal  $j$ . A inferência sobre a probabilidade do pai verdadeiro do animal com paternidade incerta pelo HIER foi baseada na frequência das amostras de MCMC nas quais  $s_j^*$  foi igual ao touro verdadeiro, designada como  $\text{Prob}(s_j^* = s_j^{(true)} | \mathbf{y})$  e pela determinação de  $E(\pi_j^{(true)} | \mathbf{y})$ , a média a posteriori de  $\pi_j^{(true)}$ , isto é, a probabilidade do parâmetro identificando  $s_j^{(true)}$  o touro verdadeiro de  $j$ . Esses resultados são apresentados separadamente para animais com paternidade incerta com (pais) e sem progênie (não pais) na tabela 2 para característica 1 e 2.

A probabilidade média a posteriori atribuída ao touro verdadeiro (ex. baseado na  $\text{Prob}(s_j^* = s_j^{(true)} | \mathbf{y})$ ) foi entre 1 a 10% maior que a sua

respectiva priori ( $1/v_j$ ) para característica 1 e entre 4 a 13% maior que as prioris da característica 2. A inferência sobre a paternidade incerta usando  $\text{Prob}(s_j^* = s_j^{(true)} | \mathbf{y})$  tem uma performance levemente superior que

a baseada em  $E(\pi_j^{(true)} | \mathbf{y})$ . A maior diferença entre a média a posteriori e as probabilidades a priori na característica 2 pode ser resultado da alta herdabilidade. Essa diferença foi estatisticamente significativa ( $P < 0,05$ ).

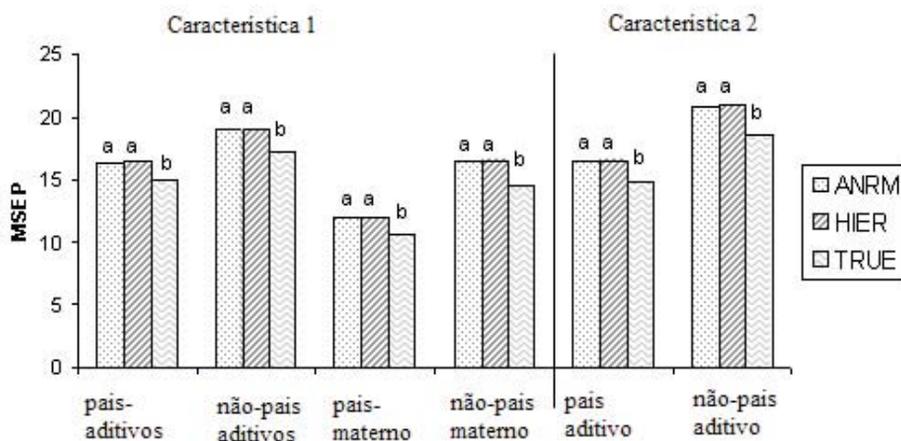
**Tabela 2.** Médias a posteriori das probabilidades dos pais verdadeiros  $(E(\pi_j^{(true)} | \mathbf{y}))$  e probabilidade do pai verdadeiro ser designado ao indivíduo  $j$   $(\text{Prob}(s_j^* = s_j^{(true)} | \mathbf{y}))$  para características 1 e 2 pelo tamanhos de grupos de reprodutores múltiplos de animais com (pais) e sem progênie (não pais).

Parâmetros	Categoria do Animal	Tamanho do grupo de reprodutor múltiplo					
		2	3	4	6	8	10
<i>Característica 1</i>							
$(E(\pi_j^{(true)}   \mathbf{y}))$	Pais	0,513	0,341	0,259	0,175	0,126 <sup>a</sup>	0,105
$(E(\pi_j^{(true)}   \mathbf{y}))$	Não-pais	0,509	0,339	0,259	0,172	0,130	0,103
$\text{Prob}(s_j^* = s_j^{(true)}   \mathbf{y})$	Pais	0,525	0,349	0,269	0,183	0,127	0,110
$\text{Prob}(s_j^* = s_j^{(true)}   \mathbf{y})$	Não-pais	0,517	0,345	0,268	0,178	0,134	0,105
<i>Característica 2</i>							
$(E(\pi_j^{(true)}   \mathbf{y}))$	Pais	0,510	0,343	0,265	0,177	0,132	0,105
$(E(\pi_j^{(true)}   \mathbf{y}))$	Não-pais	0,520	0,346	0,270	0,179	0,134	0,106
$\text{Prob}(s_j^* = s_j^{(true)}   \mathbf{y})$	Pais	0,521	0,352	0,280	0,188	0,138	0,111
$\text{Prob}(s_j^* = s_j^{(true)}   \mathbf{y})$	Não-pais	0,540	0,360	0,289	0,191	0,143	0,111

<sup>a</sup>Probabilidade a posteriori não é diferente estatisticamente para diferentes tamanhos de grupos a  $\alpha = 0,05$

Probabilidades consistentemente mais altas foram designadas ao  $s_j^{(true)}$  pelo HIER indicando que o modelo tende a inferir na direção do modelo correto. Entretanto, a pequena magnitude dessas diferenças sugere que fenótipos possam não conter informação suficiente para inferir precisamente sobre a paternidade nessas duas características. A média de  $\text{Prob}(s_j^* = s_j^{(true)} | \mathbf{y})$  para grupos de acasalamentos de tamanho  $v_j = 3$ , formados por touros exclusivamente com progênie desconhecida foi 0,348 para característica 1 e 0,360 para característica 2. Essas probabilidades foram consistentes com aquelas determinadas para outros grupos de touros  $v_j = 3$ , mas incluindo touros que tinham também progênie conhecida. O HIER se comportou de modo similar em termos de probabilidade de atribuição de paternidade para animais com ou sem progênie conhecida.

Em termos de MBIAS, nenhum dos três modelos foi significativamente diferente para cada uma de todas as situações analisadas, assim esses resultados não são apresentados no presente trabalho. A MSEP e a correlação de ordem de efeitos genéticos aditivos e maternos são apresentadas nas Figuras 3 e 4, respectivamente. Como esperado, o MSEP foi sempre menor e correlações de ordem maiores para TRUE comparado com HIER e ANRM, verificando que a utilização de acasalamentos com reprodutores múltiplos apresenta efeitos negativos na acurácia nas avaliações genéticas (Sullivan, 1995).



**Figura 3.** Quadrado médio do erro de predição (MSEP) dos efeitos aditivos e maternos de animais pais e não pais com incerteza de paternidade para as características 1 e 2 sobre os três modelos. 1) HIER: Modelo de paternidade incerta (Cardoso e Tempelman, 2003); 2) ANRM: baseado na matriz média de parentesco de Henderson; 3) TRUE: baseado no conhecimento de todos os pais como controle positivo. Barras com mesma letra, não apresentaram diferença significativa entre si a  $\alpha=0,05$ .

Médias a posteriori de efeitos genéticos diretos e maternos foram similares para HIER e ANRM, com diferença não significativa em MSEP e correlação de ordem sobre essas médias a posteriori entre esses modelos. Houve, no entanto, uma tendência de se ter uma menor MSEP e maior correlação de ordem sob o HIER para animais com paternidade incerta. Parece não haver informação suficiente nos dados simulados para discriminar entre ANRM e HIER em termos de MSEP e correlação de ordem das avaliações genéticas, usando somente dados fenotípicos. Esses resultados podem ser devido às pequenas diferenças entre as

probabilidades a priori e a posteriori nas atribuições de paternidade sob o HIER (Tabela 2).

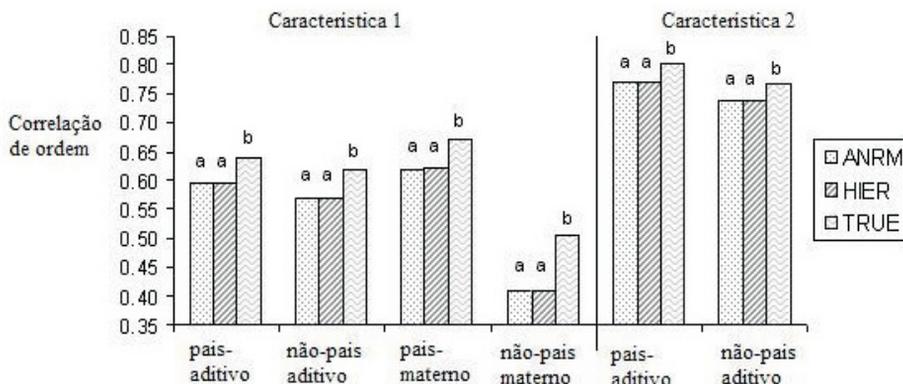


Figura 4. Correlação de ordem de efeitos aditivos e maternos de animais pais e não pais com paternidade incerta para característica 1 e 2 sobre três modelos: (1) HIER: modelo de paternidade incerta; (2) ANRM: matriz de parentesco médio de Henderson; (3) TRUE: modelo de grupos genéticos;

Aplicaram-se dois critérios de escolha de modelo, o PBF e o DIC, descritos anteriormente, para comparar o ajuste estatístico dos dois modelos, ANRM e HIER. O PBF para todas as replicações foi sempre favorável a HIER comparado com ANRM, com magnitude variando de  $2,1 \times 10^2$  a  $2,4 \times 10^7$  para característica 1, e de  $6,3 \times 10^7$  a  $2,6 \times 10^{24}$  para característica 2. O DIC calculado foi também sempre em favor de HIER comparado com o ANRM com diferenças de 9 a 41 para característica 1, e de 33 a 115 para característica 2. Esses resultados são decisivamente em favor do modelo HIER. Spiegelhalter et al. (2002), sugeriram que uma diferença DIC de 7 já pode ser considerada importante no ajuste do modelo aos dados. Para característica 1, a média de DIC sobre as dez replicas dos 17.843 para HIER ( $\bar{D}_{HIER} = 17.135$ ) e  $p_{D(HIER)} = 709$  e 17.866 para ANRM ( $\bar{D}_{ANRM} = 17.164$  e  $p_{D(ANRM)} = 702$ ); e para característica 2, obteve-se média de DIC de 17.533 para HIER ( $\bar{D}_{HIER} = 16.605$  e  $p_{D(HIER)} = 949$ ) e de 17.630 para ANRM ( $\bar{D}_{ANRM} = 16.704$  e  $p_{D(ANRM)} = 926$ ).

A razão primordial para menor DIC do HIER comparado com o ANRM, foi a menor média de deviance ( $\overline{D}_r$ ) do HIER. A diferença em termos de  $\overline{D}_r$  foi suficiente para compensar o maior número efetivo de parâmetros ( $P_{D(r)}$ ) observado para HIER. Os dois critérios (PBF e DIC) concordam em apontar o HIER como modelo de melhor ajuste em relação ao ANRM para simulações de dados envolvendo animais com paternidade incerta.

## 3.2 Inferência nos dados do rebanho Hereford

### 3.2.1 Ganho de peso pós-desmama

O DIC e seus componentes para cada modelo, ANRM e HIER, analisados para o GPD são apresentado na tabela 3. Resultados similares aos obtidos no estudo de simulação foram observados, nos quais  $\overline{D}$  para HIER foi menor que para ANRM e  $P_D$  foi maior para HIER, contudo resultando em DIC menor para o HIER, já que a vantagem alcançada  $\overline{D}$  sobressai-se ao menor  $P_D$  associado com ANRM. Além disso, o PBF para ANRM/ HIER obtido para GPD foi de  $9,695 \times 10^2$ , indicando que o modelo HIER ajusta os dados, aproximadamente, 10 vezes melhor do que ANRM.

**Tabela 3.** Deviance penalização por número efetivo de parâmetros ( $p_D$ ) e critério de informação da deviance (DIC) para os modelos ANRM e HIER usados na análise do ganho de peso pós-desmama (GPD) e do peso à desmama (PD) de um rebanho Hereford brasileiro.

Método	$\overline{D}$	$p_D$	DIC
<b>GPD</b>			
ANRM	28,379	829	29,208
HIER	28,329	860	29,189
<b>PD</b>			
ANRM	44,072	2136	46208
HIER	44,043	2148	46190

Ambos os modelos provêm inferências similares para componentes de variâncias e parâmetros genéticos (Tabela 4). A herdabilidade determinada no presente trabalho, não inclui  $\sigma_{cg}^2$  como parte da variância fenotípica para facilitar a comparação com estimativas da literatura onde os grupos de contemporâneos são considerados como efeito fixo (De Mattos et al., 2000; Koots et al., 1994; Meyer, 1992). A herdabilidade média a posteriori ( $h_a^2$ ) foi de 0,23 e 0,24 para o ANRM e HIER, respectivamente. Essas estimativas, no entanto, foram menores que os valores médios (0,31) baseados em ampla revisão de literatura (Koots et al., 1994), mas dentro da amplitude esperada em dados de produção extensiva e, consideravelmente maior, que as estimativas de herdabilidade para peso ao ano de animais Hereford na Austrália (Meyer, 1992). Para cada componente de variância, verificou-se o intervalo de probabilidade a posteriori (IPP), observando sobreposição entre os dois modelos, sugerindo que não existe diferença prática entre os modelos para os componentes de variâncias e parâmetros genéticos de GPD.

**Tabela 4.** Mediana e desvio padrão (DP) a posteriori, percentis de 2.5% e 97.5% e tamanho efetivo da amostra (TEA) de componentes de variância (em  $kg^2$ ) e parâmetros genéticos para ganho de peso pós-desmama em um rebanho Hereford brasileiro, obtidos pelos modelos de ANRM e HIER.

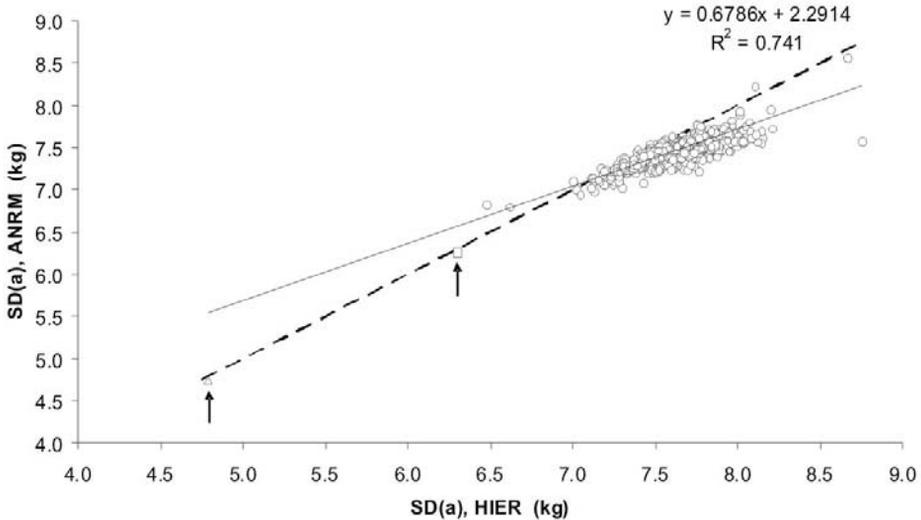
Parâmetro <sup>1</sup>	Mediana	DP	2.5%	97.5%	TEA
<b>ANRM</b>					
$h_a^2$	0,231	0,042	0,153	0,316	254
$\sigma_a^2$	73,8	14,4	48,0	103,6	254
$\sigma_e^2$	246,5	40,9	221,5	271,2	368
$\sigma_{cg}^2$	404,5	12,7	334,3	494,0	34.871
<b>HIER</b>					
$h_a^2$	0,244	0,044	0,162	0,336	337
$\sigma_a^2$	78,2	15,1	51,1	111,2	334
$\sigma_e^2$	242,9	40,8	216,5	268,2	487
$\sigma_{cg}^2$	404,5	13,1	333,9	493,8	35.691

<sup>1</sup>  $h_a^2$  = herdabilidade aditiva;  $\sigma_a^2$  = variância aditiva;  $\sigma_e^2$  = variância residual; e  $\sigma_{cg}^2$  = variância do grupo de contemporâneos.

A correlação de ordem entre a média à posteriori do efeito genético aditivo de todos os animais avaliados obtidos pelos dois diferentes modelos foi maior que 0,99. Entretanto, para animais com paternidade incerta, a correlação de ordem caiu para 0,98 e 0,94, para animais classificados como 10% e 5% superiores, respectivamente. Apesar desses valores mais baixos da correlação de ordem, existiu conformidade geral entre os dois modelos em termos de ordem com as maiores diferenças relacionadas a animais com incerteza de paternidade. Em particular, a diferença em médias a posteriori do efeito genético aditivo entre o ANRM e HIER foi de -1,22 Kg e 2,94 Kg o que não é necessariamente trivial em relação a estimativa do desvio padrão genético  $\sqrt{\sigma_a^2}$ .

Ambos os modelos, HIER e ANRM, tiveram o mesmo desvio padrão à posteriori do valor genético aditivo para a população base e para animais com pais conhecidos (dados não apresentados), entretanto, para indivíduos com incerteza de paternidade, o modelo HIER teve maiores desvios padrão de valores genéticos aditivos, particularmente quando esses desvios foram altos (Figura 5). O desvio padrão a posteriori usando o modelo ANRM foi, em geral, menor que do HIER acima do desvio padrão posteriori de 7,13 kg. Esse limiar é baseado na intersecção de duas linhas apresentadas na Figura 5. A primeira linha é a linha de melhor ajuste entre os dois conjuntos de desvios padrão a posteriori, enquanto a outra linha é uma linha unitária (isto é, com inclinação um e intercepto nulo). A estimativa da inclinação (0,68) para a linha de melhor ajuste indica que além do desPor isso, potencialmente, há um viés que sobre a confiabilidade das estimativas de valores genéticos usando ANRM para animais tendo incerteza de paternidade. Esses resultados são previsíveis já que o ANRM trata atribuições de paternidade como probabilidades conhecidas, enquanto o HIER faz inferência sobre essas probabilidades. É interessante observar na Figura 5 que dois animais tendo menores desvios padrão a posteriori para mérito genético aditivo (como indicado com a seta) têm seus correspondentes desvio padrão a posteriori por ANRM e HIER sobre a linha unitária. Ambos os indivíduos são touros com grande número de progênes conhecidas no banco de dados. O touro com menor desvio padrão (representado pelo triangulo na Figura 5) tem 50 progênes conhecidas enquanto que o outro touro (representado pelo quadrado na Figura 5) tem nove progênes conhecidas com registros de GPD. Conseqüentemente, a confiabilidade da estimativa da sua avaliação genética para GPD foi quase idêntica entre os dois modelos, com pode-se esperar nesta situação.vio padrão a posteriori de 7,13 kg, que inclui quase todos

os animais (Figura 1), o desvio padrão a posteriori para o ANRM baseado na avaliação genética aumenta somente 0,68 kg para cada kg aumentado no desvio padrão a posteriori do HIER. Esses desvios padrões a posteriori podem ser interpretados como sendo análogos aos erros padrão da predição usados para determinar a confiabilidade da estimativa dos valores genéticos baseados em BLUP.



**Figura 5.** Desvio padrão (SD) do efeito genético aditivo do ganho pós-desmama, em kg, de um rebanho Hereford brasileiro com incerteza de paternidade, obtidos pelo ANRM vs. HIER. A linha sólida representa o ajuste de quadrados mínimos obtido pela equação de regressão apresentada no gráfico, enquanto a linha tracejada tem inclinação um e intercepto nulo e  $R^2$  é o coeficiente de determinação. As setas pretas apontam para os touros que tem progênie próprias.

### 3.2.2. Peso à desmama

Como para GPD, observou-se menores  $\overline{D}$  e DIC para o HIER favorecendo este modelo em comparação com o ANRM no ajuste dos dados de peso à desmama (Tabela 3). Além disso, o PBF do ANRM para HIER obtido para PD foi  $1.654 \times 10^{-1}$ , isto é, o HIER provê um ajuste a esses dados aproximadamente seis vezes melhor que o ANRM.

Inferências a posteriori sobre os componentes de variâncias e parâmetros genéticos usando ANRM e HIER foram muito similares (Tabela 5 e 6). Há extensamente sobreposição no IPP a 95% para cada componente de variância e parâmetro genético e as medias a posteriori foram similares entre dois modelos (Tabela 5).

A média a posteriori para  $h_a^2$  foi de 0,12 nos dois modelos (Tabela 6) foi baixa comparada aos resultados encontrados com bovinos Angus (0,26) criados e controlados sobre as mesmas condições (Cardoso, *et al.*, 2001) e comparado com resultados de outros estudos com Hereford (Koots *et al.*, 1994; Meyer, 1992). A média a posteriori para  $h_m^2$  foi 0,18 no ANRM e 0,19 no HIER, sendo similar as estimativas derivados de estudos de populações Hereford da Austrália, Canadá e Uruguay (De Mattos *et al.*, 2000). A média posteriori da  $\sigma_{pe}^2$  foi surpreendentemente alta, totalizando quase 40% da variância fenotípica (excluindo  $\sigma_{cg}^2$  sobre os dois modelos (Tabela 5 e 6). Estudos prévios tem determinado que o ambiente materno totaliza de 15-23% da variância fenotípica para PD em Hereford (Meyer, 1992; de Mattos *et al.*, 2000). Contudo, estimativas relativamente altas de  $\sigma_{pe}^2$  tendem a ser também encontradas em outros rebanhos Hereford (Meyer, 1992, 1993), indicando que, talvez, a produção de leite seja um fator crítico em vacas Hereford.

**Tabela 5.** Mediana e desvio padrão (DP) a posteriori, percentis de 2.5% e 97.5% e tamanho efetivo da amostra (TEA) de componentes de variância (em kg<sup>2</sup>) para peso à desmama em um rebanho Hereford brasileiro, obtidos usando o modelo hierárquico Bayesiano (HIER) e matriz de parentesco médio (ANRM).

Parametro <sup>1</sup>	Mediana	DP	2.5%	97.5%	TEA
<b>ANRM</b>					
$\sigma_a^2$	58,1	16,1	33,8	95,9	153
$\sigma_{am}$	-21,1	15,7	-58,3	4,1	131
$\sigma_m^2$	80,4	23,1	40,7	131,5	130
$\sigma_{pe}^2$	194,5	19,1	157,4	231,6	331
$\sigma_e^2$	206,3	45,4	183,1	225,2	219
$\sigma_{cg}^2$	295,9	10,8	223,2	401,1	34.234
<b>HIER</b>					
$\sigma_a^2$	58,0	19,1	34,1	108,3	167
$\sigma_{am}$	-23,4	16,3	-62,6	1,2	127
$\sigma_m^2$	82,7	23,1	45,5	135,6	171
$\sigma_{pe}^2$	195,0	18,9	157,1	230,7	394
$\sigma_e^2$	205,9	45,5	177,2	224,9	221
$\sigma_{cg}^2$	296,4	12,1	223,9	401,4	48.448

<sup>1</sup>  $\sigma_a^2$  = Variância aditiva;  $\sigma_{am}$  = covariância aditiva-materna;  $\sigma_m^2$  = variância materna;  $\sigma_{pe}^2$  = variância de ambiente permanente materno;  $\sigma_e^2$  = variância

residual;  $\sigma_{cg}^2$  = variância de grupo de contemporâneo.

Como para as análises de GPD, existiu uma concordância geral entre os dois modelos para médias a posteriori dos efeitos genéticos para PD. A correlação de ordem entre média a posteriori de efeitos genéticos aditivos foi maior que 0,99 para todos os animais com paternidade conhecida ou desconhecida. Para animais classificados como 10% e 5% superiores pelas médias a posteriori de efeitos genéticos aditivos pelo HIER, a correlação de ordem entre os dois modelos foi maior que 0,97 para animais com paternidade conhecida. Entretanto, para animais base e animais com incerteza de paternidade, a correlação de ordem diminuiu para 0,90 e 0,94. Para os efeitos genéticos maternos, a correlação correspondente foi sempre de 0,99 ou maior, independente de como os animais foram estratificados. Diferenças em médias a posteriori e efeitos genéticos aditivos para PD entre ANRM e HIER variam de -2,17 a 1,57 kg para animais com incerteza de paternidade.

**Tabela 6.** Mediana e desvio padrão (DP) a posteriori, percentis de 2,5% e 97,5% e tamanho efetivo da amostra (TEA) de parâmetros genéticos para peso à desmama em rebanho Hereford brasileiro, obtidos usando o modelo hierárquico Bayesiano (HIER) e matriz de parentesco média (ANRM).

Parâmetro <sup>1</sup>	Mediana	DP	2,5%	97,5%	TEA
<b>ANRM</b>					
$h_a^2$	0,117	0,035	0,067	0,201	151
$r_{am}$	-0,320	0,050	-0,617	0,076	142
$h_m^2$	0,183	0,178	0,089	0,323	129
$h_t^2$	0,132	0,037	0,083	0,194	248
$p^2$	0,393	0,028	0,317	0,461	290
<b>HIER</b>					
$h_a^2$	0,117	0,041	0,068	0,229	156
$r_{am}$	-0,350	0,050	-0,632	0,021	164
$h_m^2$	0,190	0,170	0,101	0,344	157
$h_t^2$	0,129	0,037	0,079	0,196	363
$p^2$	0,396	0,030	0,319	0,463	337

<sup>1</sup>  $h_a^2$  = herdabilidade aditiva;  $r_{am}$  = correlação aditiva-materna;  $h_m^2$  = herdabilidade materna;  $h_t^2$  = herdabilidade total; e  $p^2$  = fração de ambiente materno permanente

Novamente foi observado que indivíduos com incerteza de paternidade tendem a ter desvios padrão maiores dos valores genéticos sobre o HIER comparado com o ANRM (Figura 6). Como com o GPD, isso também sugere um potencial viés nas confiabilidades reportadas para valores genéticos estimados usando ANRM para PD, apesar de neste caso, dado ao baixo R2 na Figura 6, a relação de correspondência entre desvios padrão obtidos para o mesmo animal por ANRM e HIER não ser tão evidente. Os mesmos dois animais tendo efeitos genéticos aditivos com menor desvio padrão a posteriori para GPD na Figura 5, foram também observados na Figura 6 (como indicado pelas duas setas) tendo menores desvios padrão a posteriori para PD. O respectivo número de progênie conhecidas para dados de PD para dois touros foi 87 (touro representado pelo triângulo na Figura 6) e 17 (touros representados pelo quadrado na Figura 6). Os correspondentes desvios padrão desses touros pelos dois modelos cai próximo à linha unitária na Figura 6 indicando que, como GPD, a incerteza sobre o mérito genético de touros com substancial número de progênie conhecidas é praticamente idêntica entre os dois modelos.

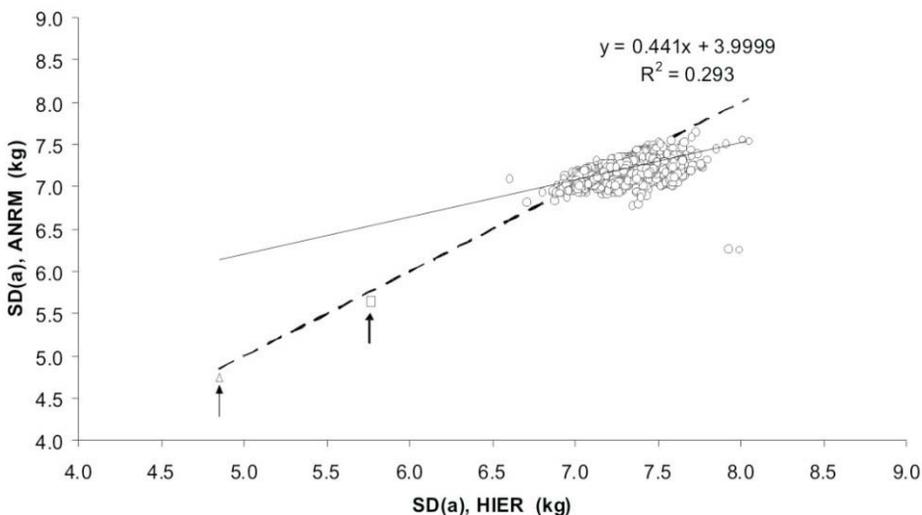


Fig. 6. Desvio padrão (SD) do efeito genético aditivo do peso à desmama, em kg, de um rebanho Hereford brasileiro com incerteza de paternidade, obtidos pelo ANRM vs. HIER. A linha sólida representa o ajuste de quadrados mínimos obtido pela equação de regressão apresentada no gráfico, enquanto a linha tracejada tem inclinação um e intercepto nulo e R2 é o coeficiente de determinação. As setas pretas apontam para os touros que tem progênie próprias.

## 4. Discussão

No presente trabalho, propôs-se um modelo bayesiano para predição do mérito genético dos animais tendo paternidade incerta. O método apresentado pode ser prontamente estendido para características múltiplas ou outros modelos da genética quantitativa (ex: regressão aleatória) sem grande dificuldade conceitual.

Os resultados obtidos por simulação de dados e nas análises do rebanho Hereford (Cardoso; Tempelman, 2004), indicam que o modelo considerando a incerteza nas atribuições de paternidade (HIER) proporciona um melhor ajuste a dados de populações contendo paternidade incerta em relação ao modelo baseado na utilização da matriz de parentesco médio (Henderson, 1988). Poderia-se esperar que a diferença relativa entre os dois modelos aumentasse com a herdabilidade, já que a capacidade de discriminação entre touros candidatos também é aumentada. Foi observado que quando  $h^2 = 0,10$ , não ocorre diferença significativa entre a probabilidade a priori e a posteriori dos touros candidatos (Cardoso; Tempelman, 2001)

Entretanto, com baixa herdabilidade aumenta a importância de informações da progênie com paternidade incerta na predição do mérito genético dos touros. A diferença entre os dois modelos, não necessariamente aumenta com maiores herdabilidades, já que, neste caso, a importância relativa da informação de pedigree e de dados fenotípicos diminuem na predição do mérito genético. O presente trabalho, sugere que a maior diferença de performance entre os dois modelos existe para características com média herdabilidade. Todavia, devido à similaridade em termos de correlação de ordem e especialmente em ausência de informações a priori (por exemplo, marcadores genéticos) o ANRM pode ser escolhido para avaliação genética de grandes populações dado seus menores requerimentos computacionais.

Na presença de informações a priori sobre as atribuições de paternidade, o modelo hierárquico apresentado no presente estudo, representa uma importante alternativa para predições genéticas. Pois, além da incorporação de probabilidades de paternidade a priori em predições do mérito genético, que foi possível tanto com ANRM quanto HIER, o HIER permite

integrar a confiabilidade sobre essas probabilidades a priori na predição do mérito genético.

Uma outra vantagem do HIER comparado com o ANRM é a possibilidade de contabilizar a redução da precisão na inferência do mérito genético devido à incerteza de atribuições de paternidade. Em avaliações genéticas seriam praticamente nulas as diferenças entre os modelos no ordenamento dos valores genéticos. Entretanto as estimativas de acurácia associada com valores genéticos do animal com paternidade incerta seriam sobreestimadas por ANRM. Em estudos de simulação, Kerr et al. (1994) encontraram que o ANRM tendeu a superestimar a acurácia da predição e a resposta da seleção comparado com a matriz de parentesco verdadeiro, o que pode ser conseqüência do incorreto modelo gamético assumido (ex. genes não podem ser transmitidos para uma animal por múltiplos touros).

O HIER representa uma importante alternativa para predição genética sobre a população de bovinos de corte que utilizam acasalamentos com touros múltiplos. Essas populações são geralmente criadas em condições de pastejo usando acasalamentos com reprodutores múltiplos para aumentar a probabilidade de prenhez, com o número de touros dentro do grupo de reprodução com as vacas definido em função do tamanho do piquete, o qual é geralmente muito maior do que pode ser coberto por um único touro. Outro caso de incerteza de paternidade inclui uso de inseminação artificial seguida pelo monta natural ou acasalamentos não planejados. Grandes rebanhos estabelecem um melhor potencial para seleção e progresso genético, mas muito freqüente o exclusivo uso de acasalamento dirigido tem custo muito alto, às vezes impraticável, devido ao tamanho da operação e trabalho requerido. Alguns programas de avaliações genéticas, entretanto, não consideram animais com paternidade incerta. No entanto, resultados simulados mostraram que a diminuição na resposta à seleção devido ao uso de acasalamento de reprodutores múltiplos comparado com o acasalamento dirigido é menor que 10%, particularmente quando as fêmeas tem dados (Kerr; Hammond et al., 1994; Sullivan, 1995) enquanto que a exclusão da performance de dados do animal com paternidade incerta da avaliação pode representar uma redução na resposta a seleção maior que 33% quando comparada ao uso de ANRM de Henderson ou 24% quando comparada com grupos genéticos (Sullivan, 1995), dependendo da porcentagem de animais com paternidade incerta.

## 5. Conclusões

O modelo hierárquico de paternidade incerta proposto neste estudo fornece uma alternativa para avaliação genética de população que utilizam acasalamento com reprodutores múltiplos, permitindo incorporação de informação a priori, proporcionando melhor ajuste aos dados e melhor tratamento da incerteza nas atribuições de paternidade, assim evitando a superestimação da precisão das estimativas dos valores genéticos como ocorre com o uso da matriz de parentesco médio.

Por outro lado, para grandes conjuntos de dados, especialmente na ausência de informações a priori, devido à similaridade em termos de correlação de ordem dos valores genéticos, a matriz de parentesco média pode ser escolhida para avaliação genética pelos seus menores requerimentos computacionais comparada com o modelo hierárquico de paternidade incerta.

## Agradecimentos

Os autores agradecem à Associação Nacional de Criadores e Gensys Consultores Associados e especialmente a Estância Guatambu pela concessão dos dados de seu rebanho.

## Referências Citadas

Cantet, R. J. C., D. Gianola, I. Misztal, et al. Estimates of dispersion parameters and of genetic and environmental trends for weaning weight in Angus cattle using a maternal animal-model with genetic grouping. **Livestock Production Science**, v.34, n.3-4, Apr, p.203-212. 1993.

Cardoso, F. F. **Manual de utilização do Programa INTERGEN – Versão 1.0 em estudos de genética quantitativa animal**. Embrapa Pecuária Sul.

Cardoso, F. F., R. A. Cardellino e L. T. Campos. (Co)Variance components and genetic parameters for weaning production traits of Angus calves raised in the state of Rio Grande do Sul. **Brazilian Journal of Animal Science**, v.30, n.1, p.41-48. 2001.

Cardoso, F. F. e R. J. Tempelman. Bayesian inference on uncertain paternity for prediction of genetic merit. **Journal of Animal Science**, v.79 Suppl. 1, p.111. 2001.

Cardoso, F. F. e R. J. Tempelman. Bayesian inference on genetic merit under uncertain paternity. **Genetics Selection Evolution**, v.35, n.5, Sep-

Cardoso, F. F. e R. J. Tempelman. Genetic evaluation of beef cattle accounting for uncertain paternity. **Livestock Production Science**, v.89,

De Mattos, D., I. Misztal e J. K. Bertrand. Variance and covariance components for weaning weight for Herefords in three countries. **Journal of Animal Science**, v.78, n.1, Jan, p.33-37. 2000.

Foulley, J. L., D. Gianola e D. Planchenault. Sire evaluation with uncertain paternity. **Genetics Selection Evolution**, v.19, n.1, p.83-102. 1987.

Gelfand, A. E. Model determination using sampling-based methods. In: W. R. Gilks, S. Richardson, et al (Ed.). **Markov Chain Monte Carlo in practice**. London: Chapman & Hall, 1996. Model determination using sampling-based methods, p.145-161

Geyer, C. J. Practical Markov Chain Monte Carlo. **Statistical Science**, v.7, n.4, p.473-511. 1992.

Henderson, C. R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. **Biometrics**, v.32, p.69-83. 1976.

Henderson, C. R. Use of an average numerator relationship matrix for multiple-sire joining. **Journal of Animal Science**, v.66, n.7, Jul, p.1614-1621. 1988.

Kerr, R. J., H. U. Graser, B. P. Kinghorn, et al. Implications of using an average relationship matrix in genetic evaluation for a population using multiple-sire matings. **Journal of Animal Breeding and Genetics-Zeitschrift Fur Tierzucht Und Zuchtungsbiologie**, v.111, n.3, p.199-208. 1994.

Kerr, R. J., K. Hammond e B. P. Kinghorn. Effects of multiple-sire matings on genetic evaluations, selection response and rates of inbreeding. **Livestock Production Science**, v.38, n.3, p.161-168. 1994.

Koots, K. R., J. P. Gibson, C. Smith, et al. Analyses of published genetic parameter estimates for beef cattle production traits. 1. Heritability. **Animal Breeding Abstracts**, v.62, n.5, May, p.309-338. 1994.

Meyer, K. Variance-components due to direct and maternal effects for growth traits of Australian beef-cattle. **Livestock Production Science**, v.31, n.3-4, Jun, p.179-204. 1992.

Meyer, K. Covariance matrices for growth traits of Australian Polled Hereford cattle. **Animal Production**, v.57, Aug, p.37-45. 1993.

Perez-Enciso, M. e R. L. Fernando. Genetic evaluation with uncertain parentage - A comparison of methods. **Theoretical and Applied Genetics**, v.84, n.1-2, p.173-179. 1992.

Quaas, R. L. Additive genetic model with groups and relationships. **Journal of Dairy Science**, v.71, n.5, May, p.1338-1345. 1988.

Quaas, R. L. e E. J. Pollak. Mixed model methodology for farm and ranch beef cattle testing programs. **Journal of Animal Science**, v.51, n.6, p.1277-1287. 1980.

Sorensen, D. A., S. Andersen, D. Gianola, et al. Bayesian-inference in threshold models using Gibbs sampling. **Genetics Selection Evolution**, v.27, n.3, p.229-249. 1995.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, et al. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society Series B-Statistical Methodology**, v.64, p.583-616. 2002.

Sullivan, P. G. Alternatives for genetic evaluation with uncertain parentage. **Canadian Journal of Animal Science**, v.75, n.1, p.31-36. 1995.

Wang, C. S., J. J. Rutledge e D. Gianola. Bayesian-analysis of mixed linear-models via Gibbs sampling with an application to litter size in Iberian pigs. **Genetics Selection Evolution**, v.26, n.2, p.91-115. 1994.

**Embrapa**

---

*Pecuária Sul*

**MINISTÉRIO DA AGRICULTURA,  
PECUÁRIA E ABASTECIMENTO**

