

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA”  
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS  
CÂMPUS DE JABOTICABAL**

**AVALIAÇÃO DE MODELOS ESTATÍSTICOS CONSIDERANDO  
A INCERTEZA DA PATERNIDADE**

**Luciana Shiotsuki**

Zootecnista

JABOTICABAL - SÃO PAULO - BRASIL  
FEVEREIRO DE 2011

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA”  
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS  
CÂMPUS DE JABOTICABAL**

**AVALIAÇÃO DE MODELOS ESTATÍSTICOS CONSIDERANDO  
A INCERTEZA DA PATERNIDADE**

**Luciana Shiotsuki**

Orientadora: Profa. Dra. Lucia Galvão de Albuquerque

Co-orientadores: Dr. Fernando Flores Cardoso

Prof. Dr. Josineudson Augusto II de Vasconcelos Silva

Tese de doutorado apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Campus de Jaboticabal, como parte das exigências para obtenção do título de Doutor em Genética e Melhoramento Animal.

JABOTICABAL - SÃO PAULO - BRASIL

FEVEREIRO DE 2011

Shiotsuki, Luciana  
S556a      Avaliação de modelos estatísticos considerando a incerteza da  
paternidade / Luciana Shiotsuki. -- Jaboticabal, 2011  
iv, 85 f. : il. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de  
Ciências Agrárias e Veterinárias, 2011

Orientador: Lucia Galvão de Albuquerque

Banca examinadora: Fernando Flores Cardoso, Roberto Augusto  
de Almeida Torres Junior, Maria Eugenia Zerlotti Mercadante, Danísio  
Prado Munari, Fernando Sebastián Baldi Rey

Bibliografia

1. Bovino de corte. 2. Paternidade incerta. 3. Reprodutor Múltiplo. I.  
Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.2: 636.082

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação –  
Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.

**CERTIFICADO DE APROVAÇÃO**

**TÍTULO:** AVALIAÇÃO DE MODELOS ESTATÍSTICOS CONSIDERANDO A INCERTEZA DA PATERNIDADE

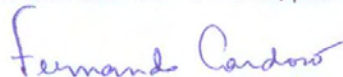
**AUTORA:** LUCIANA SHIOTSUKI

**ORIENTADORA:** Profa. Dra. LUCIA GALVAO DE ALBUQUERQUE

**CO-ORIENTADOR:** Prof. Dr. FERNANDO FLORES CARDOSO

**CO-ORIENTADOR:** Prof. Dr. JOSINEUDSON AUGUSTO II DE VASCONCELOS SILVA

Aprovada como parte das exigências para obtenção do Título de DOUTOR EM GENÉTICA E MELHORAMENTO ANIMAL , pela Comissão Examinadora:



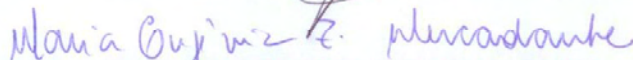
Prof. Dr. FERNANDO FLORES CARDOSO

Centro de Pesquisa de Pecuária dos Campos Sul-Brasileiros / Empresa Brasileira de Pesquisa Agropecuária / Bage/RS



Prof. Dr. ROBERTO AUGUSTO DE ALMEIDA TORRES JUNIOR

Centro Nacional de Pesquisa de Gado de Corte / Embrapa Gado de Corte / Campo Grande/MS




Profa. Dra. MARIA EUGENIA ZERLOTTI MERCADANTE

Estação Experimental de Zootecnia de Sertãozinho / Instituto de Zootecnia



Prof. Dr. DANISIO PRADO MUNARI

Departamento de Ciências Exatas / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal



Prof. Dr. FERNANDO SEBASTIÁN BALDI REY

Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Data da realização: 25 de fevereiro de 2011.

## **DADOS CURRICULARES DO AUTOR**

**LUCIANA SHIOTSUKI**, brasileira, nascido no dia 31 de outubro de 1982, em Ribeirão Preto, SP, filha de Antonio Mitsuo Shiotsuki e Helia Lopes Matias Shiotsuki, iniciou o curso de graduação em Zootecnia na Universidade Federal do Mato Grosso do Sul, no campus de Campo Grande, MS, em agosto de 2001. Em agosto de 2005 ingressou no curso de pós-graduação em Genética e Melhoramento Animal, na Faculdade de Ciências Agrárias e Veterinárias, FCAV, UNESP, campus de Jaboticabal, SP, obtendo o grau de Mestre em 27/02/2007, sob orientação da professora Dra. Lucia Galvão de Albuquerque. Em março de 2007 iniciou o doutorado em Genética e Melhoramento Animal na FCAV, Unesp, campus de Jaboticabal, SP, sob orientação da professora Dra. Lucia Galvão de Albuquerque. No ano de 2009 fez estágio de doutorado "sandwich" na Universidade de Wisconsin, campus de Madison, WI, EUA, sob orientação da professor Dr. Guilherme J. M. Rosa. Em outubro de 2010 fez um curso de especialização em Genética e Melhoramento Animal no "Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria", INIA, em Madrid na Espanha. Obteve o grau de doutora em 25/02/2011.

*"Sorte? eu não sei nada sobre sorte. Eu nunca contei com isto e eu tenho medo das pessoas que o fazem. Sorte para mim é qualquer outra coisa: Trabalho duro, e perceber o que é oportunidade e o que não é." (Lucille Ball)*

## *O SUCESSO É CONSTRUÍDO À NOITE*

*Não conheço ninguém que conseguiu realizar seu sonho, sem sacrificar feriados e domingos pelo menos uma centena de vezes. Da mesma forma, se você quiser construir uma relação amigável com seus filhos, terá que se dedicar a isso, superar o cansaço, arrumar tempo para ficar com eles, deixar de lado o orgulho e o comodismo. Se quiser um casamento gratificante, terá que investir tempo, energia e sentimentos nesse objetivo.*

*O sucesso é construído à noite! Durante o dia você faz o que todos fazem. Mas, para obter resultado diferente da maioria, você tem que ser especial. Se fizer igual a todo mundo, obterá os mesmos resultados. Não se compare à maioria, pois infelizmente ela não é modelo de sucesso.*

*Se você quiser atingir uma meta especial, terá que estudar no horário em que os outros estão tomando chopp com batatas fritas. Terá de planejar, enquanto os outros permanecem à frente da televisão. Terá de trabalhar enquanto os outros tomam sol à beira da piscina, ou passam horas com jogos no micro.*

*A realização de um sonho depende de dedicação. Há muita gente que espera que o sonho se realize por mágica. Mas toda mágica é ilusão. A ilusão não tira ninguém de onde está. Ilusão é combustível de perdedores.*

*"Quem quer fazer alguma coisa, encontra um meio. Quem não quer fazer nada, encontra uma desculpa".*

*Roberto Shinyashiki*

**DEDICO**

*Às pessoas que chamo de minha vida,  
meus pais **HELIA** e **ANTONIO***



## **AGRADECIMENTOS**

À Deus por ser o meu guia.

À Prof. Lucia por ter acreditado, pela amizade e a convivência

Ao Dr. Fernando Flores Cardoso, pela co-orientação, a grande amizade, a família e por todos os conselhos.

Ao Dr. Josineudson Augusto, pela co-orientação desde o início na pós-graduação.

À agropecuária Jacarezinho pelo fornecimento dos dados

À Fundação de Apoio e Pesquisa do estado de São Paulo (FAPESP) pela bolsa concedida.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de doutorado sanduíche concedida.

Aos componentes da banca examinadora de qualificação e defesa Dra. Maria Eugênia Z. Mercadante, Dr. Roberto Augusto de A. Torres Junior, Dr. Danísio P. Munari e Dr. Fernando S. B. Rey que puderam de alguma forma, acrescentar em meu trabalho.

Aos professores da pós-graduação da UNESP/ FCAV.

À todos os funcionários da UNESP/ FCAV, o meu muito obrigada, por tudo que puderam me ajudar para a realização desse trabalho

Aos meus pais, pela amizade, paciência e compreensão pela minha ausência. Pelo apoio incondicional, amizade e amor.

Aos familiares, Ditchan (*in memorian*), Batchan, Tia Irene, Tia To, Tia Lia, Tia América, Tio Carlos, aos primos Carol, Thadeu, Matheus, Rodrigo, Gabriel, Carla, Osvaldinho e a irmã caçula Ana Luiza.

Aos Avós Eurídice e José (*in memoriam*), ao Luiz, a madrinha Dete, aos tios Willian, Maria, Washington, Samanta, Margô, Ricardo e Rogério e aos primos Simone, Bruno e Breno e Luiz Carlos Guedes.

Aos amigos de Campo Grande Ana Maria, Larissa, Fabíola, Luciana Reis, Maurício e Fenelon, e que torceram muito para enfrentar a nova etapa.

Aos amigos irmãos Denise, André (Murotte), Gregório, Natália Sobreira, Paulinha, Diego Guidolin, Airon e Tati.

Aos amigos de Madison, Guilherme Rosa, Thaís, Théo, Fernando Dallas, Fabyano, Fernanda, Germana, Maurício, Jay, Ann Smiley, Leane, Diogo, Miller, Bruno, Ana Angélica, Sandra, Glaucio, Palilo, 'Meninão', Hao, José, Alessandra, Anibal, Walquiria, Elisson, Dave Vitale e todos as pessoas que eu possa ter me esquecido e que me ajudaram em Madison.

Aos amigos do departamento e pós-graduação Durva, Arione, Fabio, Diogo, Tirinha, Rodrigo, Mario, Fuxica, Daniel Gordo, Raul, Naudin, Priscilla, Beto, Fernando, Raphael (Consolo), Rafael, Inaê, Monyka, Marcio Cinachi, Marcus, Marcos (2mi2), Tomás, Liliam.

Aos amigos de Bagé-RS: Michele Mattos, Michelle S. da Rosa, Ubaldina Suzana, Dinda, Leonardo Mattos, Narinha, Diogo, Maurício, Alvim, Lúcio, Marcelo, Carol, Marcelo Blois, Magali, Nicole, Helena.

Aos amigos que fiz na Espanha Carla, Priscila S. Oliveira, Priscila, Natalia, Polyanna, Júlio, Dona Maria, Divier pela boa convivência dos dias do curso.

A todos os amigos de Jaboticabal e aos amigos da pós-graduação da FCAV/UNESP.

À todas as pessoas que, de alguma forma, contribuíram para a realização desse trabalho, o meu **muito obrigada!**

## ÍNDICE

|   |           |
|---|-----------|
| <b>CAPÍTULO 1- CONSIDERAÇÕES GERAIS.....</b>  | <b>01</b> |
| <i>Introdução.....</i>  | <i>01</i> |
| <i>Objetivo geral.....</i>  | <i>05</i> |
| <i>Revisão de literatura.....</i>   | <i>06</i> |
| <b>CAPÍTULO 2- AVALIAÇÃO DO MODELO DE GRUPOS GENÉTICOS E<br/>PATERNIDADE DESCONHECIDA PARA CARACTERÍSTICAS DE CRESCIMENTO<br/>DE ANIMAIS DA RAÇA NELORE.....</b>                                      | <b>17</b> |
| <i>Resumo.....</i>  | <i>17</i> |
| <i>Abstract.....</i>  | <i>18</i> |
| <i>Introdução.....</i>  | <i>19</i> |
| <i>Material e métodos.....</i>  | <i>20</i> |
| <i>Resultados e discussão.....</i>  | <i>26</i> |
| <i>Conclusão.....</i>   | <i>39</i> |
| <i>Referências bibliográficas.....</i>  | <i>40</i> |
| <b>CAPÍTULO 3- COMPARAÇÃO DE PROCEDIMENTOS ESTATÍSTICOS PARA<br/>CONSIDERAR PATERNIDADE DESCONHECIDA PARA AVALIAÇÃO GENÉTICA<br/>DE CARACTERÍSTICAS DE CRESCIMENTO DE ANIMAIS DA RAÇA NELORE.....</b> | <b>44</b> |
| <i>Resumo.....</i>  | <i>44</i> |
| <i>Abstract.....</i>  | <i>45</i> |
| <i>Introdução.....</i>  | <i>46</i> |
| <i>Material e métodos.....</i>  | <i>47</i> |
| <i>Resultados e discussões.....</i>   | <i>55</i> |
| <i>Conclusão.....</i>   | <i>63</i> |
| <i>Referências bibliográficas.....</i>  | <i>64</i> |
| <b>CAPÍTULO 4- METODOLOGIA PARA ESTIMAÇÃO DO MÉRITO GENÉTICO DE<br/>ANIMAIS COM PATERNIDADE INCERTA SOB INFERÊNCIA BAYESIANA.....</b>   | <b>66</b> |
| <i>Resumo.....</i>  | <i>66</i> |
| <i>Abstract.....</i>  | <i>67</i> |
| <i>Introdução.....</i>  | <i>68</i> |
| <i>Material e métodos.....</i>  | <i>69</i> |
| <i>Resultados e discussão.....</i>  | <i>75</i> |
| <i>Conclusão.....</i>   | <i>81</i> |
| <i>Referências bibliográficas.....</i>  | <i>82</i> |
| <b>CAPÍTULO 5- IMPLICAÇÕES.....</b>   | <b>84</b> |

**Resumo:** O sistema de acasalamento mais empregado na pecuária de corte extensiva é a monta a campo com o uso de reprodutores múltiplos. Este sistema traz diversas facilidades ao produtor, entretanto não permite a identificação da paternidade das crias, impossibilitando a comparação do desempenho reprodutivo e produtivo dos diferentes touros, o que afeta negativamente as avaliações genéticas e o progresso genético pela seleção. Objetivou-se com este trabalho comparar metodologias estatísticas que permitam considerar rotineiramente o uso da informação de acasalamentos com touros múltiplos na avaliação genética. Para tanto, foram analisados registros de ganhos de pesos pós-desmama e peso ao sobreano, mensurados em machos e fêmeas de animais da raça Nelore, nascidos no período de 1984 a 2006. Foram aplicados diferentes procedimentos estatísticos que consideram dados de animais com incerteza de paternidade na estimação de parâmetros e valores genéticos. As estimativas dos componentes de (co)variâncias genéticas e dos valores genéticos para os animais presentes no pedigree foram comparadas entre os diferentes procedimentos estatísticos. No capítulo 2, foram utilizados o modelo de paternidade desconhecida e o modelo de grupos genéticos para comparação. Verificou-se que o modelo de grupo genético, definido pelo intervalo de gerações dos machos deve ser usado para classificar e predizer o mérito genético dos filhos de reprodutores múltiplos. No capítulo 3, comparou-se o modelo com base na matriz de parentesco médio e o modelo hierárquico bayesiano (HIER) para incerteza de paternidade. Concluiu-se que o HIER é o modelo que melhor ajustou os dados para estimar os parâmetros genéticos de animais que possuem paternidade incerta. No capítulo 4, propôs-se uma aproximação do modelo hierárquico Bayesiano, usando procedimentos Bayesianos empíricos e máxima verossimilhança. O modelo proposto representou uma alternativa computacionalmente viável para calcular as probabilidades dos touros candidatos e os efeitos genéticos dos animais em avaliações genéticas de grandes bases de dados, quando existe incerteza nas atribuições de parentesco de alguns animais.

**Palavras-chave:** Bovino de corte, ganho de peso pós-desmama, paternidade incerta, peso ao sobreano, reprodutor múltiplo

**Abstract:** Multiple-sire (MS) mating is the most used mating system in extensive beef cattle systems. This system provides several conveniences to the farmers, but does not allow for paternity identification of the offspring, making impossible to compare the reproductive and productive performance of different bulls, negatively affecting the genetic evaluations and therefore the genetic progress. The aim of this work was to compare statistical approaches that allow for considering information from mating with multiple sires in genetic evaluation. Records from postweaning gain and long yearling weight, measured in males and females of Nellore animals, born from 1984 to 2006 were used. We applied different statistical approaches, which consider data from animals with uncertain paternity in the estimation of genetic parameters and values. The estimated values of the (co) variance components and breeding values for the animals in the pedigree were compared from the different statistical procedures. In chapter 2, we compared unknown parentage and genetic groups models. It was found that the genetic group model defined by the generation interval of males was more appropriate to predict the genetic merit of animals with unknown paternity. In chapter 3, we compared the model based on the use of an average numerator relationship matrix and a hierarchical Bayes model (HIER) accounting for uncertain paternity. We concluded that HIER was the best data fitting model to estimate genetic parameters of animals with uncertain paternity. In chapter 4, we proposed a Bayesian hierarchical model approximation, using empirical Bayesian procedures and maximum likelihood. The proposed model represented a computationally feasible alternative to calculate the probabilities of candidate sires and animal genetic effects on genetic evaluations of large datasets, when there are uncertainty paternity assignments for some animals.

**Keywords:** Beef cattle, long yearling weight, multiple-sire, postweaning gain, uncertain paternity

## **CAPÍTULO 1- CONSIDERAÇÕES GERAIS**

A pecuária de corte extensiva é muito expressiva no Brasil, devido a diversos fatores favoráveis, entre eles, os baixos custos de produção. O sistema de acasalamento mais empregado na pecuária de corte extensiva é a monta a campo, já que a inseminação artificial em grande escala se restringe atualmente a poucos rebanhos, por se tratar de uma tecnologia que requer equipe habilitada, acarreta maiores perdas de cio e aumenta a despesa com mão-de-obra e equipamentos (FERREIRA et al., 2006).

Na monta a campo, em sua maioria, é adotado o sistema de reprodutor múltiplo, que se caracteriza pela utilização de mais de um touro em um mesmo lote de matrizes durante a estação de monta. Este sistema apresenta as vantagens de minimizar dificuldades quanto ao manejo individual de animais requerido pela monta controlada, economizar na mão-de-obra e aumentar a chance de que a maioria das fêmeas conceba durante uma determinada estação de monta, já que a perda de identificação das fêmeas em cio pode ser reduzida (VALLE et al., 1998). Embora apresente diversas vantagens práticas para o produtor, o acasalamento com reprodutores múltiplos não permite a identificação da paternidade das crias, impossibilitando a comparação do desempenho reprodutivo e produtivo dos diferentes touros, o que afeta negativamente a acurácia das avaliações genéticas e o diferencial de seleção (CARDOSO & TEMPELMAN, 2003).

Diferentes modelos estatísticos têm sido propostos para predição do mérito genético de animais com paternidade incerta. A estratégia mais antiga proposta é a de grupos genéticos (WESTELL et al., 1988; QUAAS et al., 1988), na qual, grupos de “pais fictícios” são designados aos animais que foram acasalados no mesmo piquete. Embora existam outras alternativas, os grupos genéticos 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 (QUAAS, 1988; CANTET et al., 1993).

Outro modelo utilizado para avaliação genética de animais com paternidade incerta inclui o melhor preditor linear não viesado (BLUP), fundamentado na matriz do numerador de parentesco médio (ANRM) proposta por HENDERSON (1988). Este procedimento se baseia na pressuposição do conhecimento da probabilidade de cada touro ser o pai verdadeiro de um determinado animal. 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.

Uma alternativa proposta por FOULLEY et al. (1987) é o procedimento bayesiano empírico, que utiliza, além das probabilidades de paternidade a priori, informações contidas nos dados para calcular probabilidades a posteriori de cada touro, no grupo de reprodutores múltiplos, ser o pai verdadeiro de animais com paternidade incerta para calcular o mérito genético desses animais. Esse procedimento, entretanto, apresenta a limitação de restringir-se ao uso de modelo touro.

Com o desenvolvimento da Cadeia de Markov de Monte Carlo (MCMC) aplicada ao melhoramento animal (WANG et al., 1994), foi possível estender o método de FOULLEY et al. (1987) para o modelo animal, sob enfoque bayesiano (CARDOSO & TEMPELMAN, 2003). Este modelo hierárquico bayesiano permite uma avaliação estatística mais formal da imprecisão do mérito genético e das probabilidades de atribuição de paternidade. Entretanto, a adoção dessa metodologia em processos rotineiros de avaliação genética, pode ser dificultada pela demanda computacional associada aos métodos MCMC. Assim, o desenvolvimento de um procedimento Bayesiano empírico para implementar o modelo de CARDOSO & TEMPELMAN (2003) via máxima verossimilhança, seguindo os conceitos propostos por FOULLEY et al. (1987), facilitará o cálculo de probabilidade de paternidade a posteriori e sua utilização em avaliações genéticas, com potencial redução no tempo de computação.

Atualmente, não se utilizam modelos que consideram a incerteza da paternidade para predição do mérito genético dos animais em programas de melhoramento animal. No entanto, animais sem identificação dos pais são avaliados, e têm seus valores com base nos pais fundadores, ou seja, na população base. Isso faz

com que as predições não sejam tão acuradas, interferindo nos resultados das avaliações genéticas e conseqüentemente no progresso genético.

A utilização de modelos estatísticos que considerem a incerteza de paternidade na estimação dos valores genéticos dos animais é de grande relevância para a pecuária nacional, dado que o sistema de acasalamento com uso de reprodutor múltiplo no Brasil é amplamente utilizado em propriedades de criação extensiva, a qual representa 96,84% (ANUALPEC, 2010) e a inferência sobre paternidade levará a uma maior segurança nas avaliações genéticas e a possibilidade da inclusão de rebanhos comerciais em programas de melhoramento genético.

## **OBJETIVOS**

O objetivo geral do presente estudo foi comparar metodologias estatísticas que permitam considerar rotineiramente o uso da informação de acasalamentos com touros múltiplos na avaliação genética para o ganho de peso pós-desmama e peso ao sobreano de animais da raça Nelore.



## REVISÃO DE LITERATURA

Em avaliações genéticas, a ausência da correta identificação da paternidade ocasiona diminuição da acurácia e da intensidade de seleção dos animais, fazendo-se necessária a utilização de modelos apropriados que minimizem o erro de predição e que permitam predizer corretamente ou com maior segurança o mérito genético de cada animal.

Vários estudos foram realizados com a utilização de diferentes procedimentos estatísticos com o objetivo de minimizar o erro de predição, quando os arquivos de dados continham observações de animais com paternidade incerta. Na maioria dos casos esses animais com paternidade incerta eram provenientes de acasalamento com uso de reprodutor múltiplo. Entre as metodologias verificadas na literatura, foi proposto a utilização de grupos genéticos (QUAAS, 1988; WESTELL et al., 1988), modelo com base na matriz de parentesco médio (HENDERSON, 1988), modelo animal hierárquico Bayesiano, que utiliza informações de desempenho para calcular probabilidades a posteriori dos pais candidatos (CARDOSO & TEMPELMAN, 2003) e modelo que não requer a inversa da matriz de parentesco médio, desta forma, demanda menor tempo computacional proposto por SAPP (2005).

Modelos estatísticos que incluem grupos genéticos, segundo KENNEDY (1981), apresenta benefícios como diminuição do erro de predição, especialmente quando a seleção ocorre com o uso de dados de desempenho ou arquivo de pedigree incompleto. Esta metodologia considera, para animais que não possuem paternidade conhecida, o efeito de um grupo, que é expresso como a média ponderada de um pequeno número de efeitos, como proposto por THOMPSON (1979) para modelo touro e WESTELL (1984) e ROBINSON (1986) para modelo animal. As equações de modelo misto (MME) para o modelo animal com efeitos de grupos genéticos definidos desta forma, foram simplificadas por WESTELL et al. (1988), usando técnicas de transformação descritas por QUAAS & POLLAK (1981) para modelo touro com grupos genéticos. Realizada essa simplificação, fica fácil adicionar os efeitos de grupos genéticos, utilizando definições de grupos muito complexos. QUAAS (1988) advertiu

para dificuldades quanto ao confundimento entre grupos e outros efeitos fixos no modelo, especialmente quando se trata de grupos com poucos animais.

WESTEL et al. (1988) descreveram as equações de modelo misto incluindo grupos genéticos, da seguinte maneira:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu}^* + \mathbf{ZQg} + \mathbf{e};$$

$$\text{Assumindo que: } \text{var} \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & 0 \\ 0 & \mathbf{I}\sigma_e^2 \end{bmatrix};$$

em que  $\mathbf{y}$  é o vetor de desempenho,  $\mathbf{b}$  é o vetor de efeitos não-genéticos,  $\mathbf{X}$  e  $\mathbf{Z}$  são matrizes de incidência que relacionam as observações aos efeitos não-genéticos e efeito genético aditivo, respectivamente;  $\mathbf{u} = \mathbf{u}^* + \mathbf{Qg}$  (vetor dos efeitos aleatórios genéticos aditivos diretos mais os efeitos aditivos de grupo);  $\mathbf{Q}$  = matriz de incidência dos efeitos de grupo;  $\mathbf{g}$  = vetor de efeitos fixos genéticos aditivos de grupo;  $\mathbf{e}$  é o vetor de efeito residual;  $\mathbf{A}$  é a matriz de parentesco médio;  $\mathbf{I}$  é a matriz identidade;  $\sigma_a^2$  é a variância genética aditiva;  $\sigma_e^2$  é a variância residual;

No entanto, PEREZ-ENCISO & FERNANDO (1992) e SULLIVAN (1995) descreveram a ineficiência da utilização de grupos genéticos na avaliação genética de animais com paternidade incerta, em comparação com outros métodos. Os mesmos autores afirmam que o modelo com grupos genéticos quando se tem acasalamentos com reprodutor múltiplo, assume um número infinito de animais não endogâmicos, isto é, os pais candidatos são considerados não aparentados. Além disto, assume-se que dentro de cada grupo, cada candidato tenha a mesma probabilidade de ser o pai correto do animal com incerteza de paternidade.

Em outro método proposto por HENDERSON (1988), a matriz de parentesco médio (ANRM) auxilia a correta especificação da matriz de (co)variâncias genéticas. Esse método atribui probabilidades para cada um dos possíveis pais dos animais com paternidade incerta, ou seja, probabilidades são presumidamente conhecidas, e desse modo facilita a melhor predição linear não-viesada (BLUP) do mérito genético. Essas

probabilidades são utilizadas para obtenção do coeficiente de parentesco aditivo do animal com paternidade incerta.

O cálculo segue a seguinte estrutura:

$$\mathbf{a}_{ij} = \frac{1}{2} \left[ \mathbf{a}_{is_1}^{(j)} \mathbf{p}_1^{(j)} + \dots + \mathbf{a}_{is_g}^{(j)} \mathbf{p}_g^{(j)} + \mathbf{a}_{id}^{(j)} \right] = \frac{1}{2} \left[ \sum_{k=1}^g \mathbf{a}_{is_k}^{(j)} \mathbf{p}_k^{(j)} + \mathbf{a}_{id}^{(j)} \right]$$

em que  $\mathbf{a}_{ij}$  é o parentesco aditivo entre o animal  $i$  e  $j$ ;  $\mathbf{a}_{is_k}^{(j)}$  é o parentesco aditivo entre o animal  $i$  e cada um dos pais possíveis ( $k = 1, \dots, g$ ) do animal  $j$ ;  $\mathbf{p}_k^{(j)}$  é a probabilidade computada do touro  $k$  ser o pai do animal  $j$ ;  $\mathbf{a}_{id}^{(j)}$  é o parentesco aditivo entre o animal  $i$  e a mãe do animal  $j$ .

Realizado este procedimento para todos os animais com paternidade incerta, a matriz de parentesco, agora chamada de matriz de parentesco médio, segue o procedimento de solução das equações de modelos mistos (PATTERSON & THOMPSON, 1971).

Diversos trabalhos (PEREZ-ENCISO & FERNANDO, 1992; SULIVAN, 1995) avaliaram o uso do modelo animal utilizando a matriz de parentesco médio (ANRM) proposta por HENDERSON (1988). Neste procedimento, os animais com paternidade incerta, têm seus valores genéticos estimados com base na média dos possíveis pais fundadores. Essas estimativas dos valores genéticos dos animais têm propriedades “BLUP – best linear unbiased prediction” e são utilizadas para seleção com resultados satisfatórios.

Um algoritmo simples e rápido para computar a inversa da ANRM está disponível (FAMULA, 1992) e foi demonstrado, em estudos de simulação (PEREZ-ENCISO & FERNANDO, 1992; SULIVAN, 1995), que é possível obter melhores respostas à seleção utilizando ANRM em comparação aos grupos genéticos, quando informações sobre os possíveis pais foram registradas. Na ausência de informações, como tipagem sanguínea, marcadores genéticos, fertilidade, período reprodutivo e gestação, probabilidades a priori iguais podem ser assumidas para os touros pertencentes ao mesmo grupo de reprodutores múltiplos (FOULLEY et al., 1987; HENDERSON, 1988).

CARDOSO & TEMPELMAN (2003) propuseram o modelo animal hierárquico Bayesiano (HIER), que considera a incerteza nas atribuições de paternidade de animais participantes do sistema de reprodutor múltiplo (RM), em que probabilidades a priori fixas são atribuídas a cada possível pai dos animais com paternidade incerta, sendo que as probabilidades a posteriori são alteradas de acordo com a informação contida nos dados.

Esse modelo pode ser descrito em quatro estágios:

*1º Estágio:* O desempenho dos animais é descrito como uma função linear de efeitos genéticos aditivos e não-genéticos.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_3\mathbf{gc} + \mathbf{Z}_4\mathbf{pe} + \mathbf{e};$$

resultando na seguinte distribuição condicional de amostragem dos dados:

$$\mathbf{y} | \mathbf{a}, \mathbf{b}, \mathbf{m}, \mathbf{gc}, \sigma_e^2 \sim \mathbf{N}(\mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_3\mathbf{gc} + \mathbf{Z}_4\mathbf{pe}, \mathbf{I}\sigma_e^2);$$

em que  $\mathbf{y}$  é o vetor de desempenho,  $\mathbf{b}$  é o vetor de efeitos não-genéticos,  $\mathbf{a}$  é o vetor de efeitos genéticos aditivos;  $\mathbf{m}$  é o vetor de efeitos genéticos maternos e  $\mathbf{pe}$  é o vetor de efeito de ambiente permanente materno, ambos foram considerados somente para o PS;  $\mathbf{gc}$  é o vetor de efeitos de grupos de contemporâneos;  $\mathbf{e}$  é o vetor de efeito residual, com  $\mathbf{e} | \sigma_e^2 \sim \mathbf{N}(0, \mathbf{I}\sigma_e^2)$ ;  $\mathbf{X}$ ,  $\mathbf{Z}_1$ ,  $\mathbf{Z}_2$ ,  $\mathbf{Z}_3$  e  $\mathbf{Z}_4$  são matrizes de incidência que relacionam as observações aos efeitos não-genéticos, efeito genético aditivo, materno, de grupo contemporâneo e ambiente permanente materno, respectivamente;  $\sigma_e^2$  é a variância residual,  $\mathbf{I}$  é a matriz identidade;

*2º Estágio:* A distribuição *a priori* para efeitos genéticos aditivos e não-genéticos são descritos por:

$$\text{Efeito não-genético } (\mathbf{b}): \mathbf{b} \sim \mathbf{N}(\mathbf{b}_0, \mathbf{V}_b);$$

Efeito genético aditivo (**a**):  $\mathbf{a} | \mathbf{s}, \sigma_a^2 \sim \mathbf{N}(0, \mathbf{A}\sigma_a^2)$ ;

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{m} \end{bmatrix} | \mathbf{S}, \mathbf{G}_0 \sim \mathbf{N}(0, \mathbf{G});$$

$$\mathbf{G} = \mathbf{A} \otimes \mathbf{G}_0$$

$$\mathbf{G}_0 = \mathbf{N}(0, \mathbf{G});$$

$$\mathbf{G}_0 = \begin{bmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{bmatrix};$$

Efeito de grupos de contemporâneos:  $\mathbf{gc} \sim \mathbf{N}(0, I\sigma_{gc}^2)$ ;

Efeito de ambiente permanente materno:  $\mathbf{pe} \sim \mathbf{N}(0, I\sigma_{pe}^2)$ ;

Para a predição do valor genético do animal, (co)variâncias do efeito genético aditivo baseiam-se na matriz de parentesco ( $\mathbf{A}$ ), touros candidatos ( $s$ ), variância genética ( $\sigma_a^2$ );  $\sigma_{gc}^2$  é a variância dos efeitos de grupos de contemporâneos;  $\sigma_{pe}^2$  é a variância efeito de ambiente permanente materno;  $\mathbf{V}_b$  é a uma matriz diagonal da variância *a priori* de  $\mathbf{b}$ , assumindo-se  $\mathbf{V}_b \propto \infty$ .

A distribuição qui-quadrado escalonada invertida é assumida *a priori* para a variância residual ( $\sigma_e^2$ ), ou seja:

$$\sigma_e^2 \sim \mathbf{S}_e^2 \chi_{\nu}^{-2};$$

em que  $\mathbf{S}_e^2$  é o valor *a priori* da variância residual baseado em informações da literatura e  $\nu$  é o grau de confiança nesse valor (SORENSEN & GIANOLA, 2002).

*3º Estágio:* Para a distribuição *a priori* da variância genética, assume-se que:

A densidade *a priori* Wishart invertida:  $\mathbf{G}_0 \sim \mathbf{IW}_{\nu(a)}(\mathbf{S}_a)$

$$\sigma_{pe}^2 \sim \mathbf{S}_{pe}^2 \chi_{\nu(pe)}^2;$$

$$\sigma_{gc}^2 \sim \mathbf{S}_{gc}^2 \chi_{\nu(gc)}^2$$

em que  $\mathbf{S}_a^2, \mathbf{S}_{pe}^2, \mathbf{S}_{gc}^2$  são os valores *a priori* da variância genética aditiva, variância de efeito de ambiente permanente materno e variância dos efeitos de grupos de contemporâneos, respectivamente, baseados na literatura e  $v_{(a)}, v_{(pe)}, v_{(gc)}$  são os graus de confiança desses valores.

Além disso, neste estágio são especificadas a probabilidade que  $s_j$  seja o pai verdadeiro do animal  $j$  dado por:

$$\pi_j^{(k)} = \text{Prob}(\mathbf{s}_j = \mathbf{s}_j^{(k)}) \text{ para os pais candidatos } \mathbf{k} = 1, 2, \dots, \mathbf{v}_j;$$

sendo o conjunto de probabilidades *a priori* de cada touro  $\mathbf{v}_j$  ser o pai verdadeiro do animal  $j$ :

$$\boldsymbol{\pi}_j = \{\pi_j^{(1)}, \pi_j^{(2)}, \dots, \pi_j^{(v_j)}\} \text{ para os } \mathbf{v}_j \text{ pais candidatos do animal } j.$$

4<sup>o</sup> Estágio: A distribuição *a posteriori* para probabilidades de paternidade é descrita nesse estágio. O conjunto inteiro de probabilidades para todos os animais que não pertencem à base, raramente é conhecido com absoluta certeza, podendo ser considerado como quantidades aleatórias da distribuição de Dirichlet, no quarto estágio do modelo:

$$p(\boldsymbol{\pi}|\boldsymbol{\alpha}) = \prod_{j=q_b+1}^q \prod_{k=1}^{v_j} (\pi_j^{(k)})^{\alpha_j^{(k)}}.$$

em que  $\alpha_j = \{\alpha_j^{(k)}\}_{k=1}^{v_j}$ ,  $\alpha_j^{(k)} > 0$  para  $\mathbf{k} = 1, 2, \dots, \mathbf{v}_j$  e  $\pi_j^{(v_j)} = 1 - \sum_{k=1}^{v_j-1} \pi_j^{(k)}$ .

Especificações do conjunto de hiper-parâmetros  $\boldsymbol{\alpha} = \{\alpha_j\}_{j=q_b+1}^q$  podem ser baseadas na avaliação da confiabilidade de uma fonte de informação externa sobre as probabilidades *a priori* de cada touro designado.

SAPP et al. (2007), trabalhando com dados simulados, estudaram metodologias para aumentar a acurácia de predição da paternidade em casos que a paternidade incerta existia para alguns animais, em análises uni e multi-característica. Esses autores concluíram que calcular probabilidades *a posteriori* de pais candidatos utilizando informações de desempenho, permite uma estimação do mérito genético de

forma mais acurada para todos os animais, principalmente para animais com paternidade desconhecida. O método proposto por SAPP (2005) não requer a construção da inversa da matriz de parentesco e, conseqüentemente, demanda menor tempo computacional, entretanto, as estimativas não foram significativamente diferentes comparado ao modelo hierárquico bayesiano (SAPP et al., 2007).

CARDOSO & TEMPELMAN (2003), trabalhando com estudo de simulação, compararam duas modelagens estatísticas, modelo baseado na matriz de parentesco médio (ANRM) e modelo hierárquico com incerteza de paternidade (HIER). Foram consideradas duas características, a primeira atribuiu-se herdabilidade direta média ( $h^2=0,30$ ), herdabilidade materna média ( $h^2=0,20$ ) e correlação entre efeito genético direto e materno baixa e negativa ( $r_{am} = -0,20$ ), semelhante ao que ocorre com peso à desmama. Para a segunda característica, foi atribuído herdabilidade direta alta ( $h^2=0,50$ ) e nenhuma herdabilidade materna, como é o caso do ganho de peso pós-desmama. Esses autores observaram probabilidades *a posteriori* atribuídas ao pai verdadeiro entre 1 a 13% maiores que os valores *a priori* (inverso do número de pais candidatos), com diferenças dependendo da herdabilidade e do tamanho do grupo de reprodutores múltiplos. Além disso, o quadrado médio do erro de predição foi sempre menor e as correlações de ordem foram maiores, comparando HIER à ANRM, em relação ao modelo com uso do pai verdadeiro. Os autores verificaram que a utilização de acasalamentos com reprodutores múltiplos apresenta efeitos negativos na acurácia das avaliações genéticas, mesmo que modelada apropriadamente. Dois critérios de escolha do modelo foram utilizados, ambos apontaram o modelo hierárquico bayesiano melhor para prever o mérito genético dos animais comparado ao modelo baseado na matriz de parentesco médio.

CARDOSO & TEMPELMAN (2004), trabalhando com animais da raça Hereford, observaram que as diferenças relativas entre o HIER e o ANRM variam dependendo do valor da herdabilidade da característica. Entretanto, a maior diferença de desempenho entre os dois modelos ocorreu quando a característica foi de média herdabilidade. Devido à similaridade em termos de correlação de ordem e especialmente pela ausência de informações *a priori*, por exemplo, marcadores genéticos, o modelo com base na matriz de parentesco médio pode ser escolhido para

avaliação genética de grandes populações, dado a menor demanda computacional. Por outro lado, na presença de informações *a priori* sobre as atribuições de paternidade, o modelo hierárquico animal, representa uma importante alternativa para predições genéticas, pois, além da incorporação de probabilidades de paternidade *a priori*, que também foi possível no modelo baseado na matriz de parentesco médio, permite integrar a confiabilidade sobre essas probabilidades *a priori* na predição do mérito genético.

Embora algumas limitações ainda existam, informações fenotípicas podem ser utilizadas para aumentar a probabilidade de determinar a paternidade de animais em acasalamentos múltiplos e podem ser aplicadas a uma gama extensiva de características e implementadas em diferentes cenários dependendo dos dados disponíveis. SAPP et al. (2007) incluíram informações fenotípicas em análises uni e tricaracterística para diferentes estimativas de herdabilidade. Os autores concluíram que, em análises unicaracterística, a probabilidade média do touro verdadeiro ser identificado aumentou quando a herdabilidade foi maior. Já para análises multicaracterísticas, em que informações fenotípicas de três características foram incluídas, maiores probabilidades de identificar o touro verdadeiro foram observadas, assim como, se verificaram que a acurácia dos valores genéticos preditos foram maiores que nas análises unicaracterística. Os mesmos autores estimaram correlação entre os valores genéticos foi maior quando as diferentes estimativas de probabilidades da paternidade para os touros candidatos calculadas pelo modelo foram usadas.

Grandes rebanhos apresentam um melhor potencial para seleção e progresso genético, mas o acasalamento dirigido muito frequente ou exclusivo tem custo muito alto, às vezes impraticável, devido ao tamanho da operação e mão-de-obra requerida. O mesmo acontece com identificação de paternidade por métodos moleculares, embora seja o método mais acurado para identificar a paternidade em grupos de reprodutores múltiplos, é de alto custo e atualmente impraticável na pecuária comercial. Alguns programas de avaliações genéticas excluem animais com paternidade incerta ou assumem touros desconhecidos para esses animais. Resultados de trabalhos de simulação mostraram perda na resposta à seleção quando os dados de desempenho de animais com paternidade incerta foram excluídos



(SULLIVAN, 1995). O mesmo autor concluiu que a perda na resposta a seleção depende da porcentagem de animais com paternidade incerta.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANUALPEC. Anuário da pecuária Brasileira, AgraFNP Consultoria e Informação em Agronegócio, editora Argos, 62 p. 2010.

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. **Liv. Prod. Sci.**, v.34, n.3-4, Apr, p.203-212. 1993.

CARDOSO, F. F.; R. J. TEMPELMAN. Bayesian inference on genetic merit under uncertain paternity. **Gen. Sel. Evol.** v.35, p. 469-487. 2003.

CARDOSO, F. F.; R. J. TEMPELMAN. Genetic evaluation of beef cattle accounting for uncertain paternity. **Liv. Prod. Sci.**, v. 89, p. 109-119. 2004.

FAMULA, T.R. Simple and rapid inversion of additive relationship matrices incorporating parental uncertainty. **J. Anim. Sci.**, v. 70, p. 1045-1048. 1992.

FERREIRA, A. M.; SÁ, W.F. *et al.* **Inseminação Artificial**. <[http://www.agencia.cnptia.embrapa.br/Agencia8/AG01/arvore/AG01\\_25\\_217200392357.html](http://www.agencia.cnptia.embrapa.br/Agencia8/AG01/arvore/AG01_25_217200392357.html)>. Acesso em 25 jul. 2006.

FOULLEY, J.L.; GIANOLA, D.; PLANCHENAULT, D. Sire evaluation with uncertain paternity. **Gen. Sel. Evol.** 19: 83-102. 1987.

HENDERSON C.R. Use of an average numerator relationship matrix for multiple-sire joining, **J. Anim. Sci.**, v. 66, p. 1614-1621. 1988.

KENNEDY, B.W. Bias and mean square error from ignoring genetic groups in mixed model sire evaluation. **J. Dairy Sci.** v. 64, p.689-697. 1981.

KERR, R. J., K. HAMMOND E B. P. KINGHORN. Effects of multiple-sire matings on genetic evaluations, selection response and rates of inbreeding. **Liv. Prod. Sci**, v.38, n.3, p.161-168. 1994.

PATTERSON, H.D. THOMPSON, R. Recovery of interblock information when block sizes are unequal. **Biometrika**. v. 58, p.545-554. 1971.

PEREZ-ENCISO M.; FERNANDO R.L. Genetic evaluation with uncertain parentage- a comparison of methods, **Theor. Appl. Gen.** v. 84 p. 173-179. 1992.

QUAAS, R.L.; POLLAK, E.J. Modified equations for sire models with groups. **J.Dairy Sci.**, v. 64, p. 1868-1872. 1981.

QUAAS, R. L. Additive genetic model with groups and relationships. **J. Dairy Sci.**, v. 71, p. 1338-1345. 1988.

ROBINSON, G.K. Group effects and computing strategies for models for estimating breeding values. **J. Dairy Sci.**, v. 69, p.3106-3111. 1986.

SAPP, R. L. 2005. Statistical approach for dealing with uncertain paternity, **PhD Diss.** Univ. Georgia, Athens.

SAPP, R.L.; ZHANG, W.; BERTRAND, J.K. et al. Genetic evaluation in the presence of uncertain additive relationships. I. Use of phenotypic information to ascertain paternity. **J. Anim. Sci.**, v. 85, p. 2391-2400. 2007.

SORENSEN, D., GIANOLA, D. 2002. **Likelihood, bayesian, and MCMC methods in quantitative genetics.** ed.Springer-Verlag, New York.

SULLIVAN P.G. Alternatives for genetic evaluation with uncertain parentage, **Can. J. Anim. Sci.** v. 75, p. 31-36. 1995.

THOMPSON, R. Sire evaluation. **Biometrics.** v. 35, p.339-353. 1979.

VALLE, E.R.; ANDREOTTI, R.; THIAGO, L.R.S. **Estratégias para aumento da eficiência reprodutiva e produtiva em bovinos de corte.** Campo Grande- MS, EMBRAPA, Doc 71, 1998.

WANG, C. S., J. J. RUTLEDGE, AND D. GIANOLA. Bayesian-analysis of mixed linear-models via Gibbs sampling with an application to litter size in Iberian pigs. **Gen. Sel. Evol.** v.26, p. 91-115. 1994.

WESTELL, R.A. **Simultaneous evaluation of sires and cows for a large population.** Ph.D. Thesis, Cornell Univ., Ithaca, NY. 1984.

WESTELL R.A., QUAAS R.L., VAN VLECK L.D. Genetic groups in an animal-model, **J. Dairy Sci.**, v.71, p. 1310-1318. 1988.

## **CAPÍTULO 2 – AVALIAÇÃO DO MODELO DE GRUPOS GENÉTICOS E PATERNIDADE DESCONHECIDA PARA CARACTERÍSTICAS DE CRESCIMENTO DE ANIMAIS DA RAÇA NELORE**

**Resumo:** O objetivo do presente trabalho foi comparar os modelos de paternidade desconhecida e de grupos genéticos a fim de indicar o procedimento estatístico mais adequado para melhor estimar o valor genético de animais que possuem incerteza de paternidade. Após a consistência dos dados, foram mantidos 62.212 animais da raça Nelore com registros de ganho de peso pós-desmama e o peso ao sobreano. O arquivo de pedigree continha 75.088 animais, sendo 20.827 (37,10%) animais filhos de RM e 12.876 animais pertencentes à população base do arquivo com pais desconhecidos. Três abordagens foram adotadas no tratamento de paternidade desconhecida para os animais filhos de reprodutores múltiplos (RM). No modelo de paternidade desconhecida, considerou-se que os pais eram desconhecidos e pertencentes a uma única população base. Na abordagem de grupos genéticos, foram utilizadas duas definições de grupos genéticos para comparação com o modelo de paternidade desconhecida. Na primeira (GGa), atribuíram-se “pais fictícios” aos animais com incerteza de paternidade, designando que cada grupo genético foi formado pelo grupo de RM. Na outra definição (GGb), também atribuíram-se “pais fictícios”, entretanto, os animais provenientes de acasalamentos com RM foram agrupados de acordo com seus anos de nascimento, a cada três anos, baseado no intervalo de gerações dos machos. Foram realizadas análises unicaracterística, sob enfoque bayesiano, utilizando métodos Monte Carlo via cadeias de Markov (MCMC). Para verificar o modelo de melhor ajuste foram utilizados o critério de informação da deviance e a ordenada preditiva condicional (CPO). De acordo com os resultados dos critérios de escolha, o modelo de grupo genético, definido pelo intervalo de gerações dos machos (GGb) foi mais adequado para predizer o mérito genético dos animais com incerteza de paternidade. Recomenda-se o uso desse modelo para predizer o mérito genético e classificar filhos de reprodutores múltiplos.

**Palavras-chave:** Incerteza de paternidade, inferência bayesiana, modelo de grupo genético

**Abstract:** The objective of this work was to compare unknown paternity model and genetic group model to indicate the most appropriate statistical procedure to better estimate the genetic value of animals that have uncertain paternity, using growth traits. After editing, the data set contained records of 62,212 Nelore animals with postweaning gains and long yearling weight. The pedigree file contained records of 75,088 Nelore animals, 20,827 (37.10%) offspring of multiple-sire (MS) and 12,876 animals belonging to the population base file with uncertain paternity. Three different approaches were adopted to deal with uncertain paternity of multiple-sire (MS) offspring. In the unknown paternity model, sire of MS were considered unknown and pertaining to a single base population. In the genetic group approach, two definitions were used. In the first definition (GGa), "phantom parents" for animals with uncertain paternity were attributed, defining the genetic group as the group of multiple-sire. In the other approach, "phantom parents" for animals with uncertain paternity were also attributed, however, MS offspring were clustered in genetic groups according to their year of birth, every three years, based on the average of males generation interval. Univariate analyses were carried out, under Bayesian approach via Markov Chain Monte Carlo (MCMC) methods. Models were compared by deviance information criteria (DIC) and the conditional predictive ordinate (CPO). According to the choice criteria results, the genetic group model defined by the generation interval of males was more appropriate to predict the genetic merit of animals with uncertain paternity. Therefore, this model should be used to predict and rank the genetic merit of MS offspring.

**Keyword:** Bayesian inference, genetic group model, uncertain paternity

## INTRODUÇÃO

O sistema de acasalamentos adotado por grande parte dos produtores de bovinos de corte é o de reprodutores múltiplos (RM) devido às suas facilidades quanto à mão de obra e ao baixo custo, comparado aos outros sistemas. Embora seja comumente utilizado, o sistema de expor um grupo de vacas a vários touros apresenta grande desvantagem sob o ponto de vista de avaliações genéticas, já que a paternidade das crias é considerada desconhecida para os programas de melhoramento animal. Desta forma, a comparação do desempenho reprodutivo e produtivo dos diferentes touros é impossibilitada, afetando negativamente a acurácia da seleção.

A metodologia de grupos genéticos, proposta por WESTELL et al. (1988), foi desenvolvida para avaliações genéticas em que os pais desconhecidos são supostamente derivados de diferentes populações. Nesse modelo, por meio de atribuições de “pais fictícios”, permitem-se médias de valores genéticos distintas para as diferentes populações base ou grupos genéticos. No caso de reprodutores múltiplos, os grupos de pais fictícios são tipicamente definidos como o próprio grupo de touros que foi acasalado com um determinado grupo de vacas. Esta metodologia minimiza o erro de predição (QUAAS, 1988; CANTET et al., 1993). Porém, as estimativas dos efeitos dos grupos genéticos, definidas como o valor genético médio do grupo de RM, podem estar confundidas com outros efeitos do modelo, por exemplo, grupos de contemporâneos (GC), especialmente se todos os filhos de um determinado grupo de RM compõem única e exclusivamente um GC ou se grupos muito pequenos são formados (QUAAS, 1988; CARDOSO & TEMPELMAN, 2004).

A atribuição de “pais fictícios” aos animais com paternidade incerta pode ser uma solução simples e eficaz para aumentar a acurácia das avaliações genéticas comparada à estratégia comumente utilizada, a qual adota paternidade desconhecida para os animais filhos de RM (HENDERSON, 1988; CARDOSO & TEMPELMAN, 2003). Portanto, com o presente trabalho, objetivou-se comparar o modelo assumindo paternidade desconhecida e o modelo de grupos genéticos a fim de indicar o

procedimento estatístico mais adequado para melhor estimar o valor genético de animais que possuem incerteza da paternidade.

## **MATERIAL E MÉTODOS**

Os dados utilizados foram de animais da raça Nelore, nascidos no período de 1984 a 2006, pertencentes ao arquivo zootécnico da Agropecuária Jacarezinho Ltda, localizada no município de Valparaíso, São Paulo. Os objetivos principais na Agropecuária são a venda de reprodutores jovens e animais para abate, para isto conta com um rebanho de 10.000 matrizes e produção de cerca de 8.000 bezerros desmamados por ano.

Touros e vacas são mantidos à pasto com sal mineral à vontade. Realizam-se duas estações de monta, sendo a primeira (antecipada) entre os meses de fevereiro e abril, somente para as novilhas de 16 a 18 meses (aproximadamente, 60 dias), e a segunda em novembro para vacas e as novilhas com dois anos de idade que falharam na monta antecipada (aproximadamente, 70 dias). Todas as novilhas são expostas à reprodução independente do peso e da condição corporal. Os sistemas de acasalamentos utilizados são: inseminação artificial, monta controlada e reprodutor múltiplo (RM), com relação touro:vaca de 1:30. Acasalamentos com RM são definidos pela exposição de um grupo de fêmeas, a mais de um touro simultaneamente na mesma estação de monta. As épocas de nascimento dos bezerros se concentram de agosto a outubro e novembro a janeiro, e os mesmos são mantidos com suas mães até os sete meses de idade à pasto. As novilhas são avaliadas quanto à prenhez por palpação retal, aproximadamente, 60 dias após o término da estação de monta antecipada.

As características consideradas neste estudo foram o ganho de peso pós-desmama (GPD) entre 205 e 550 dias de idade e o peso ao sobreano (PS). As médias e os desvios-padrão foram  $98,40 \pm 33,33$  kg e  $274,34 \pm 42,02$  kg para GPD e PS, respectivamente. Ao sobreano os animais possuíam idades entre 420 e 640 dias, respectivamente.

Durante a consistência dos dados foram consideradas como perdidas as informações de grupos contemporâneos (GC) com menos de três animais e de animais cujas medidas foram superiores ou inferiores à média de seu GC, mais ou menos 3,5 desvios padrão, respectivamente, para as características GPD e PS. Todos os animais tinham pelo menos a mãe conhecida. Os GC foram formados pela combinação dos efeitos sexo, fazenda, ano e estação de nascimento, fazenda, mês e grupo de manejo à desmama, fazenda, mês e grupo de manejo ao sobreano.

Testou-se a conectabilidade entre os GC, com base no número total de laços genéticos, utilizando o programa AMC (ROSO & SCHENKEL, 2006). O programa permite escolher entre várias opções para testar conectabilidade, tais como o modelo (touro e vaca, animal, etc.), o número mínimo de laços genéticos e observações para considerar o grupo contemporâneo conectado. No presente estudo foi utilizado o modelo animal, e aceitou-se GC com, no mínimo, 10 laços genéticos com os demais conectados.

A descrição do arquivo de dados, pedigree e o de grupos de reprodutores múltiplos (RM) foram apresentados na Tabela 1. Após a consistência dos dados foram mantidos 62.212 animais da raça Nelore, filhos de 581 touros e 27.743 vacas. O arquivo de pedigree continha 75.088 animais, sendo 20.827 (37,10%) animais filhos de RM e 12.876 animais pertencentes à população base do arquivo com pais desconhecidos. Em média, os grupos de reprodutores múltiplos possuíam 4,8 touros (variando de 2 a 16).

As características GPD e PS foram modeladas como uma função linear de efeitos fixos (efeitos com *a priori* subjetiva uniforme limitada) e efeitos aleatórios (efeitos com distribuição *a priori* normal multivariada). Os efeitos aleatórios considerados foram: o efeito genético aditivo direto do animal, de GC, efeito genético materno e efeito de ambiente permanente materno. Os dois últimos efeitos foram considerados somente para o PS. Como efeitos fixos incluíram-se, para ambas as características, o efeito da idade do animal ao sobreano, com idade mínima e máxima de 420 e 660 dias, respectivamente, e da idade da vaca (2-16 anos), linear e quadrático.



**Tabela 1-** Descrição dos arquivos pedigree e de dados considerando grupo de reprodutores múltiplos (RM) e grupo de contemporâneos (GC)

|  |         |
|--|---------|
| Número de animais no arquivo de dados        | 62.2121 |
| Número de touros                             | 581     |
| Número de vacas                              | 27.743  |
| Número de animais no arquivo de pedigree     | 75.088  |
| Animais pertencentes a população base        | 12.876  |
| Animais com paternidade conhecida            | 39.402  |
| Animais provenientes de acasalamentos com RM | 22.810  |
| Número de grupos de RM                       | 1.024   |
| Número de grupos de contemporâneos           | 2.301   |

Três abordagens foram adotadas no tratamento de paternidade desconhecida para os animais filhos de RM:

(1) No modelo de paternidade desconhecida, os grupos de RM foram ignorados e para os filhos de RM considerou-se que os pais eram desconhecidos e pertencentes a uma única população base.

Na abordagem de grupos genéticos, foram utilizadas duas definições de grupos genéticos para comparação com o modelo de paternidade desconhecida.

(2) Na primeira definição de grupo genético (GGa), atribuíram-se “pais fictícios” aos animais com incerteza da paternidade, designando que cada grupo genético era formado pelo grupo de RM (formando 444 grupos). Neste caso, o efeito de grupo genético representa uma estimativa do valor genético médio dos touros pertencentes ao grupo de reprodutores múltiplos.

(3) Na outra definição (GGb), também atribuíram-se “pais fictícios” aos animais com incerteza da paternidade. Entretanto, os animais provenientes de acasalamentos com RM foram agrupados de acordo com o ano de nascimento, a cada três anos (baseado no intervalo de gerações dos machos), e neste caso, seis grupos genéticos foram formados.

Embora não houvesse um número pequeno de animais dentro do grupo, esta última definição foi para minimizar possíveis problemas de confundimento de grupos

genéticos com grupo de contemporâneos e o pequeno número de animais dentro dos grupos genéticos, conforme apontado por QUAAS (1988).

O modelo de pais desconhecidos, seguindo as equações de modelos mistos de HENDERSON (1984) pode ser representado pela seguinte equação:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e};$$

em que  $\mathbf{y}$  é o vetor de desempenho,  $\mathbf{b}$  é o vetor de efeitos fixos,  $\mathbf{u}$  é o vetor de efeitos aleatórios;  $\mathbf{e}$  é o vetor de efeito residual e  $\mathbf{X}$  e  $\mathbf{Z}$  são as matrizes de incidência que relacionam as observações aos efeitos não-genéticos e efeito genético aditivo, respectivamente. Assumindo que:

$$p(\mathbf{y}) \sim \mathbf{N}(\mathbf{Xb} + \mathbf{Zu}, I\sigma_e^2);$$

$$p(\mathbf{b}) = \text{constante};$$

$$p(\mathbf{u}) \sim \mathbf{N}(0, \mathbf{G}), \text{ com priores:}$$

$$p(\mathbf{G}_0) \sim IW;$$

$$p(\sigma_p^2) \sim IG;$$

$$p(\sigma_{gc}^2) \sim IG;$$

$$E \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = 0 \quad \text{var} \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & 0 \\ 0 & I\sigma_e^2 \end{bmatrix};$$

Para o GPD:

$$\mathbf{u} = \begin{bmatrix} \mathbf{a} \\ \mathbf{gc} \end{bmatrix} \text{ e } \mathbf{G} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & 0 \\ 0 & I\sigma_{gc}^2 \end{bmatrix};$$

enquanto para o PS o modelo definiu-se por:

$$\mathbf{u} = \begin{bmatrix} \mathbf{a} \\ \mathbf{m} \\ \mathbf{p} \\ \mathbf{gc} \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{am} & 0 & 0 \\ \mathbf{A}\sigma_{am} & \mathbf{A}\sigma_m^2 & 0 & 0 \\ 0 & 0 & \mathbf{I}\sigma_p^2 & 0 \\ 0 & 0 & 0 & \mathbf{I}\sigma_{gc}^2 \end{bmatrix};$$

em que  $\mathbf{a}$  é o vetor de efeito genético aditivo,  $\mathbf{m}$  é o vetor de efeito genético materno,  $\mathbf{p}$  é o vetor de efeito de ambiente permanente materno,  $\mathbf{gc}$  é o vetor de efeito de grupo contemporâneo;  $\mathbf{A}$  é a matriz de parentesco médio,  $\mathbf{I}$  é a matriz identidade,  $\sigma_e^2$  é a variância residual,  $\sigma_a^2$  é a variância genética aditiva,  $\sigma_m^2$  é a variância genética materna,  $\sigma_p^2$  é a variância de ambiente permanente,  $\sigma_{am}$  é a covariância entre os efeitos genéticos aditivo e materno;  $\sigma_{gc}^2$  = variância do grupo de contemporâneos;

As equações de modelo misto incluindo grupos genéticos (WESTELL et al., 1988) podem ser descritas da seguinte maneira:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu}^* + \mathbf{Z}_Q\mathbf{Qg} + \mathbf{e};$$

em que,  $\mathbf{a} = \mathbf{u}^* + \mathbf{Qg}$ , sendo que  $\mathbf{u}^*$  é o vetor dos efeitos aleatórios genético aditivo direto,  $\mathbf{g}$  é o vetor de efeitos fixos aditivos de grupo e  $\mathbf{Q}$  é a matriz de incidência dos efeitos de grupo. Assumindo-se que:

$$\mathbf{p}(\mathbf{g}) = \text{constante};$$

Foram realizadas análises unicaracterística, com todos os modelos, sob enfoque bayesiano, utilizando métodos Monte Carlo via cadeias de Markov (MCMC). Foram considerados 400.000 ciclos. O intervalo de amostragem do ciclo foi de 10 amostras. Os 10.000 primeiros ciclos foram descartados. Para verificar a convergência da cadeia foi realizada uma análise baseada na seqüência inicial monotônica de GEYER (1992). Para cada componente de variância, a seqüência inicial monotônica foi utilizada para determinar o número efetivo de amostras, que é uma estimativa do número de amostras independentes com conteúdo equivalente às informações contidas dentro das amostras dependentes (SORENSEN et al., 1995). O comprimento

da cadeia foi determinado, tal que, todos os parâmetros tivessem pelo menos 100 amostras independentes.

Para verificar o modelo de melhor ajuste foram utilizados o critério de informação da *deviance* (DIC) (SPIEGELHALTER et al., 2002) e a ordenada preditiva condicional (CPO) (GELFAND, 1996). A ordenada preditiva condicional (CPO) foi definida como a aproximação de MCMC para o CPO dos modelos  $M_1$  e  $M_2$ , com parâmetros  $\theta$ , foram obtidos pela média harmônica de  $G$  ciclos de MCMC:

$$p(\mathbf{y}_{ij} | \mathbf{y}_{(-ij)}, \mathbf{M}_r) \approx \frac{1}{\frac{1}{G} \sum_{l=1}^G p^{-1}(\mathbf{y}_{ij} | \theta^{(l)}, \mathbf{M}_r)}$$

em que  $p(\mathbf{y}_{ij} | \mathbf{y}_{(-ij)}, \mathbf{M}_1)$  é a ordenada preditiva condicional (CPO) do modelo 1 para observação  $\mathbf{y}_{ij}$ , é a densidade de validação cruzada, que sugere quais valores de  $\mathbf{y}_{ij}$  são prováveis quando o modelo é ajustado a todas as outras observações  $\mathbf{y}_{(-ij)}$  exceto  $\mathbf{y}_{ij}$ . Enquanto  $p(\mathbf{y}_{ij} | \mathbf{y}_{(-ij)}, \mathbf{M}_2)$  refere-se à ordenada preditiva condicional (CPO) do modelo 2.

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  $\mathbf{M}_r$  foi estimada por  $\mathbf{DEV} = \frac{1}{G} \sum_{l=1}^G -\log p(\mathbf{y} | \theta^{(l)}, \mathbf{M}_r)$ , como proposto por Spiegelhalter et al. (2002). A complexidade do modelo foi determinada pelo número efetivo de parâmetros  $\mathbf{p}_{D(r)} = \mathbf{DEV} - \mathbf{D}_r(\bar{\theta})$ , em que  $\mathbf{DEV}(\bar{\theta}) = -2 \log p(\mathbf{y} | \bar{\theta}, \mathbf{M}_r)$  com  $\bar{\theta}$  sendo a média *a posteriori* do  $\vartheta$ , isto é,  $\mathbf{p}_{D(r)}$  representa a diferença entre a média *a posteriori* da deviance e a deviance baseada na média *a posteriori* dos parâmetros sob o modelo. Finalmente, o DIC para o modelo foi determinado por:

$$\mathbf{DIC} = \mathbf{DEV} + \mathbf{p}_{D(r)}$$

Para comparar modelos, menores valores de DIC e CPO são indicativos de melhor ajuste do modelo. A diferença maior que sete entre os valores de DIC preditos pelos diferentes modelos indicam que a diferença foi importante para o ajuste do modelo (SPIEGELHALTER et al, 2002).

Correlações de Spearman obtidas entre as médias *a posteriori* dos efeitos genéticos aditivos, dos diferentes modelos foram calculadas para o GPD e o PS para diferentes arquivos: (1) contendo somente touros e os touros ranqueados como os 5, 10, 20 e 50% melhores touros para cada característica; e um segundo arquivo (2) contendo animais filhos de RM e filhos de RM classificados da mesma forma que em (1). Além disso, verificou-se o número de coincidentes entre os arquivos contendo os touros e animais com paternidade incerta, para as porcentagens descritas.

Os valores genéticos dos animais filhos de reprodutores múltiplos, para cada característica, foram utilizados para obter estimativas das tendências genéticas anuais, empregando-se o programa computacional Statistic Analysis System (SAS 9.0, SAS Institute, Cary, NC, USA).

## **RESULTADOS E DISCUSSÃO**

Os critérios de seleção de modelos utilizados indicaram que o modelo que assumiu grupos genéticos formados pelos reprodutores do próprio grupo de RM, que é a definição mais fácil, obteve um ajuste inferior comparado ao modelo que não incluiu grupos genéticos (Tabela 2). Na maioria dos trabalhos em que se compararam modelos incluindo grupos genéticos ou não, verificou-se que a inclusão de grupos genéticos levou à predição do mérito genético aditivo dos animais com maior acurácia (FAMULA & VAN VLECK, 1982; PIERAMATI & VAN VLECK, 1993; GOLDEN et al., 1994). Entretanto, FAMULA & VAN VLECK (1982) concluíram que a definição de grupo genético mais apropriada depende do critério de seleção e que uma definição única de grupos genéticos não existe.

Para as duas características, os menores valores de médias da deviance (DEV) foram encontrados para o modelo em que os grupos genéticos foram formados pelos touros pertencentes ao grupo de RM (GGa). Entretanto, o critério de escolha DIC leva

em conta tanto os valores de média da deviance como a complexidade do modelo (ou número efetivo de parâmetros). Assim, devido à sua complexidade, este modelo (GGa) resultou em um critério da informação da deviance superior aos demais (Tabela 2). Desta forma, o resultado do critério de escolha do DIC indicou, tanto para o GPD como para o PS, que o modelo de grupo genético definido com base no intervalo de gerações entre os machos de três anos (GGb), foi mais apropriado para predizer o valor genético dos animais comparado aos outros modelos do presente trabalho (Tabela 2).

O resultado do critério de escolha CPO, avaliado pela deviance baseada na ordenada preditiva condicional (CPO), apontou diferenças entre os modelos de maiores magnitudes que o DIC, mas, da mesma forma, indicou o modelo GGb como o mais apropriado para predizer o valor genético dos animais com paternidade desconhecida para ambas as características .

**Tabela 2** – Média da deviance (DEV), penalidade para número efetivo de parâmetros ( $P_D$ ), critério de informação da deviance (DIC) e deviance baseada na ordenada preditiva condicional (CPO) para o ganho de peso pós-desmama (GPD) e o peso ao sobreano (PS)

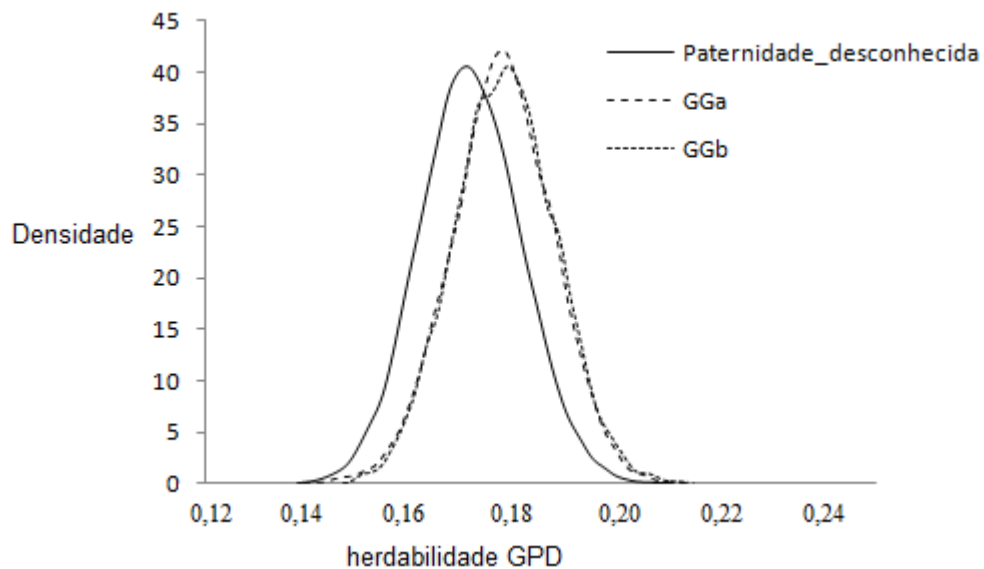
| <b>Característica</b> | <b>Modelo</b>                   | <b>DEV</b> | <b><math>P_D</math></b> | <b>DIC</b> | <b>CPO</b> |
|-----------------------|---------------------------------|------------|-------------------------|------------|------------|
| <b>GPD</b>            |                                 |            |                         |            |            |
|                       | <i>Paternidade desconhecida</i> | 520.553    | 10.710                  | 531.263    | 532.390    |
|                       | <i>GGa</i>                      | 520.174    | 11.223                  | 531.397    | 532.635    |
|                       | <i>GGb</i>                      | 520.268    | 10.917                  | 531.185    | 532.354    |
| <b>PS</b>             |                                 |            |                         |            |            |
|                       | <i>Paternidade desconhecida</i> | 532.437    | 17.874                  | 550.311    | 553.572    |
|                       | <i>GGa</i>                      | 529.399    | 20.249                  | 549.648    | 553.954    |
|                       | <i>GGb</i>                      | 529.957    | 19.485                  | 549.442    | 553.399    |

GGa= modelo em que cada grupo genético era formado pelo grupo de reprodutores múltiplos; GGb= modelo em que os grupos genéticos foram baseados no intervalo de gerações dos machos (3 anos)

Com esses resultados dos critérios de escolha, espera-se que os modelos que incluem grupos genéticos apropriados sejam úteis para predizer o valor genético dos animais com maior acurácia, quando estes possuem alguma informação de parentesco perdida, assim como descrito nos trabalhos de FAMULA & VAN VLECK (1982) e WESTELL et al. (1988). Além disso, esses resultados concordam com PIERAMATI & VAN VLECK (1993) que concluíram que a inclusão de grupos genéticos no modelo é uma alternativa para predizer o valor genético quando informações de pedigree são perdidas.

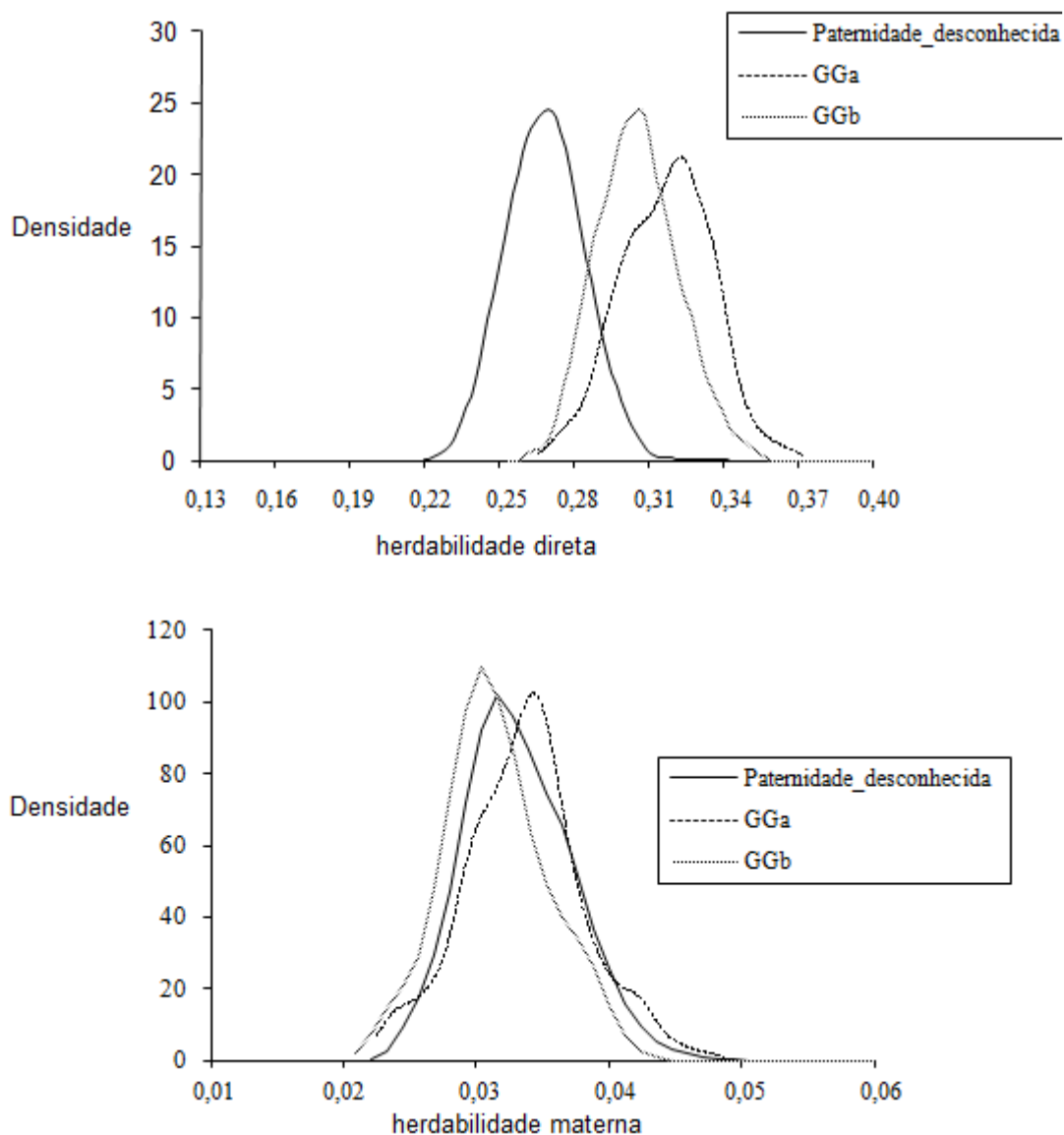
No modelo de paternidade desconhecida, os touros dos filhos de RM são considerados como desconhecidos e pertencentes a uma única população base. Ou seja, os animais pertencentes à população base e os touros participantes do sistema RM terão o mesmo valor genético médio, impossibilitando a comparação do desempenho reprodutivo e produtivo dos diferentes touros. No entanto, essa pressuposição não pode ser considerada verdadeira, uma vez que os animais da população base que não possuem nenhum dos pais conhecidos, podem pertencer a diferentes regiões e/ou gerações de reprodutores, e assim, o valor genético médio difere de população para população. Segundo PIERAMATI & VAN VLECK (1993), a inclusão de grupos genéticos no modelo, pode corrigir parte do viés nas estimativas de variância genética, se os animais da população base não forem todos da mesma geração.

Por meio do teste de convergência realizado com base na seqüência inicial monotônica de GEYER (1992), verificou-se que o número de amostras independentes obtidas, acima de 200 para todos os parâmetros, indicando que o comprimento da cadeia foi adequado. As estimativas da distribuição marginal *a posteriori* das herdabilidades (Figuras 1 e 2) apresentaram dispersão simétrica das amostras em torno da média *a posteriori*, após o descarte das primeiras 10.000 amostras.



**Figura 1-** Densidades das estimativas de herdabilidade *a posteriori* do ganho de peso pós-desmama (GPD) em animais Nelore obtidas pelo modelo de paternidade desconhecida (Paternidade\_desconhecida), modelo considerando os grupos de reprodutores múltiplos como grupos genéticos (GGa) e o modelo considerando o intervalo de gerações de machos (GGb)





**Figura 2-** Densidades das estimativas de herdabilidade *a posteriori* de efeito direto (acima) e materno (abaixo) do peso ao sobreano (PS) em animais Nelore obtida pelo modelo de paternidade desconhecida (Paternidade\_desconhecida), modelo considerando os grupos de reprodutores múltiplos como grupos genéticos (GGa) e o modelo considerando o intervalo de gerações de machos (GGb)

Verificou-se com o presente trabalho que as estimativas de todos os parâmetros e os seus respectivos desvios-padrão, para o ganho de peso, obtidas com os diferentes modelos, foram muito similares (Tabela 3). Assim, a escolha do modelo

não deve interferir na estimação de parâmetros genéticos para esta característica. As estimativas de herdabilidade, variâncias genética aditiva e residual para o GPD estão próximas às descritas na literatura (ELER et al., 1996; MARCONDES et al., 1999; PANETO et al., 2002; CARDOSO et al., 2004; MALHADO et al., 2005; BOLIGON et al., 2010). FERNANDES & FERREIRA (2000) compararam modelos estatísticos para o ganho de peso pós-desmama, obtendo estimativas de herdabilidade superiores (0,36 a 0,37) às encontradas no presente trabalho para animais da raça Charolês. No entanto, as estimativas do presente trabalho são superiores às encontradas por HOLANDA et al. (2004), trabalhando com animais da raça Nelore em Pernambuco (0,06).

**Tabela 3** – Médias das estimativas dos parâmetros genéticos com seus respectivos desvios padrão para ganho de peso pós-desmama obtido pelo modelo de paternidade incerta e modelos com grupos genéticos

| Parâmetros   | Modelo                          |               |               |
|--------------|---------------------------------|---------------|---------------|
|              | <i>Paternidade desconhecida</i> | <i>GGa</i>    | <i>GGb</i>    |
| $\sigma_a^2$ | 52,85 ± 2,91                    | 53,75 ± 2,87  | 54,19 ± 2,95  |
| $\sigma_e^2$ | 252,04 ± 2,70                   | 250,51 ± 2,68 | 250,88 ± 2,72 |
| $h_d^2$      | 0,17 ± 0,01                     | 0,18 ± 0,01   | 0,18 ± 0,01   |

GGa= modelo que considerou que cada grupo genético foi formado pelo grupo de reprodutores múltiplos; GGb= modelo em que os grupos genéticos foram baseados no intervalo de gerações dos machos (3 anos);  $\sigma_a^2$  = estimativa da variância genética aditiva;  $\sigma_e^2$  = estimativa da variância residual;  $h_d^2$  = estimativa da herdabilidade

Para o peso ao sobreano (PS), as estimativas da variância genética aditiva foram próximas entre os modelos que incluíram grupos genéticos sendo que a estimativa obtida com o modelo de paternidade desconhecida foi de menor magnitude (Tabela 4). Esses resultados indicam que os modelos que incluíram grupos genéticos são mais adequados, uma vez que o modelo de paternidade desconhecida pode ter subestimado a variância genética aditiva. Entretanto, PIERAMATI & VAN VLECK

(1993) descreveram estimativas de variância genética aditiva menores com os modelos que incluíram grupos genéticos.

**Tabela 4** - Médias das estimativas dos parâmetros genéticos com seus respectivos desvios-padrão para o peso ao sobreano obtido pelo modelo de paternidade desconhecida e pelos modelos de grupos genéticos

| Parâmetros      | Modelo                          |                |                 |
|-----------------|---------------------------------|----------------|-----------------|
|                 | <i>Paternidade desconhecida</i> | <i>GGa</i>     | <i>GGb</i>      |
| $\sigma_a^2$    | 111,58 ± 7,818                  | 134,74 ± 8,641 | 131,25 ± 8,250  |
| $\sigma_e^2$    | 305,10 ± 5,073                  | 290,55 ± 5,473 | 293,19 ± 5,291  |
| $\sigma_m^2$    | 10,70 ± 1,422                   | 9,77 ± 1,379   | 9,67 ± 1,337    |
| $\sigma_{am}$   | 8,06 ± 2,675                    | 2,26 ± 3,027   | 4,02 ± 2,709    |
| $\sigma_{pe}^2$ | 20,12 ± 1,425                   | 20,21 ± 1,462  | 20,09 ± 1,409   |
| $\sigma_p^2$    | 455,54 ± 3,264                  | 457,59 ± 3,245 | 458,093 ± 3,609 |
| $\sigma_{gc}^2$ | 1015,60 ± 31,48                 | 920,24 ± 25,57 | 983,49 ± 32,08  |
| $h_d^2$         | 0,24 ± 0,016                    | 0,29 ± 0,018   | 0,29 ± 0,017    |
| $h_m^2$         | 0,02 ± 0,002                    | 0,02 ± 0,003   | 0,02 ± 0,003    |
| $rg_{dm}$       | 0,24 ± 0,085                    | 0,07 ± 0,086   | 0,12 ± 0,079    |

GGa= modelo que considerou que cada grupo genético foi formado pelo grupo de reprodutores múltiplos; GGb= modelo que os grupos genéticos foram baseados no intervalo de gerações dos machos (3 anos);  $\sigma_a^2$  = estimativa da variância genética aditiva;  $\sigma_e^2$  = estimativa da variância residual;  $\sigma_m^2$  = estimativa da variância genética materna;  $\sigma_{am}$  = estimativa da covariância entre os efeitos genético direto e materno;  $\sigma_{pe}^2$  = estimativa da variância de ambiente permanente materno;  $\sigma_{gc}^2$  = variância do grupo de contemporâneos;  $\sigma_p^2$  = estimativa da variância fenotípica;  $h_d^2$  = herdabilidade direta;  $h_m^2$  = herdabilidade genética materna;  $rg_{dm}$  = correlação entre os efeitos genéticos direto e materno

A maior magnitude desse parâmetro foi estimado pelo modelo de paternidade desconhecida (Tabela 4). Este modelo proporcionou o pior ajuste aos dados comparado aos modelos que incluíram grupos genéticos, superestimando a variância residual.

As estimativas da covariância e da correlação entre os efeitos genético direto e materno foram positivas (Tabela 4) como as descritas por MEYER (1992) em animais Angus. No entanto, muitos trabalhos da literatura descrevem estimativas de correlações altas e negativas entre efeito genético direto e materno (GUNSKI et al., 2001; RIBEIRO et al., 2001; BOLIGON et al., 2008). No presente trabalho, as estimativas de covariância e correlação entre os efeitos genético direto e materno obtidos pelos diferentes modelos diferiram, sendo o modelo de paternidade desconhecida o que estimou os maiores valores. No entanto, o intervalo de credibilidade a 95% entre os modelos tiveram sobreposição (dados não apresentados), indicando que a escolha entre os modelos pode não interferir na estimação desse parâmetro. A pequena magnitude da correlação entre o efeito genético direto e materno sugere que este não é um efeito muito importante sob ponto de vista biológico (ELER et al., 1995).

Observou-se um acréscimo no valor da herdabilidade direta quando o modelo incluiu grupos genéticos, o que pode ser atribuído, em parte, ao melhor ajuste do modelo com grupos genéticos (Tabela 2). As herdabilidades de efeito direto estimadas no presente trabalho (Tabela 4) são próximas às descritas na literatura para animais de raças Zebuínas (MERCADANTE et al., 1995; BITTENCOURT et al., 2002; GIANNOTTI et al., 2005; BOLIGON et al., 2008) que variaram de 0,22 à 0,48. Com esses resultados, o PS poderá responder mais eficientemente à seleção que o GPD, concordando com o trabalho de GUNSKY et al. (2001). Por outro lado, o PS possui maior correlação com as outras características de crescimento como peso ao nascer e peso adulto que o GPD (ELER et al., 1996; SIQUEIRA et al., 2003; MERCADANTE et al., 2004), o que nem sempre é desejável, pois a seleção para aumento de peso ao sobreano pode gerar bezerros maiores, dificultando o parto e animais adultos mais pesados, que são mais exigentes nutricionalmente.

Baixas herdabilidades maternas foram estimadas para o peso ao sobreano (Tabela 3). Estas são semelhantes às descritas na literatura (GUNSKY et al., 2001; RIBEIRO et al., 2001). Esses resultados concordam com RIBEIRO et al. (2001), os quais sugeriram que a variação fenotípica total para o peso aos 550 dias de idade depende, em grande proporção, do efeito genético aditivo direto.

As correlações de Spearman entre os valores genéticos obtidos pelos modelos de paternidade desconhecida e os modelos de grupos genéticos (GGa e GGb) para o arquivo incluindo somente touros foram altas (acima de 0,80) para ambas as características avaliadas, GPD e PS (Tabela 5). Esses resultados indicam que a utilização destes modelos não causaria mudanças importantes na classificação dos touros. Isto é confirmado pelo grande número de touros coincidentes (entre parênteses na Tabela 5) observado em todas as pressões de seleção (superior à 78%). Com base nos valores genéticos preditos pelo modelo de paternidade desconhecida, GGa e GGb, observou-se que 23 a 24 touros foram coincidentes entre os modelos quando os 5% melhores reprodutores foram avaliados (arquivo total contendo 26 touros) (Tabela 5).

Incluindo-se somente filhos de reprodutores múltiplos, verificou-se para o GPD, que as correlações entre os valores genéticos preditos pelo modelo de paternidade desconhecida e o GGb foram todas acima de 0,80. Assim, a escolha entre esses modelos (paternidade desconhecida e GGb) não provocou mudanças substanciais nas classificações dos animais. Isto é confirmado pelo grande número de filhos de RM coincidentes (entre parênteses na Tabela 6), que seriam selecionados por esses dois modelos. Entretanto, as correlações entre valores genéticos preditos considerando outros modelos (paternidade desconhecida x GGa e GGa x GGb) foram menores que 0,63. Desta forma, a utilização destes modelos pode promover mudança importante nas classificações dos filhos de RM.

Para PS, as correlações entre os valores genéticos preditos pelo modelo de paternidade desconhecida e os modelos de grupos genéticos (GGa e GGb) foram moderadas (inferiores a 0,70). Além disso, para essa característica, as maiores correlações foram entre os modelos de grupo genético (Tabela 6). Estas baixas correlações entre os valores genéticos sugerem que os modelos de grupo genético

(GGa e GGb) podem favorecer a classificação correta dos filhos de RM em comparação com modelo paternidade desconhecida quando a seleção é baseada no PS. Cabe ressaltar que o modelo GGb é o que melhor ajustou os dados observados.

**Tabela 5-** Número de touros coincidentes (entre parêntesis) e as estimativas de correlação de Spearman entre os valores genéticos preditos pelo modelo de paternidade desconhecida e modelos de grupos genéticos para o ganho de peso pós-desmama (GPD) e peso ao sobreano (PS)

|                               | <b>Porcentagem de touros selecionados</b> |            |            |              |
|-------------------------------|---|------------|------------|--------------|
|                               | <b>10%</b>                                | <b>20%</b> | <b>50%</b> | <b>Todos</b> |
| <b>Número total de touros</b> | (52)                                      | (104)      | (260)      | (519)        |
| <b>Características</b>        |   |            |            |              |
| <b>GPD</b>                    |   |            |            |              |
| <i>Paternidade Desc x GGa</i> | (49)                                      | (98)       | (250)      | (519)        |
|                               | 0,869                                     | 0,961      | 0,979      | 0,990        |
| <i>Paternidade Desc x GGb</i> | (50)                                      | (100)      | (252)      | (519)        |
|                               | 0,941                                     | 0,978      | 0,987      | 0,996        |
| <i>GGa x GGb</i>              | (51)                                      | (100)      | (254)      | (519)        |
|                               | 0,961                                     | 0,986      | 0,993      | 0,994        |
| <b>PS</b>                     |   |            |            |              |
| <i>Paternidade Desc x GGa</i> | (41)                                      | (89)       | (244)      | (519)        |
|                               | 0,836                                     | 0,818      | 0,905      | 0,971        |
| <i>Paternidade Desc x GGb</i> | (44)                                      | (94)       | (245)      | (519)        |
|                               | 0,835                                     | 0,861      | 0,933      | 0,981        |
| <i>GGa x GGb</i>              | (49)                                      | (97)       | (254)      | (519)        |
|                               | 0,945                                     | 0,940      | 0,974      | 0,992        |

*Paternidade Desc*= modelo que considerou crias de RM com paternidade desconhecida; GGa= modelo em que cada grupo genético era formado pelo grupo de reprodutores múltiplos; GGb= modelo em que os grupos genéticos foram baseados no intervalo de gerações dos machos (3 anos);

Verificou-se que para o GPD, as correlações de Spearman entre os valores genéticos obtidos pelos modelos do presente trabalho foram, em todos os casos, levemente superiores às observadas para o peso ao sobreano. Provavelmente, isso deve-se ao menor número de efeitos que foram incluídos no modelo para o GPD.

O controle da paternidade por monta controlada e a inseminação artificial, assim como a identificação da paternidade por da tipagem sanguínea e exames de DNA são ferramentas que podem ter alto custo para os produtores de gado de corte, principalmente quando se trabalha com grandes rebanhos. Muitos programas de melhoramento genético desconsideram animais com paternidade desconhecida. No entanto, SULLIVAN, (1995) verificou que a exclusão dos dados de desempenho de animais com paternidade incerta da avaliação pode representar uma redução de 24% na resposta à seleção, comparando o modelo que exclui os dados de desempenho e os modelos considerando grupos genéticos. O mesmo autor concluiu que a perda na resposta à seleção pode depender da porcentagem de animais com paternidade incerta. Portanto, a utilização de modelos que incluam grupos genéticos é uma medida alternativa para melhor predizer o mérito genético de animais com paternidade desconhecida.

As médias dos valores genéticos estimados para o ganho em peso pelos três modelos foram muito próximas à zero até o ano de 1995 (Figura 2). Após esse ano, maiores tendências foram observadas, principalmente, para o modelo de grupo genético formado com base no intervalo de gerações dos machos, o qual apresentou, para o ganho de peso pós-desmama, a maior tendência genética média anual (0,22 kg/ano).

**Tabela 6** - Número de filhos de reprodutores múltiplos (RM) coincidentes (entre parêntesis) e correlação de Spearman entre os valores genéticos preditos pelo modelo de paternidade desconhecida e modelos de grupos genéticos para o ganho de peso pós-desmama (GPD) e peso ao sobreano (PS)

|                               | <b>Porcentagem de filhos de RM selecionados</b> |                 |                  |                  |
|-------------------------------|---|-----------------|------------------|------------------|
|                               | <b>10%</b>                                      | <b>20%</b>      | <b>50%</b>       | <b>Todos</b>     |
| <b>Número de filhos de RM</b> | (2281)  | (4562)          | (11405)          | (22810)          |
| <b>Características</b>        |   |                 |                  |                  |
| <b>GPD</b>                    |   |                 |                  |                  |
| <i>Paternidade Desc x GGa</i> | 0,412<br>(1176)                                 | 0,487<br>(2931) | 0,574<br>(9086)  | 0,743<br>(22810) |
| <i>Paternidade Desc x GGb</i> | 0,822<br>(1862)                                 | 0,852<br>(3940) | 0,895<br>(10543) | 0,954<br>(22810) |
| <i>GGa x GGb</i>              | 0,522<br>(1391)                                 | 0,568<br>(3127) | 0,623<br>(9259)  | 0,762<br>(22810) |
| <b>PS</b>                     |   |                 |                  |                  |
| <i>Paternidade Desc x GGa</i> | 0,294<br>(1142)                                 | 0,354<br>(2685) | 0,442<br>(8555)  | 0,634<br>(22810) |
| <i>Paternidade Desc x GGb</i> | 0,532<br>(1489)                                 | 0,594<br>(3252) | 0,691<br>(9402)  | 0,853<br>(22810) |
| <i>GGa x GGb</i>              | 0,701<br>(1518)                                 | 0,713<br>(3298) | 0,734<br>(9500)  | 0,802<br>(22810) |

*Paternidade Desc*= modelo que considerou crias de RM com paternidade desconhecida; *GGa*= modelo em que cada grupo genético era formado pelo grupo de reprodutores múltiplos (RM); *GGb*= modelo em que os grupos genéticos foram baseados no intervalo de gerações dos machos (3 anos);



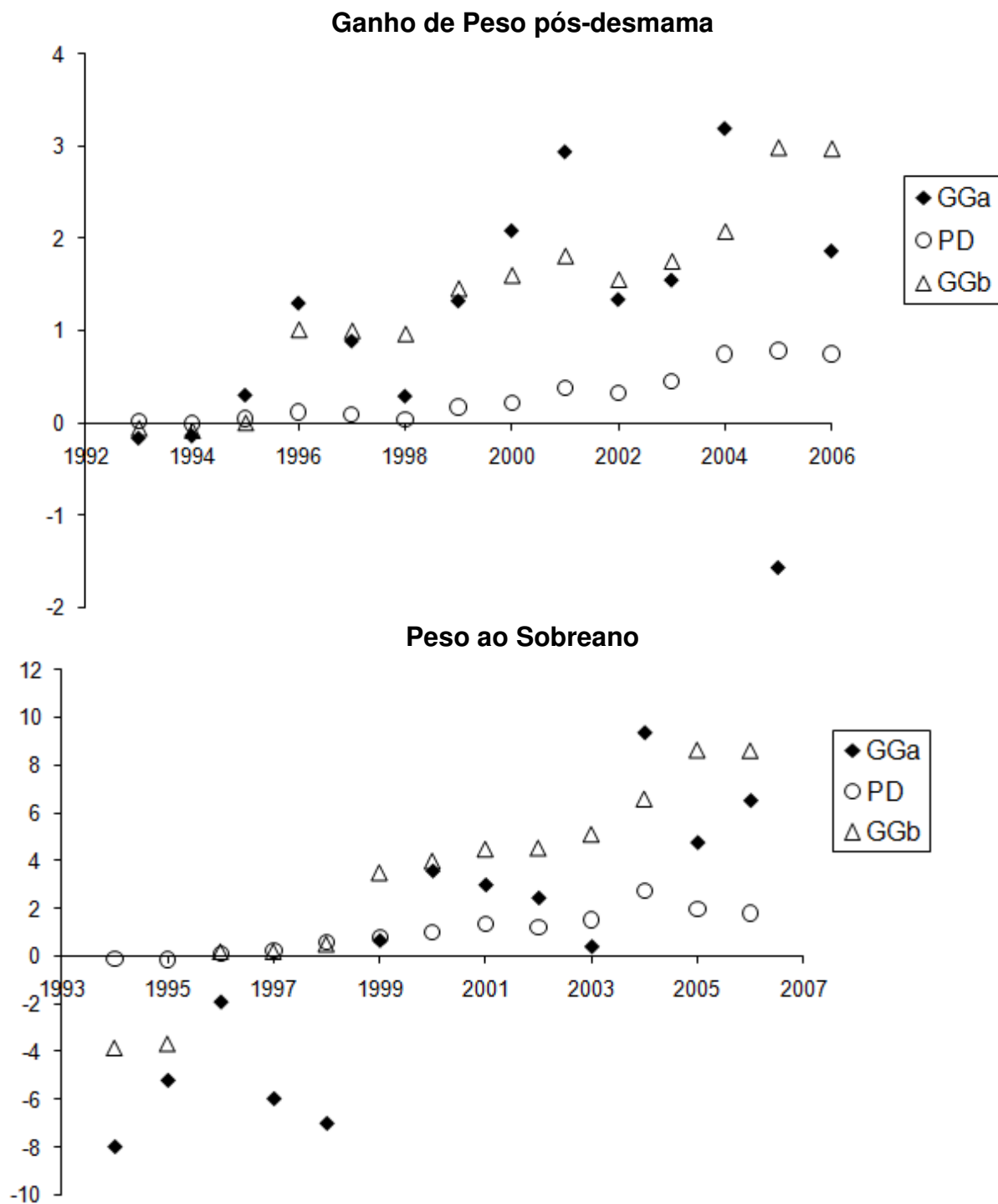


Figura 2- Tendências genéticas anuais dos filhos de reprodutores múltiplos para o ganho de peso pós-desmama e peso ao sobreano obtidos pelo modelo que considerou crias de reprodutores múltiplos (RM) com paternidade desconhecida (PD), modelo em que cada grupo genético era formado pelo grupo de RM (GGa) e o modelo em que os grupos genéticos foram baseados no intervalo de gerações dos machos (3 anos) (GGb)

Para peso ao sobreano, até o ano de 1998, os menores valores genéticos médios foram estimados pelos modelos que incluíram grupos genéticos (GGa e GGb). No entanto, a maior tendência genética foi obtida com o modelo de grupos genéticos formados pelo grupo de RM (1,22 kg/ano). Por outro lado, as médias anuais dos valores genéticos obtidas com este modelo oscilaram muito para as duas características não seguindo uma tendência linear. Isto pode ser devido a contaminação dos valores genéticos por efeitos ambientais neste modelo, em consequência de confundimento entre grupos de RM e GC. As médias anuais estimadas com o modelo GGb foram superiores às estimadas com o modelo de paternidade desconhecida com tendência genética de 1,020 e 1,016 kg/ano, respectivamente (Figura 2). O modelo de paternidade desconhecida por assumir que os touros incluídos nos grupos de RM pertencem à população base subestima o ganho genético, uma vez que, esses touros são selecionados e espera-se tenham um valor genético médio superior ao da população base.

## **CONCLUSÕES**

O modelo de grupo genético, agrupando os animais de acordo com o intervalo de gerações dos machos foi mais adequado para prever o mérito genético dos animais com incerteza de paternidade. Recomenda-se o uso desse modelo para prever o mérito genético e classificar filhos de reprodutores múltiplos.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BERTAZZO, R. P., FREITAS, R. T. F., GONÇALVES, T. M., et al., 2004. Parâmetros genéticos de longevidade e produtividade de fêmeas da raça Nelore. **Rev. Bras. Zootec.**, 33(5), 118-1127.
- BITTENCOURT, T. C. C., ROCHA, J. C. M. C., LÔBO, R. B. Et al., 2002. Estimação de componentes de (co)variâncias e predição de DEP's para características de crescimento pós-desmama de bovinos da raça Nelore, usando diferentes modelos estatísticos. **Arq. Bras. Med. Vet. Zoot.** 54, 303-308.
- BOLIGON, A. A., ALBUQUERQUE, L. G.; RORATO, P. R. N., 2008. Associações genéticas entre pesos e características reprodutivas em rebanhos da raça Nelore. **Rev. Bras. Zoot.**, 37, 596-601.
- BOLIGON, A. A., ALBUQUERQUE, L. G., MERCADANTE, M. E. Z., LÔBO, R. B. 2010. Study of relations among age at first calving, average weight gains and weights from weaning to maturity in Nelore cattle. **Rev. Bras. Zootec.**, 39(4),746-751.
- CANTET, R. J. C., GIANOLA, D., MISZTAL, I., et al., 1993. Estimates of dispersion parameters and of genetic and environmental trends for weaning weight in Angus cattle using a maternal animal-model with genetic grouping. **Livest. Prod. Sci.**, 34(3-4), 203-212.
- CARDOSO, F. F., TEMPELMAN, R. J., 2003. Bayesian inference on genetic merit under uncertain paternity. **Gen. Sel. Evol.**, 35, 469-487.
- CARDOSO, F. F., CARDELLINO, R. A.; CAMPOS, L. T., 2004. (Co)Variance components and genetic parameters of post-weaning traits in Angus cattle. **Rev. Bras. Zootec.**, 33(2), 313-319.
- CARDOSO, F. F.; TEMPELMAN, R. J., 2004. Genetic evaluation of beef cattle accounting for uncertain paternity. **Livest. Prod. Sci.**, 89(2-3), 109-120.
- CARDOSO, F. F., 2008. Manual de utilização do programa INTERGEN – Versão 1.0 em estudos de genética quantitativa animal. p.74 in Série Documentos. E. P. Sul, ed. Embrapa Pecuária Sul, Bagé.

ELER, J. P., VAN VLECK, L. D., FERRAZ, J. B., et al., 1995. Estimation of variances due to direct and maternal effects for growth traits of Nelore cattle. **J. Anim. Sci.**, 73(11), 3253-3258.

ELER, J. P., FERRAZ, J. B. S., SILVA, P. R., 1996. Parâmetros genéticos para peso, avaliação e circunferência escrotal na raça Nelore, estimados por modelo animal. **Arq. Bras. Med. Vet. Zootec.**, 48(2), 203-213.

FAMULA, T. R.; VAN VLECK, D. 1982. Monte Carlo Study of genetic groups in Sire Evaluation. **J. Dairy Sci.**, 65:1286-1293.

FERNANDES H. D., FERREIRA, G. B., 2000. Estudo comparativo de sete diferentes modelos estatísticos para a característica ganho de peso em bovinos de corte. **Rev. Bras. Zootec.**, 29(5), 1340-1348.

GARNERO, A. D. V., LÔBO, R. B., BEZERRA, L. A. F. et al., 2001. Comparação entre Alguns Critérios de Seleção para Crescimento na Raça Nelore. **Rev. Bras. Repr. Ani.**, 30(3), 714-718.

GELFAND, A. E. 1996. Model determination using sampling-based methods. Pages 145-161 in Markov Chain Monte Carlo in practice. 1st ed. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, ed. Chapman & Hall, London.

GEYER, C. J. 1992. Practical Markov Chain Monte Carlo. **Stat. Sci.**, 7(4), 473-511.

GIANNOTTI, J. D. G., PACKER, I. U., MERCADANTE, M. E. Z., et al., 2005. Análise de agrupamento para implementação da meta-análise em estimativas de herdabilidade para características de crescimento em bovinos de corte. **Rev. Bras. Zootec.**, 34, 1165-1172.

GOLDEN, B. L., BOURDON, R. M., SNELLING, W. M. Additive genetic groups for animals evaluated in more than one breed association national cattle evaluation. **J. Anim. Sci.**, 72, 2559-2567.

GUNSKY, R. J., GARNEIRO, A. D. V., BORJAS, A. R. L. et al., 2001. Estimativas de parâmetros genéticos para características incluídas em critérios de seleção em gado Nelore. **Ciência Rural**, 31(4), 603-607.

HENDERSON, C. R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph.

HENDERSON, C. R., 1988. Use of an average numerator relationship matrix for multiple-sire joining, **J. Anim. Sci.**, 66, 1614-1621.

HOLANDA, M. C. R., BARBOSA, S. B. P., RIBEIRO, A. C. et al., 2004. Tendências genéticas para crescimento em bovinos Nelore em Pernambuco, Brasil. **Arch. Zootec.**, 53, 185-194.

KERR, R. J., K. HAMMOND E B. P. KINGHORN. Effects of multiple-sire matings on genetic evaluations, selection response and rates of inbreeding. **Liv. Prod. Sci**, v.38, n.3, p.161-168. 1994.

MALHADO, C. H. M., MARTINS FILHO, R., LÔBO, R. N. B., et al., 2005. Tendências genéticas para características relacionadas à velocidade de crescimento em bovinos Nelore na região nordeste do Brasil. **Rev. Bras. Zootec.**, 34, 60-65.

MERCADANTE, M. E. Z., LÔBO, R. B., BORJAS, A. DE LOS R. 1995 Parâmetros genéticos para características de crescimento em zebuínos de carne. **Arch. Latin. Prod. Anim**, 3(1):45-89.

MERCADANTE, M. E. Z., LÔBO, R. B. 1997. Estimativas de (co)variâncias e parâmetros genéticos dos efeitos direto e materno de características de crescimento de fêmeas de um rebanho Nelore. **Rev. Bras. Zootec.**, 26(6), 1124-1133.

MERCADANTE, M. E. Z., RAZOOK, A. G., FREITAS, J. B. et al., 2004. Parâmetros Genéticos do Peso no Início da Estação de Monta, Considerado Indicativo do Peso Adulto de Matrizes Nelore. **Rev. Bras. Zootec.**, 33(5), 1135-1144.

PANETO, J. C. D. C., LEMOS, D. C., BEZERRA, L. A. F., et al., 2002. Estudo de Características Quantitativas de Crescimento dos 120 aos 550 Dias de Idade em Gado Nelore. **Rev. Bras. Zootec.** 31, 668-674.

QUAAS, R. L. 1988. Additive Genetic Model with Groups and Relationships. **J. Dairy Sci.** 71(5), 1338-1345.

PIERAMATI, C. and VAN VLECK, L. D. 1993. Effect of genetic groups on estimates of additive genetic variance. **J. Anim. Sci.**, 71, 66-70.

POLLAK, E. J., UFFORD, G. R., GROSS, S. J. 1977 Comparison of alternative models for within-herd genetic evaluation of beef cattle. **J. Anim. Sci.**, 45, 1010-1014.

RIBEIRO, M. N., PIMENTA FILHO, E. C., MARTINS, J. L. R. G. A., et al., 2001. Herdabilidade para efeitos direto e materno de características de crescimento de bovinos Nelore no Estado da Paraíba. **Rev. Bras. Zootec.** 30, 1224-1227.

ROSO, V. M. ;F. S. SCHENKEL. 2006. AMC – A computer program to assess the degree of connectedness among contemporary groups. **In: 8th World Congress on Genetics Applied to Livestock Production**, Belo Horizonte, MG, Brasil. Pages communication nº 27-26

SIQUEIRA, R. L. P. G., OLIVEIRA, J. A., Lôbo, R. B., et al., 2003. Análise da variabilidade genética aditiva de características de crescimento na raça Nelore. **Rev. Bras. Zootec.** 32, 99-105.

SORENSEN, D. A., ANDERSEN,S., GIANOLA, D., et al., 1995. Bayesian-inference in threshold models using Gibbs sampling. **Gen. Select. Evol.** 27(3), 229-249.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P., et al., 2002. Bayesian measures of model complexity and fit. **J. R. Statist. Soc. Ser. B-Stat. Methodol.** 64, 583-616.

SULLIVAN P.G. Alternatives for genetic evaluation with uncertain parentage, **Can. J. Anim. Sci.** v. 75, p. 31-36. 1995.

TONG, A. K. W., KENNEDY, B. W., MOXLEY, J. E. 1980. Potential errors in sire evaluation from regional genetic differences. **J. Dairy Sci.**, 63, 627-633.

WESTELL, R. A., QUAAS, R. L.; VAN VLECK, L. D., 1988. Genetic Groups in an Animal Model. **J. Dairy Sci.** 71(5), 1310-1318.

### **CAPÍTULO 3 – COMPARAÇÃO DE PROCEDIMENTOS ESTATÍSTICOS PARA CONSIDERAR PATERNIDADE DESCONHECIDA PARA AVALIAÇÃO GENÉTICA DE CARACTERÍSTICAS DE CRESCIMENTO DE ANIMAIS DA RAÇA NELORE.**

**Resumo:** O objetivo do presente trabalho foi comparar um modelo com base na matriz de parentesco médio (ANRM) e um modelo hierárquico bayesiano (HIER) a fim de indicar o procedimento estatístico mais adequado para melhor estimar o valor genético para característica de crescimento de animais da raça Nelore que possuem incerteza da paternidade, quando os pais candidatos estão disponíveis. Após a consistência dos dados, foram mantidos 62.212 animais da raça Nelore com informações de ganho de peso pós-desmama e peso ao sobreano. O arquivo de pedigree continha 75.088 animais. Duas abordagens foram adotadas no tratamento da incerteza da paternidade. No modelo ANRM foram atribuídas probabilidades iguais para cada um dos possíveis pais dos animais com paternidade incerta. A outra metodologia adotada foi o modelo HIER, que considera a incerteza nas atribuições de paternidade de animais participantes do sistema de RM, em que probabilidades *a priori* fixas são atribuídas a cada possível pai dos animais com paternidade incerta, sendo que as probabilidades *a posteriori* são estimadas de acordo com a informação contida nos dados. Foram realizadas análises unicaracterística, sob enfoque bayesiano, utilizando métodos Monte Carlo via cadeias de Markov (MCMC). Foram considerados 400.000 ciclos, descartando os 10.000 primeiros ciclos (período de aquecimento). Para verificar o modelo de melhor ajuste foram utilizados o critério de informação da deviance (DIC) e a ordenada preditiva condicional (CPO). O modelo que melhor ajustou os dados para estimar os parâmetros genéticos de animais que possuem paternidade incerta foi o modelo hierárquico bayesiano. Entretanto, para avaliações genéticas, qualquer um dos dois modelos pode ser utilizado já que praticamente não houve diferenças entre os valores genéticos estimados

**Palavras-chave:** ganho de peso pós-desmama, modelo hierárquico bayesiano, paternidade desconhecida, peso ao sobreano

**Abstract:** The objective of this work was to compare model based on the use of an average numerator relationship matrix (ANRM) and a hierarchical animal model to indicate the most appropriate statistical procedure to better estimate the genetic value of Nelore animals that have uncertain paternity, when candidate date sires are available. The data set contained records of 62,212 Nelore animals. The pedigree file contained a total of 75,088 animals. Two approaches were adopted to handle uncertain paternity. In the model based on the use of the ANRM equal probabilities were attributed to each of the possible parents of the animals with uncertain paternity. The other method adopted in the present study was the HIER that considers uncertainty in the assignment of paternity of animals participating in the MS system. Within this context, prior probabilities are assigned to each possible sire of animals with uncertain paternity, which are altered according to information present in the data for the estimate corresponding posterior probabilities. Univariate analyses were carried out under Bayesian approach via Markov Chain Monte Carlo (MCMC) methods, implementing a chain of 400,000 rounds where the first 10,000 rounds were discarded (burn-in period). Models were compared by deviance information criteria (DIC) and the conditional predictive ordinate (CPO). The hierarchical animal model was more appropriate to fit the data to estimate genetic parameters with uncertain paternity. However, for genetic evaluations, both models can be used since there was no difference between estimated genetic values.

**Keyword:** Hierarchical bayes model, long yearling weight, postweaning gain, uncertain paternity



## INTRODUÇÃO

Acasalamentos com reprodutores múltiplos (RM) são muito utilizados na pecuária extensiva, uma vez que não requer mão-de-obra tão qualificada quanto outros sistemas de acasalamentos. Este sistema de acasalamentos também é utilizado para melhorar a taxa de concepção, uma vez que o tamanho médio do grupo de reprodução é formado em função do tamanho do piquete e, em geral, esses piquetes possuem grandes áreas, assim, a exposição de um grupo de vacas a um único touro se torna praticamente inviável. A consequência da inclusão de vários touros em um grupo grande de vacas é a formação de pedigrees incompletos, os quais podem afetar diretamente a acurácia das avaliações genéticas.

Diferentes modelos estatísticos têm sido propostos para predição do mérito genético de animais com paternidade incerta. Um método inicialmente proposto foi a inclusão de grupos genéticos no modelo a fim de minimizar o viés nas predições dos valores genéticos de animais com paternidade desconhecida (WESTELL et al., 1988). Outra metodologia é o modelo com base na matriz de parentesco médio (ANRM), proposto por HENDERSON (1988), que pressupõe o conhecimento da probabilidade de cada touro ser o pai verdadeiro de um determinado animal. O ANRM tem propriedades de melhor preditor linear não viesado (BLUP), fundamentado na matriz de parentesco médio, assim, auxilia a correta especificação da matriz de (co)variâncias genéticas, quando essas probabilidades são presumidamente conhecidas.

Em 2003, foi proposto um outro método para avaliação genética de animais com paternidade desconhecida (CARDOSO & TEMPELMAN, 2003), o modelo animal hierárquico (HIER). Este método combina informações a priori aos dados a fim de determinar probabilidades a posteriori dos pais candidatos. Desta forma, o modelo infere o mérito genético dos animais com paternidade incerta e seus possíveis pais. A principal vantagem do modelo HIER em relação aos modelos anteriormente descritos na literatura é a possibilidade de considerar adequadamente a incerteza sobre as atribuições de touros na inferência do mérito genético, devido à incerteza sobre as atribuições de touros. Com base nos critérios bayesianos de escolha do modelo,

CARDOSO & TEMPELMAN (2004) mostraram que o HIER foi consistentemente melhor para prever o valor genético de animais com paternidade incerta comparado ao modelo ANRM. Além disso, a inclusão de informações fenotípicas para inferir sobre a incerteza da paternidade pode aumentar a acurácia do valor genético predito, como descrito na literatura (CARDOSO & TEMPELMAN, 2003; 2004; SAPP et al., 2007).

Trabalhos comparando modelos que levam em conta a paternidade desconhecida para avaliação genética de bovinos de corte não são frequentes, principalmente na raça Nelore. Portanto, objetivou-se com esse trabalho comparar o modelo baseado na matriz de parentesco médio (ANRM) e o modelo hierárquico animal (HIER) para animais da raça Nelore criados à pasto a fim de indicar o procedimento estatístico mais adequado para estimar o valor genético de animais que possuem incerteza da paternidade e os possíveis pais estão disponíveis.

## **MATERIAL E MÉTODOS**

Foram utilizados registros de animais da raça Nelore, nascidos no período de 1984 a 2006, pertencentes ao arquivo zootécnico da Agropecuária Jacarezinho Ltda, localizada no município de Valparaíso, São Paulo.

São utilizados como sistemas de acasalamentos: inseminação artificial, monta controlada e reprodutor múltiplo (RM), com relação touro:vaca de 1:30. Acasalamento com RM é a exposição das fêmeas a mais de um touro na mesma estação de monta. São realizadas duas estações de monta. Na primeira, denominada estação de monta antecipada, todas as novilhas são expostas à reprodução independente do peso e da condição corporal entre os meses de fevereiro e abril, com duração de, aproximadamente, 60 dias. A segunda, denominada estação de monta normal, tem início em torno da segunda quinzena do mês de novembro com duração de, aproximadamente, 70 dias. Nesta estação, são expostas as novilhas que não conceberam na estação de monta antecipada e todas as vacas do rebanho. As épocas de nascimento dos bezerros se concentram de agosto a outubro e novembro a janeiro, e os mesmos são mantidos com suas mães até os sete meses de idade à pasto.

As características consideradas neste estudo foram o ganho de peso pós-desmama (GPD), dos 205 aos 550 dias de idade, com média de  $98,40 \pm 33,33$  kg e o peso ao sobreano (PS) com média de  $274,34 \pm 42,02$  kg. Ao sobreano os animais possuíam idades entre 420 e 640 dias, respectivamente.

Informações de grupos contemporâneos (GC) com menos de três animais e de animais cujas medidas foram superiores ou inferiores à média de seu GC, mais ou menos 3,5 desvios padrão foram consideradas como perdidas durante a consistência dos dados, para as características GPD e PS, respectivamente. Todos os animais tinham pelo menos a mãe conhecida. Os GC foram formados pela combinação dos efeitos de sexo, fazenda, ano e estação de nascimento, fazenda, mês e grupo de manejo à desmama, fazenda, mês e grupo de manejo ao sobreano.

Foi realizado um teste de conectabilidade de GC baseado no número total de laços genéticos (mínimo 10), em um modelo animal, usando o programa AMC (ROSO & SCHENKEL, 2006).

Os resumos dos arquivos de dados, pedigree e o de grupos de reprodutores múltiplos (RM) estão apresentados na Tabela 1. Após a consistência dos dados foram mantidos 62.212 animais, sendo que o arquivo de pedigree continha 75.088 animais. Os touros que compõem cada grupo de RM foram ou não identificados, conforme o controle zootécnico de cada propriedade. No presente trabalho, entre os 1.024 grupos de RM, apenas 177 possuíam a identificação dos touros participantes, ou seja, 847 grupos de RM não tinham nenhuma informação sobre os touros participantes. Dos 177 grupos com os touros identificados, 153 estavam conectados e possuíam informação de desempenho ou progênie. O modelo de grupos genéticos foi adotado para grupos de RM sem identificação dos touros participantes. Os grupos genéticos foram formados atribuindo-se “pais fictícios” aos animais com paternidade desconhecida, que foram agrupados de acordo com seus anos de nascimento, a cada três anos, baseado no intervalo de gerações dos machos, com base nos resultados do capítulo 2.

No arquivo com grupos de RM que continham as informações dos touros (177 grupos), foram identificados 559 touros dentro dos grupos de RM. Entre estes, 40 touros não possuíam desempenho ou progênie conhecida. Somente seis touros não possuíam desempenho, mas possuíam progênie identificada, 351 possuíam

desempenho próprio, mas não possuíam progênie identificada e 162 possuíam progênie e informação de desempenho. Em média, os grupos de reprodutores múltiplos possuíam 4,8 touros (variando de 2 a 16).

**Tabela 1** - Descrição do arquivo de dados, pedigree e grupos de reprodutores múltiplos (RM)

|  |        |
|--|--------|
| <b>Número de animais no arquivo</b>                      | 62.212 |
| <i>Número de touros</i>                                  | 581    |
| <i>Número de fêmeas</i>                                  | 27.743 |
| <b>Número de animais no pedigree</b>                     | 75.088 |
| <i>Animais pertencentes a população base</i>             | 12.876 |
| <i>Animais com paternidade conhecida</i>                 | 39.402 |
| <i>Animais provenientes de acasalamentos com RM</i>      | 22.810 |
| <b>Número de grupos de RM</b>                            | 1.024  |
| <i>Grupos com identificação dos touros participantes</i> | 177    |
| <i>Grupos sem identificação dos touros participantes</i> | 847    |

As características GPD e PS foram modeladas como uma função linear de efeitos fixos (efeitos com *a priori* subjetivas uniformes limitadas) e efeitos aleatórios (efeitos com distribuição *a priori* normal multivariada). Os efeitos aleatórios considerados foram: os efeitos genético direto dos animais, de GC, efeito genético materno e efeito de ambiente permanente. Os dois últimos efeitos foram considerados somente para o PS. Como efeitos fixos incluíram-se como covariáveis, para ambas as características, o efeito da idade do animal ao sobreano e idade da vaca (2-16 anos), efeito linear e quadrático.

Duas abordagens foram adotadas no tratamento da incerteza de paternidade desconhecida. No modelo baseado na matriz de parentesco médio (ANRM) proposto por HENDERSON (1988) foram atribuídas probabilidades iguais para cada um dos possíveis pais dos animais com paternidade incerta. Essas probabilidades são utilizadas para obtenção do coeficiente de parentesco aditivo do animal com paternidade incerta.

O cálculo do parentesco aditivo seguiu a seguinte estrutura:

$$\mathbf{a}_{ij} = \frac{1}{2} [\mathbf{a}_{is_1}^{(j)} \mathbf{p}_1^{(j)} + \dots + \mathbf{a}_{is_g}^{(j)} \mathbf{p}_g^{(j)} + \mathbf{a}_{id}^{(j)}] = \frac{1}{2} \left[ \sum_{k=1}^g \mathbf{a}_{is_k}^{(j)} \mathbf{p}_k^{(j)} + \mathbf{a}_{id}^{(j)} \right];$$

em que,  $\mathbf{a}_{ij}$  é o parentesco aditivo entre o animal  $i$  e  $j$ ;  $\mathbf{a}_{is_k}^{(j)}$  é o parentesco aditivo entre o animal  $i$  e cada um dos possíveis pais ( $k = 1, \dots, g$ ) do animal  $j$ ;  $\mathbf{p}_k^{(j)}$  é a probabilidade do touro  $k$  ser o pai do animal  $j$ ;  $\mathbf{a}_{id}^{(j)}$  é o parentesco aditivo entre o animal  $i$  e a mãe do animal  $j$ .

A outra metodologia adotada no presente trabalho foi o modelo hierárquico bayesiano proposto por CARDOSO & TEMPELMAN (2003), que considera a incerteza nas atribuições de paternidade de animais participantes do sistema de RM, em que probabilidades *a priori* fixas são atribuídas a cada possível pai dos animais com paternidade incerta, sendo que as probabilidades *a posteriori* são alteradas de acordo com a informação contida nos dados.

Esse modelo pode ser descrito em quatro estágios:

*1º Estágio:* O desempenho dos animais é descrito como uma função linear de efeitos genéticos aditivos e não-genéticos.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_3\mathbf{gc} + \mathbf{Z}_4\mathbf{pe} + \mathbf{e};$$

resultando na seguinte distribuição condicional de amostragem dos dados:

$$\mathbf{y} | \mathbf{a}, \mathbf{b}, \mathbf{m}, \mathbf{gc}, \sigma_e^2 \sim \mathbf{N}(\mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_3\mathbf{gc} + \mathbf{Z}_4\mathbf{pe}, \mathbf{I}\sigma_e^2);$$

em que  $\mathbf{y}$  é o vetor de desempenho,  $\mathbf{b}$  é o vetor de efeitos não-genéticos,  $\mathbf{a}$  é o vetor de efeitos genéticos aditivos;  $\mathbf{m}$  é o vetor de efeitos genéticos maternos e  $\mathbf{pe}$  é o vetor de efeito de ambiente permanente materno, ambos foram considerados somente para o PS;  $\mathbf{gc}$  é o vetor de efeitos de grupos de contemporâneos;  $\mathbf{e}$  é o vetor de efeito residual, com  $\mathbf{e} | \sigma_e^2 \sim \mathbf{N}(0, \mathbf{I}\sigma_e^2)$ ;  $\mathbf{X}$ ,  $\mathbf{Z}_1$ ,  $\mathbf{Z}_2$ ,  $\mathbf{Z}_3$  e  $\mathbf{Z}_4$  são matrizes de incidência que

relacionam as observações aos efeitos não-genéticos, efeito genético aditivo, materno, de grupo contemporâneo e ambiente permanente materno, respectivamente;  $\sigma_e^2$  é a variância residual,  $\mathbf{I}$  é a matriz identidade;

*2º Estágio:* A distribuição *a priori* para efeitos genéticos aditivos e não-genéticos são descritos por:

Efeito não-genético (**b**):  $\mathbf{b} \sim \mathbf{N}(\mathbf{b}_0, \mathbf{V}_b)$ ;

Efeito genético aditivo (**a**):  $\mathbf{a} | \mathbf{s}, \sigma_a^2 \sim \mathbf{N}(0, \mathbf{A}\sigma_a^2)$ ;

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{m} \end{bmatrix} | \mathbf{S}, \mathbf{G}_0 \sim \mathbf{N}(0, \mathbf{G})$$

$$\mathbf{G} = \mathbf{A} \otimes \mathbf{G}_0$$

$$\mathbf{G}_0 = \mathbf{N}(0, \mathbf{G})$$

$$\mathbf{G}_0 = \begin{bmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{bmatrix}$$

Efeito de grupos de contemporâneos:  $\mathbf{gc} \sim \mathbf{N}(0, I\sigma_{gc}^2)$

Efeito de ambiente permanente materno:  $\mathbf{pe} \sim \mathbf{N}(0, I\sigma_{pe}^2)$

Para a predição do valor genético do animal, (co)variâncias do efeito genético aditivo baseiam-se na matriz de parentesco ( $A$ ), touros candidatos ( $s$ ), variância genética ( $\sigma_a^2$ );  $\sigma_{gc}^2$  é a variância dos efeitos de grupos de contemporâneos;  $\sigma_{pe}^2$  é a variância efeito de ambiente permanente materno;  $\mathbf{V}_b$  é a uma matriz diagonal da variância *a priori* de  $\mathbf{b}$ , assumindo-se  $\mathbf{V}_b \propto \infty$ . A distribuição qui-quadrado escalonada invertida é assumida *a priori* para a variância residual ( $\sigma_e^2$ ), ou seja:

$$\sigma_e^2 \sim \mathbf{S}_e^2 \chi_v^{-2};$$

em que  $\mathbf{S}_e^2$  é o valor *a priori* da variância residual baseado em informações da literatura e  $\nu$  é o grau de confiança nesse valor (SORENSEN & GIANOLA, 2002).

*3º Estágio:* Para a distribuição *a priori* da variância genética, assume-se que:

A densidade *a priori* Wishart invertida:  $\mathbf{G}_0 \sim \mathbf{IW}_{\nu(a)}(\mathbf{S}_a)$

$$\sigma_{pe}^2 \sim \mathbf{S}_{pe}^2 \chi_{\nu(pe)}^2;$$

$$\sigma_{gc}^2 \sim \mathbf{S}_{gc}^2 \chi_{\nu(gc)}^2$$

em que  $\mathbf{S}_a^2, \mathbf{S}_{pe}^2, \mathbf{S}_{gc}^2$  são os valores *a priori* da variância genética aditiva, variância de efeito de ambiente permanente materno e variância dos efeitos de grupos de contemporâneos, respectivamente, baseados na literatura e  $\nu(a), \nu(pe), \nu(gc)$  são os graus de confiança desses valores.

Além disso, neste estágio são especificadas a probabilidade que  $s_j$  seja o pai verdadeiro do animal  $j$  dado por:

$$\pi_j^{(k)} = \text{Prob}(\mathbf{s}_j = \mathbf{s}_j^{(k)}) \text{ para os pais candidatos } \mathbf{k} = 1, 2, \dots, \mathbf{v}_j;$$

sendo o conjunto de probabilidades *a priori* de cada touro  $\mathbf{v}_j$  ser o pai verdadeiro do animal  $\mathbf{j}$ :

$$\boldsymbol{\pi}_j = \{\pi_j^{(1)}, \pi_j^{(2)}, \dots, \pi_j^{(\mathbf{v}_j)}\} \text{ para os } \mathbf{v}_j \text{ pais candidatos do animal } j.$$

*4º Estágio:* A distribuição *a posteriori* para probabilidades de paternidade é descrita nesse estágio. O conjunto inteiro de probabilidades para todos os animais que não pertencem à base, raramente é conhecido com absoluta certeza, podendo ser considerado como quantidades aleatórias da distribuição de Dirichlet, no quarto estágio do modelo:

$$p(\boldsymbol{\pi}|\boldsymbol{\alpha}) = \prod_{j=q_b+1}^q \prod_{k=1}^{\mathbf{v}_j} (\pi_j^{(k)})^{\alpha_j^{(k)}}.$$

Onde  $\alpha_j = \{\alpha_j^{(k)}\}_{k=1}^{v_j}$ ,  $\alpha_j^{(k)} > 0$  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  $\alpha = \{\alpha_j\}_{j=q_b+1}^q$  podem ser baseadas na avaliação da confiabilidade de uma fonte de informação externa sobre as probabilidades *a priori* de cada touro designado.

Foram realizadas análises unicaracterísticas, com ambos os modelos, sob enfoque bayesiano, utilizando métodos Monte Carlo via cadeias de Markov (MCMC). Foram considerados 400.000 ciclos. O intervalo entre as amostras foi de 10 unidades. Para verificar a convergência da cadeia foi realizada uma análise baseada na seqüência inicial monotônica de GEYER (1992). Foram descartados os 10.000 primeiros ciclos.

Por meio da seqüência inicial monotônica de GEYER (1992) foi calculado o tamanho efetivo da amostra (TEA) de cada componente de variância. Com o resultado do TEA pode-se verificar se o comprimento da cadeia foi adequado. O número mínimo esperado para todos os componentes foi atingido para ambas as características avaliadas, que foi acima de 100 amostras.

O critério de informação da *deviance* (DIC) (SPIEGELHALTER et al., 2002) e a ordenada preditiva condicional (CPO) (GELFAND, 1996) foram utilizados para indicar o modelo que melhor ajustou os dados. 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  $\mathbf{M}_r$  foi estimada por  $\mathbf{DEV} = \frac{1}{G} \sum_{i=1}^G -\log p(\mathbf{y}|\theta^{(i)}, \mathbf{M}_r)$ , como proposto por Spiegelhalter et al. (2002). A complexidade do modelo foi determinada pelo número efetivo de parâmetros  $\mathbf{p}_{D(r)} = \mathbf{DEV} - \mathbf{D}_r(\bar{\theta})$ , em que  $\mathbf{DEV}(\bar{\theta}) = -2\log p(\mathbf{y}|\bar{\theta}, \mathbf{M}_r)$  com  $\bar{\theta}$  sendo a média *a posteriori* do  $\theta$ , isto é,  $\mathbf{p}_{D(r)}$  representa a diferença entre a média *a posteriori* da deviance e a *deviance* baseada na média *a posteriori* dos parâmetros sob o modelo. Finalmente, o DIC para o modelo foi determinado por:



$$\text{DIC} = \text{DEV} + \mathbf{p}_{D(r)}$$

A ordenada preditiva condicional (CPO) foi definida como aproximações de MCMC para o CPO dos modelos  $M_1$  e  $M_2$ , com parâmetros  $\theta$ , foram obtidos pela média harmônica de  $G$  ciclos de MCMC:

$$\mathbf{p}(\mathbf{y}_{ij} | \mathbf{y}_{(-ij)}, \mathbf{M}_r) \approx \frac{1}{\frac{1}{G} \sum_{l=1}^G \mathbf{p}^{-1}(\mathbf{y}_{ij} | \theta^{(l)}, \mathbf{M}_r)}$$

em que  $\mathbf{p}(\mathbf{y}_{ij} | \mathbf{y}_{(-ij)}, \mathbf{M}_1)$  é a ordenada preditiva condicional (CPO) do modelo 1 para observação  $\mathbf{y}_{ij}$ , destinada a ser a densidade de validação cruzada, que sugere quais valores de  $\mathbf{y}_{ij}$  são prováveis quando o modelo é ajustado a todas as outras observações  $\mathbf{y}_{(-ij)}$  exceto  $\mathbf{y}_{ij}$ . Enquanto  $\mathbf{p}(\mathbf{y}_{ij} | \mathbf{y}_{(-ij)}, \mathbf{M}_2)$  refere-se à ordenada preditiva condicional (CPO) do modelo 2.

Os menores valores de DIC e CPO são indicativos de melhor ajuste para comparação de modelos. Segundo SPIEGELHALTER et al., (2002), os valores de DIC com diferença maior que sete entre os valores preditos pelos diferentes modelos indicam uma diferença importante para o ajuste do modelo.

Correlação de Spearman entre a média *a posteriori* dos valores genéticos obtidos pelos modelos HIER e ANRM foi calculada para o GPD e PS para seguintes arquivos: O primeiro (completo) incluiu todos os animais do pedigree. O segundo conjunto incluiu exclusivamente progênies provenientes de acasalamentos de RM, ou seja, animais com paternidade incerta. O terceiro incluiu somente os touros. Dentro de cada conjunto, os animais foram classificados pelo seu valor genético. Os 10, 30 e 50% melhores animais foram utilizados para calcular as correlações entre as médias dos valores genéticos obtidos pelos modelos HIER e ANRM.

## RESULTADOS E DISCUSSÃO

O modelo hierárquico foi mais complexo em relação ao ANRM (Tabela 2), concordando com os resultados de CARDOSO & TEMPELMAN (2004), que trabalharam com dados de animais Hereford. Apesar desta maior complexidade, os valores das médias da deviance (DEV) foram inferiores para HIER levando a um menor DIC. Verificou-se que a diferença entre os DIC dos modelos ANRM e HIER foi de 45 para o GPD e 255 para o PS (Tabela 2). Estas diferenças são maiores que sete que, segundo SPIEGELHALTER et al. (2002), é um valor suficiente para indicar uma diferença importante entre modelos, em termos de ajuste dos dados. Calculado pela *deviance* baseada na ordenada predictiva condicional (CPO), as diferenças entre os modelos foram 224 para o GPD e 193 para o PS (Tabela 2).

**Tabela 2** – Média da deviance (DEV), penalidade para número efetivo de parâmetros (PD), critério de informação da deviance (DIC) e deviance baseada na ordenada predictiva condicional (CPO) para o ganho de peso pós desmama (GPD) e o peso ao sobreano obtido pelo modelo baseado na matriz de parentesco médio (ANRM) e modelo hierárquico bayesiano (HIER)

|            | <b>Modelo</b> | <b>DEV</b> | <b>PD</b> | <b>DIC</b> | <b>CPO</b> |
|------------|---------------|------------|-----------|------------|------------|
| <b>GPD</b> | <i>ANRM</i>   | 520.267    | 10.862    | 531.129    | 522.486    |
|            | <i>HIER</i>   | 520.110    | 10.975    | 531.085    | 522.262    |
| <b>PS</b>  | <i>ANRM</i>   | 530.093    | 19.400    | 549.493    | 533.822    |
|            | <i>HIER</i>   | 529.371    | 19.867    | 549.238    | 533.629    |

ANRM= modelo baseado na matriz de parentesco médio; HIER= modelo hierárquico animal

Desta forma, os critérios de escolha de modelos utilizados no presente trabalho, DIC e CPO, consistentemente indicaram o modelo HIER como o mais apropriado para prever o valor genético dos animais com paternidade desconhecida para ambas as

características avaliadas. Estes resultados concordam com os apresentados na literatura por CARDOSO & TEMPELMAN (2003; 2004), que concluíram que o modelo HIER foi mais apropriado que o ANRM para inferir o mérito genético de animais que possuem paternidade desconhecida, quando os componentes de variância foram considerados conhecidos e desconhecidos. Além disto, CARDOSO & TEMPELMAN (2004) observaram que o modelo HIER produziu estimativas de valores genéticos para animais com paternidade incerta com menor viés.

Nas Tabelas 3 e 4 esta apresentado o tamanho efetivo da amostra (TEA) de cada componente de variância, calculado por meio da seqüência inicial monotônica de GEYER (1992). O TEA estima o número de amostras independentes com informações equivalentes às contidas dentro da amostragem dependente (SORENSEN et al., 1995). O número mínimo esperado para todos os componentes foi atingido por ambas as características avaliadas (acima de 100 amostras). Assim, verificou-se que o comprimento da cadeia (400.000 ciclos) utilizado foi adequado.

**Tabela 3** – Média *a posteriori* (MP), intervalos de credibilidade à 95% (IC\_95%), tamanho efetivo da amostra (TEA) para os parâmetros genéticos do ganho pós-desmama obtido pelo modelo baseado na matriz de parentesco médio (ANRM) e modelo hierárquico bayesiano (HIER)

| Modelo | Parâmetro    | MP     | IC_95%          | TEA  |
|--------|--------------|--------|-----------------|------|
| ANRM   | $\sigma_a^2$ | 60,62  | 54,19 à 67,29   | 1396 |
|        | $\sigma_e^2$ | 250,88 | 245,52 à 256,25 | 2133 |
|        | $h_d^2$      | 0,19   | 0,18 à 0,21     | 1406 |
| HIER   | $\sigma_a^2$ | 61,80  | 55,24 à 68,54   | 1275 |
|        | $\sigma_e^2$ | 250,24 | 244,87 à 255,59 | 1968 |
|        | $h_d^2$      | 0,20   | 0,18 à 0,22     | 1284 |

$\sigma_a^2$  = variância genética aditiva;  $\sigma_e^2$  = variância residual;  $h_d^2$  = herdabilidade de efeito direto

**Tabela 4** – Média *a posteriori* (MP), intervalos de credibilidade à 95% (IC\_95%), tamanho efetivo da amostra (TEA) para os parâmetros genéticos do peso ao sobreano (PS) obtido pelo modelo baseado na matriz de parentesco médio (ANRM) e modelo hierárquico bayesiano (HIER)

| Modelo | Parâmetro       | MP      | IC_95%           | TEA   |
|--------|-----------------|---------|------------------|-------|
| ANRM   | $\sigma_a^2$    | 147,34  | 127,70 à 168,61  | 277   |
|        | $\sigma_m^2$    | 10,33   | 7,97 à 13,44     | 210   |
|        | $\sigma_{am}$   | 3,08    | -4,02 à 9,49     | 211   |
|        | $\sigma_{pe}^2$ | 20,28   | 17,63 à 23,12    | 2135  |
|        | $\sigma_{gc}^2$ | 1013,88 | 956,23 à 1075,97 | 35642 |
|        | $\sigma_e^2$    | 293,84  | 282,40 à 305,04  | 365   |
|        | $h_d^2$         | 0,31    | 0,27 à 0,35      | 268   |
|        | $h_m^2$         | 0,02    | 0,01 à 0,03      | 215   |
|        | $rg_{dm}$       | 0,08    | -0,09 à 0,26     | 202   |
| HIER   | $\sigma_a^2$    | 155,64  | 134,65 à 177,61  | 231   |
|        | $\sigma_m^2$    | 10,35   | 8,04 à 13,70     | 178   |
|        | $\sigma_{am}$   | 1,27    | -6,26 à 8,21     | 138   |
|        | $\sigma_{pe}^2$ | 20,25   | 17,61 à 23,11    | 2533  |
|        | $\sigma_{gc}^2$ | 1013,49 | 954,20 à 1076,48 | 36928 |
|        | $\sigma_e^2$    | 290,46  | 278,50 à 301,93  | 301   |
|        | $h_d^2$         | 0,32    | 0,28 à 0,37      | 223   |
|        | $h_m^2$         | 0,02    | 0,01 à 0,03      | 184   |
|        | $rg_{dm}$       | 0,04    | -0,14 à 0,22     | 138   |

$\sigma_a^2$  = variância genética aditiva;  $\sigma_e^2$  = variância residual;  $\sigma_m^2$  = variância genética materna;  $\sigma_{am}$  = covariância entre os efeitos direto e materno;  $\sigma_{pe}^2$  = estimativa da variância de ambiente permanente materno;  $\sigma_{gc}^2$  = variância do grupo de

contemporâneos;  $h_d^2$  = herdabilidade direta;  $h_m^2$  = herdabilidade genética materna;  $rg_{dm}$  = correlação entre os efeitos genéticos direto e materno

O TEA do GPD foi acima de 1200 amostras para todos os componentes de variância, além disso, o modelo ANRM apresentou maior número de amostras independentes que o HIER. Portanto, esses resultados são diferentes dos descritos por CARDOSO & TEMPELMAN (2004), para o GPD de animais Hereford, verificaram que o maior número de amostras independentes foi para HIER comparado ao ANRM. O número de amostras independentes obtidos para os componentes de variância do PS foi menor que para o GPD, provavelmente, por ser um modelo mais complexo.

O modelo HIER considera dados de desempenho e probabilidade *a priori* dos pais candidatos para inferir probabilidades *a posteriori* de um determinado touro ser o pai verdadeiro de um animal com paternidade desconhecida. Assim, espera-se que esse modelo infira o mérito genético dos animais com maior precisão comparado ao modelo ANRM. No entanto, verificou-se no presente trabalho que IC<sub>95%</sub> e o desvio-padrão (dados não apresentados) para GPD e PS, foram um pouco maiores para o modelo HIER (Tabela 3 e 4). Além disso, o IC<sub>95%</sub> de cada componente de variância estimado por ambos os modelos, sobrepõe-se em todos os casos, indicando que praticamente não há diferenças entre os modelos ao estimar esses parâmetros.

Em geral, as médias *a posteriori* dos componentes de variâncias estimados pelos modelos ANRM e HIER foram muito próximas para ambas as características (Tabela 3 e 4). O presente trabalho está de acordo com os resultados de CARDOSO & TEMPELMAN (2004), os quais concluíram que não houve diferenças significativas para a estimação da média *a posteriori* dos componentes de variâncias estimados pelos modelos HIER e ANRM.

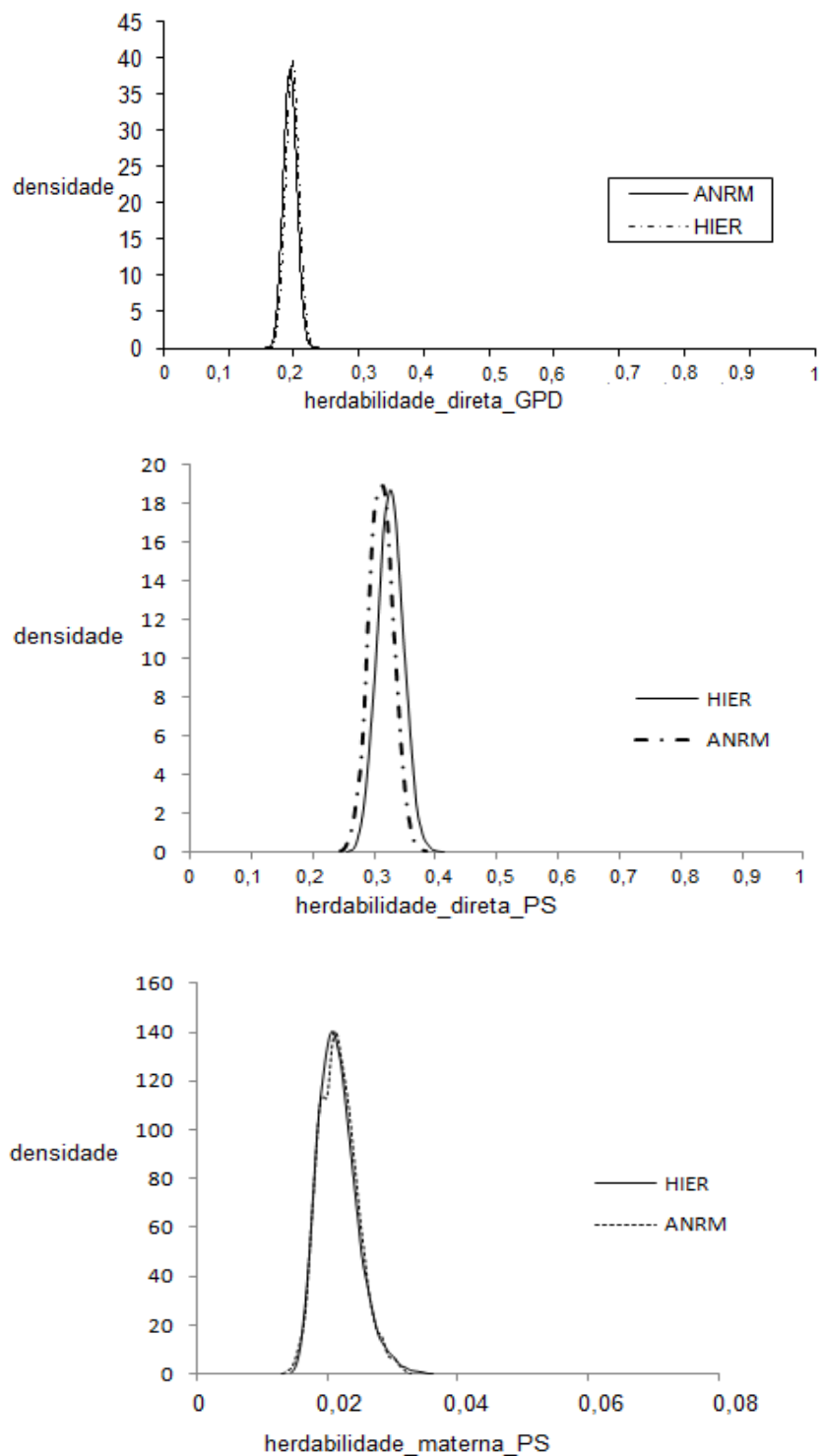
As maiores médias *a posteriori* da variância genética aditiva foram observadas quando se utilizou o modelo HIER para GPD e PS. Para esse modelo também se observou as menores estimativas da variância residual. Este resultado também foi encontrado por CARDOSO & TEMPELMAN (2004), os quais verificaram que as variâncias genéticas aditiva, para característica de herdabilidade mediana e alta, foram próximas comparando o ANRM e HIER, entretanto, verificaram uma tendência do modelo HIER de subestimar a variância residual (CARDOSO & TEMPELMAN, 2003).

Os desvios-padrão médios (DP) da variância genética aditiva foram maiores quando estimado pelo HIER (dados não apresentados). Esses resultados foram semelhantes aos descritos por CARDOSO & TEMPELMAN (2004), que descreveram os DP da variância genética aditiva, considerando somente animais da população base e animais com paternidade conhecida, semelhantes entre os modelos HIER e ANRM. No entanto, para indivíduos com incerteza de paternidade, o DP da variância genética aditiva foi maior quando estimado pelo HIER.

As estimativas da distribuição marginal *a posteriori* da herdabilidade para o GPD, obtidas pelo ANRM e HIER sobrepõem-se quase completamente (Figura 1), confirmando que praticamente não há diferenças entre os modelos ao estimar esse parâmetro. As estimativas médias *a posteriori* da herdabilidade (Tabela 4) são próximas às descritas na literatura para essa característica (ELER et al., 1996; MARCONDES, 1999; PANETO et al., 2002; MALHADO et al., 2005). De acordo com esses resultados, a seleção direta pelo GPD pode resultar em um pequeno ganho genético anual.

Em geral, as médias *a posteriori* dos componentes de variâncias estimados pelos modelos ANRM e HIER foram muito próximas para ambas as características (Tabela 3 e 4). O presente trabalho está de acordo com os resultados de CARDOSO & TEMPELMAN (2004), os quais concluíram que não houve diferenças significativas para a estimação da média *a posteriori* dos componentes de variâncias estimados pelos modelos HIER e ANRM.

As maiores médias *a posteriori* da variância genética aditiva foram observadas quando se utilizou o modelo HIER para GPD e PS. Para esse modelo também se observou as menores estimativas da variância residual. Este resultado também foi encontrado por CARDOSO & TEMPELMAN (2004), os quais verificaram que as variâncias genéticas aditiva, para característica de herdabilidade mediana e alta, foram próximas comparando o ANRM e HIER, entretanto, verificaram uma tendência do modelo HIER de subestimar a variância residual (CARDOSO & TEMPELMAN, 2003).



**Figura 1** - Densidades das estimativas de herdabilidade *a posteriori* de efeito direto do ganho de peso pós-desmama (GPD) e peso ao sobreano (PS) e herdabilidade *a posteriori* de efeito direto materno (abaixo) do peso ao sobreano

(PS) em animais Nelore obtidas pelo modelo baseado na matriz de parentesco médio (ANRM) e modelo hierárquico bayesiano (HIER)

Os desvios-padrão médios (DP) da variância genética aditiva foram maiores quando estimado pelo HIER (dados não apresentados). Esses resultados foram semelhantes aos descritos por CARDOSO & TEMPELMAN (2004), que descreveram os DP da variância genética aditiva, considerando somente animais da população base e animais com paternidade conhecida, semelhantes entre os modelos HIER e ANRM. No entanto, para indivíduos com incerteza de paternidade, o DP da variância genética aditiva foi maior quando estimado pelo HIER.

As estimativas da distribuição marginal *a posteriori* da herdabilidade para o GPD, obtidas pelo ANRM e HIER sobrepõem-se quase completamente (Figura 1), confirmando que praticamente não há diferenças entre os modelos ao estimar esse parâmetro. As estimativas médias *a posteriori* da herdabilidade (Tabela 4) são próximas às descritas na literatura para essa característica (ELER et al., 1996; MARCONDES, 1999; PANETO et al., 2002; MALHADO et al., 2005). De acordo com esses resultados, a seleção direta pelo GPD pode resultar em um pequeno ganho genético anual.

A média *a posteriori* da herdabilidade do peso ao sobreano estão próximas as encontradas na literatura (BITTENCOURT et al., 2002; BERTAZZO et al., 2004; BOLIGON et al., 2008). Estes resultados indicam que o peso ao sobreano responderá a seleção individual.

A herdabilidade média *a posteriori* para efeito genético materno obtido pelos modelos ANRM e HIER está próxima as descritas na literatura para animais Zebuínos (GUNSKY et al., 2001; RIBEIRO et al., 2001). Esses resultados indicam menor possibilidade de ganho pela seleção para essa característica do que o PS no rebanho Nelore do presente trabalho. As covariâncias e a correlações entre os efeitos genéticos direto e materno estimados com ambos os modelos foram positivas, porém de baixa magnitude e não significativas (Tabela 4).

As correlações de posição entre os valores genéticos preditos pelo ANRM e HIER para ambas as características foram altas para todos os conjuntos de animais, isto é, incluindo todos os animais do pedigree, apenas animais filhos de RM e somente



touros (Tabela 5), variando de 0,993 a 1,00. De modo geral, as correlações entre os valores genéticos medidos pelos diferentes modelos sugerem que estes não interferem na classificação dos animais quando se utilizam esses conjuntos de dados. Esses resultados estão de acordo com CARDOSO & TEMPELMAN (2003), que também encontraram altas correlações de posto entre parâmetros genéticos estimados pelo ANRM e HIER para característica de herdabilidade média e alta, sendo as maiores correlações encontradas quando a característica foi de herdabilidade mediana. Entretanto, CARDOSO & TEMPELMAN (2004) verificaram uma diminuição na correlação de posto quando selecionaram apenas 5 e 10% dos maiores valores genéticos.

**Tabela 5** – Estimativa da correlação de Spearman entre médias dos valores genéticos obtidos para diferentes conjuntos de animais pelo modelo baseado na matriz media de parentesco (ANRM) e o modelo hierárquico bayesiano (HIER) para o ganho de peso pós desmama (GPD) e o peso ao sobreano (PS)

| <b>Arquivo*</b>             | <i>Número de animais</i> | <b>Característica</b> |           |
|-----------------------------|--------------------------|-----------------------|-----------|
|                             |                          | <b>GPD</b>            | <b>PS</b> |
| <b>Com todos os animais</b> |                          |                       |           |
| <i>Completo</i>             | 75088                    | 0,999                 | 0,999     |
| <i>50%</i>                  | 37544                    | 0,999                 | 0,998     |
| <i>30%</i>                  | 22526                    | 0,998                 | 0,998     |
| <i>10%</i>                  | 7509                     | 0,997                 | 0,996     |
| <b>Filhos de RM</b>         |                          |                       |           |
| <i>Completo</i>             | 22810                    | 0,999                 | 0,999     |
| <i>50%</i>                  | 11169                    | 0,995                 | 0,998     |
| <i>30%</i>                  | 6012                     | 0,993                 | 0,997     |
| <i>10%</i>                  | 2281                     | 1,000                 | 0,990     |
| <b>Touros</b>               |                          |                       |           |
| <i>Completo</i>             | 519                      | 0,999                 | 0,999     |
| <i>50%</i>                  | 259                      | 0,994                 | 0,999     |
| <i>30%</i>                  | 155                      | 0,990                 | 0,997     |
| <i>10%</i>                  | 52                       | 0,972                 | 0,992     |

\*Os arquivos considerados continham todos os animais, os 50%, 30% e 10% superiores classificados pelos seus respectivos valores genéticos

## CONCLUSÕES

Pelos critérios de escolha do modelo verificou-se que o modelo que melhor ajustou os dados para estimar os parâmetros genéticos de animais que possuem incerteza de paternidade é o modelo hierárquico bayesiano. Entretanto, para avaliações genéticas, qualquer um dos dois modelos pode ser utilizado já que praticamente não houve diferenças entre os valores genéticos estimados.

## REFERÊNCIAS BIBLIOGRÁFICAS

BERTAZZO, R. P., FREITAS, R. T. F., GONÇALVES, T. M., PEREIRA, I. G., ELER, J. P., FERRAZ, J. B. S., OLIVEIRA, A. I. G., ANDRADE, I. F. 2004. Parâmetros Genéticos de Longevidade e Produtividade de Fêmeas da Raça Nelore. **Rev. Bras. Zootec.**, 33: 1118-1127.

BITTENCOURT, T. C. C., ROCHA, J. C. M. C., LOBO, R. B., BEZERRA, L. F. 2002. Estimação de componentes de (co)variâncias e predição de DEP's para características de crescimento pós-desmama de bovinos da raça Nelore, usando diferentes modelos estatísticos. **Arq. Bras. Med. Vet. Zootec.** 54(3): 303-308.

BOLIGON, A. A., ALBUQUERQUE, L. G., RORATO, P. R. N. 2008. Associações genéticas entre pesos e características reprodutivas em rebanhos da raça Nelore. **Rev. Bras. Zootec.**, 37(4): 596-601.

CARDOSO, F. F., TEMPELMAN, R. J., 2003. Bayesian inference on genetic merit under uncertain paternity. **Gen. Sel. Evol.**, 35, 469-487.

CARDOSO, F. F., R. J. TEMPELMAN. 2004. Genetic evaluation of beef cattle accounting for uncertain paternity. **Liv. Prod. Sci.**, v. 89, p. 109-119.

ELER, J. P., FERRAZ, J. B. S., SILVA, P. R. 1996. Parâmetros genéticos para peso, avaliação e circunferência escrotal na raça Nelore, estimados por modelo animal. **Arq. Bras. Med. Vet. Zootec.**, 48(2): 203-213.

GELFAND, A. E. 1996. Model determination using sampling-based methods. Page 145 in W. R, Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), **Markov chain Monte Carlo in practice**. London: Chapman & Hall.

GEYER, C.J. 1992. Practical Markov chain Monte Carlo. **Stat. Sci.** 7: 473- 511.

GUNSKY, R. J., GARNEIRO, A. D. V., BORJAS, A. R. L., BEZERRA, L. A. F., LÔBO, R. B. 2001. Estimativas de parâmetros genéticos para características incluídas em critérios de seleção em gado Nelore. **Ciência Rural**, Santa Maria, 31(4):603-607.

HENDERSON C. R. 1988. Use of an average numerator relationship matrix for multiple-sire joining, **J. Anim. Sci.**, 66:1614-1621.

MALHADO, C. H. M., MARTINS FILHO, R., LOBO, R. N. B., FACÓ, O., AZEVEDO, D. M. M. R., SOUZA, J. C., OLIVEIRA, S. M. P. 2005. Tendências genéticas para características relacionadas à velocidade de crescimento em bovinos Nelore na região Nordeste do Brasil. **Rev. Bras. Zootec.**, 34(1): 60-65.

MARCONDES, C. R. 1999. Análise de alguns critérios de seleção para características de crescimento na raça Nelore. Belo Horizonte: Universidade Federal de Minas Gerais. **Diss.Mestrado**. Universidade Federal de Minas Gerais, Brazil.

PANETO, J. C. C., LEMOS, D. C., BEZERRA, L. A. F., MARTINS FILHO, R., LÔBO, R. B. 2002. Estudo de características quantitativas de crescimento dos 120 aos 550 dias de idade em gado Nelore. **Rev. Bras. Zootec.**, 31(2): 668-674.

RIBEIRO, M. N., PIMENTA FILHO, E. C.; MARTINS, G. A. 2001. Herdabilidade para efeitos direto e materno de características de crescimento de bovinos Nelore no estado da Paraíba. **Rev. Bras. Zootec.**, 30(4):1224-1227.

ROSO, V. M., SCHENKEL, F. S. 2006. AMC – A computer program to assess the degree of connectedness among contemporary groups. Page 27 in **Proc.** 8th World Congress on Genetics Applied to Livestock Production. Belo Horizonte, MG, Brazil.

SAPP, R. L. 2005. Statistical approach for dealing with uncertain paternity, **PhD Diss.** Univ. Georgia, Athens.

SAPP, R. L., ZHANG, W., BERTRAND, J. K., REKAYA, R. 2007. Genetic evaluation in the presence of uncertain additive relationships. I. Use of phenotypic information to ascertain paternity. **J. Anim. Sci.**, 85:2391-2400.

SORENSEN, D. A., ANDERSEN, S., GIANOLA, D., et al., 1995. Bayesian-inference in threshold models using Gibbs sampling. **Gen. Select. Evol.** 27(3), 229-249.

SORENSEN, D., GIANOLA, D. 2002. **Likelihood, bayesian, and MCMC methods in quantitative genetics**. ed.Springer-Verlag, New York.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P., LINDE, A. 2002. Bayesian measures of model complexity and fit. **J. Royal Stat Soc. Series B-Stat. Method.**, 64: 583-616.

## **CAPÍTULO 4 – METODOLOGIA PARA ESTIMAÇÃO DO MÉRITO GENÉTICO DE ANIMAIS COM PATERNIDADE INCERTA SOB INFERÊNCIA BAYESIANA**

**Resumo:** Os objetivos do presente trabalho foram: (1) propor um procedimento bayesiano empírico iterativo para implementar o modelo animal hierárquico; (2) calcular as probabilidades da paternidade *a posteriori* via máxima verossimilhança; (3) comparar o modelo proposto com o modelo baseado na matriz de parentesco médio e ao modelo hierárquico implementado sob enfoque bayesiano pelo método Monte Carlo via Cadeias de Markov (MCMC) e predizer o mérito genético de animais com incerteza de paternidade. Após a consistência dos dados, foram mantidos 62.212 ganho de peso da desmama ao sobreano de animais da raça Nelore. O arquivo de pedigree continha 75.088 animais. Diferentes abordagens foram adotadas no tratamento de paternidade desconhecida para os animais filhos de reprodutores múltiplos. No primeiro, o modelo bayesiano empírico iterativo (ITER) foi usado para calcular as probabilidades da paternidade *a posteriori* via máxima verossimilhança. Na segunda abordagem, considerou-se o modelo hierárquico bayesiano (HIER) para inferir sobre a incerteza da paternidade. Na outra abordagem, com o modelo baseado na matriz de parentesco médio, atribuíram-se probabilidades iguais para cada um dos possíveis pais dos animais com paternidade incerta. Correlação de Spearman e Pearson entre a média *a posteriori* dos efeitos genéticos dos animais foram calculados para comparar o ordenamento dos modelos para seleção. O modelo ITER representa uma alternativa computacionalmente viável para calcular as probabilidades dos touros candidatos e os efeitos genéticos dos animais em avaliações genéticas de grandes bases de dados, quando existe incerteza nas atribuições de parentesco de alguns animais, sem a necessidade de usar MCMC.

**Palavras-chave:** Acasalamento com reprodutor-múltiplo, inferência bayesiana, paternidade desconhecida

**Abstract:** The aim of this work was: (1) to propose an empirical Bayesian procedure to implement the iterative hierarchical animal model; (2) to calculate posterior probabilities of paternity using maximum likelihood; (3) to compare the proposed model with a model based on the use of an average numerator relationship matrix (ANRM) and hierarchical animal model implemented under Bayesian approach via Markov Chain Monte Carlo (MCMC) methods to predict the genetic merit of animals with uncertain paternity. After the consistency of the data were kept records of 62,212 postweaning gain from Nelore animals. The pedigree file contained a total of 75,088 animals. Different approaches were adopted for the treatment of uncertain paternity. In the first, the Iterative empirical Bayes (ITER) was used to calculate the posterior probabilities of paternity based on maximum likelihood principles. The second approach was considered the Bayesian hierarchical model (HIER) that infers the uncertainty of paternity. The other approach, the model based on the use of an average numerator relationship matrix (ANRM), equal probabilities were attributed to each of the possible parents of the animals with uncertain paternity. Pearson and Spearman correlation between the average of the posterior breeding values were calculated to compare the ranking of models for selection. The ITER model represents a computationally feasible alternative to calculate the probabilities of candidate bulls and breeding values in genetic evaluation of large databases, when there is uncertainty in parentage assignments of some animals, without the need of using MCMC.

**Keyword:** Bayesian inference, multiple-sire mating, uncertain paternity

## 1. INTRODUÇÃO

Acasalamento com reprodutores múltiplos é um tipo de manejo reprodutivo que expõe um grupo de vacas a vários touros na mesma estação de monta. Esse tipo de acasalamento é utilizado para melhorar a taxa de concepção, no entanto, dificulta comparações dos touros quanto aos aspectos reprodutivos e produtivos. A falta da correta identificação de paternidade é uma das consequências para animais provenientes de acasalamento com reprodutores múltiplos, que pode afetar negativamente a acurácia das avaliações genéticas e, conseqüentemente, o progresso genético pela seleção.

Testes de DNA fornecem uma tecnologia acurada para identificação da paternidade de animais, quando amostras biológicas dos possíveis pais e da progênie estão disponíveis. Essa técnica é muito utilizada em grande parte dos países desenvolvidos, como Estados Unidos e alguns países da União Européia, principalmente em gado de leite. No entanto, para os países em desenvolvimento, essa técnica ainda não é acessível com relação aos custos, já que esses países possuem, em sua maioria, produção extensiva e rebanho comercial grande. Uma alternativa para inferir sobre a incerteza de paternidade é um modelo que combina informações *a priori* e de desempenho para calcular probabilidades *a posteriori* dos touros candidatos serem os pais, buscando melhorar a predição de mérito genético de animais provenientes de acasalamentos com reprodutores múltiplos (CARDOSO & TEMPELMAN, 2003). O método proposto por esses pesquisadores é superior ao uso da matriz de parentesco médio (HENDERSON, 1988) e ao uso de grupos genéticos (Westell et al., 1988), mas é baseado em Métodos Monte Carlo via Cadeias de Markov (MCMC), podendo demandar um tempo computacional longo demais para ser utilizado rotineiramente nas avaliações genéticas usando grandes bancos de dados.

Com esse trabalho, objetivou-se: (1) propor um procedimento bayesiano empírico iterativo para implementar o modelo animal hierárquico proposto por CARDOSO & TEMPELMAN (2003); (2) calcular as probabilidades da paternidade *a posteriori* via máxima verossimilhança, seguindo os conceitos propostos por FOULLEY et al. (1987); (3) comparar o modelo proposto com o modelo baseado na matriz de

parentesco médio (HENDERSON, 1988) e ao modelo hierárquico bayesiano (CARDOSO E TEMPELMAN, 2003) implementado sob enfoque bayesiano pelo método MCMC e para predizer o mérito genético de animais com incerteza de paternidade.

## **2. MATERIAL E MÉTODOS**

### **2.1 Animais**

Os registros de animais da raça Nelore, nascidos no período de 1984 a 2006, pertencentes ao arquivo zootécnico da Agropecuária Jacarezinho Ltda, localizada no município de Valparaíso, São Paulo foram utilizados.

A característica considerada neste estudo foi o ganho de peso pós-desmama (GPD) entre 205 e 550 dias de idade. A média e o desvio-padrão foram 98,40 e 33,33 kg, respectivamente.

Durante a consistência dos dados foram consideradas como perdidas as informações de grupos contemporâneos (GC) com menos de três animais e de animais cujas medidas foram superiores ou inferiores à média de seu GC, mais ou menos 3,5 desvios padrão. Todos os animais tinham pelo menos a mãe conhecida. Os GC foram formados pela combinação dos efeitos de sexo, fazenda de manejo à desmama e ao sobreano, ano e estação de nascimento, grupo de manejo à desmama e ao sobreano, mês da desmama e ao sobreano.

Após a consistência dos dados foram mantidos 62.212 animais, sendo que o arquivo de pedigree continha 75.088 animais, com 12.876 animais pertencentes à população base, 39.402 animais com paternidade conhecida e 22.810 animais eram provenientes de acasalamentos com reprodutores múltiplos (RM). Os grupos com reprodutores múltiplos foram formados com três, quatro, cinco, seis e dez touros.



## 2.2 Modelo hierárquico bayesiano com paternidade desconhecida

### 2.2.1 Metodologia

O seguinte modelo animal unicaracterística para um registro de GPD foi considerado:

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u} + e_i \quad [1]$$

em que  $y_i$  é o GPD observado para o animal  $i$ ,  $\boldsymbol{\beta}$  é o vetor de efeitos fixos (o efeito da idade do animal ao sobreano e idade da vaca linear e quadrático, 2 a 16 anos),  $\mathbf{u}$  é o vetor de efeitos genéticos diretos,  $e_i$  é o resíduo associado com o registro do animal  $i$  e, finalmente,  $\mathbf{x}'_i$  e  $\mathbf{z}'_i$  são vetores linha de incidência, relacionando, respectivamente, os efeitos em  $\boldsymbol{\beta}$  e  $\mathbf{u}$  com o registro do animal  $i$ .

Assumiu-se normalidade ( $N$ ) e que as variâncias genética aditiva ( $\sigma_u^2$ ) e residual ( $\sigma_e^2$ ) foram conhecidas, a distribuição condicional dos registros  $\mathbf{y}$  ( $y_i = 1, 2, \dots, n$ ) dado os parâmetros  $\boldsymbol{\beta}$ ,  $\mathbf{u}$ , e  $\sigma_e^2$ , pode ser escrita como:

$$\mathbf{y} | \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2 \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R} = \mathbf{I}\sigma_e^2) \quad [2]$$

em que  $\mathbf{X} = \{\mathbf{x}'_i\}_{i=1}^n$  e  $\mathbf{Z} = \{\mathbf{z}'_i\}_{i=1}^n$  são matrizes de incidência e  $\mathbf{I}$  uma matriz identidade. Diferentemente do modelo touro de FOULLEY et al., (1987), a especificação do primeiro estágio do modelo hierárquico na equação [2] acima, por ser baseada em um modelo animal, não depende do conhecimento do pai verdadeiro do animal  $i$ , mesmo quando alguns animais tem incerteza nas atribuições de paternidade. Todos os elementos de  $\mathbf{Z}$  são conhecidos, pois relacionam os efeitos dos animais em  $\mathbf{u}$  (e não de seus pais) com os seus registros.

Desta forma a verossimilhança dos dados de acordo com a Equação [2] pode ser escrito como:

$$p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2) = \frac{1}{(2\pi\sigma_e^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_e^2} \sum_{i=1}^n (y_i - \mathbf{x}'_i\boldsymbol{\beta} - \mathbf{z}'_i\mathbf{u})^2\right) \quad [3]$$

O segundo estágio do modelo é definido pelas distribuições *a priori* dos parâmetros de locação. Para os efeitos fixos, a ausência de conhecimento prévio é representada por:

$$p(\boldsymbol{\beta}) \sim \text{constante} \quad [4]$$

A especificação *a priori* de  $\mathbf{u}$  para o modelo animal depende do conhecimento da matriz do numerador de parentesco ( $\mathbf{A}$ ). Como no conjunto de dados analisado, determinados indivíduos têm incerteza de paternidade, o pai verdadeiro da progênie  $i$  é desconhecido e essa situação é representada pela variável aleatória  $s_{i^*}$  (CARDOSO & TEMPELMAN, 2003). Assumindo-se que  $s_{i^*} = s_{ij}$  é a situação na qual o touro  $j$  é o pai verdadeiro do animal  $i$ , temos a seguinte especificação:

$$\mathbf{u} | \mathbf{s}^* = \mathbf{s}_j, \sigma_u^2 \sim N(\mathbf{0}, \mathbf{A}_s \sigma_u^2) \quad [5]$$

em que  $\mathbf{s}^* = \mathbf{s}_j$  representa o conjunto de todas as atribuições de paternidade e  $\mathbf{A}_s$  a matriz do numerador de parentesco condicional nessas atribuições.

A densidade dessa priori é representada por:

$$p(\mathbf{u} | \mathbf{s}^* = \mathbf{s}_j, \sigma_u^2) = (2\pi\sigma_u^2)^{-\frac{p}{2}} \prod_{i=1}^p d_{ij}^{-1} \exp\left(-\frac{\left(u_i - 0,5(u_{s_{ij}} + u_{d_i})\right)^2}{2d_{ij}\sigma_u^2}\right), \quad [6]$$

em que,  $p$  é número total de animais e  $d_{ij}$  é fração da variância genética devido a segregação mendeliana, calculada por  $u_i - 0,5(u_{s_{ij}} + u_{d_i})$ , assumindo-se que o touro  $j$  é o pai verdadeiro do animal  $i$ . Aqui  $u_{s_{ij}}$  e  $u_{d_i}$  são respectivamente os valores genéticos

do pai e da mãe do animal  $i$  e são iguais a zero para animais com pais desconhecidos. Além disso, a inversa  $d_{ij}^{-1}$  corresponde ao  $i$ -ésimo elemento da matriz  $\mathbf{D}^{-1}$  em  $\mathbf{A}^{-1} = \mathbf{T}\mathbf{D}^{-1}\mathbf{T}$ , a decomposição usada para computar  $\mathbf{A}^{-1}$  (QUAAS, 1976; HENDERSON, 1976). O valor de  $d_{ij}^{-1}$  pode ser calculado pela inversa de  $d_{ij} = 1 - 0,25a_{s_{ij}s_{ij}} - 0,25a_{d_id_i}$ , sendo  $a_{kk}$  um elemento da diagonal  $\mathbf{A}$  correspondente aos pais de  $i$  e definido como zero para animais pertencentes a geração base.

Finalmente, no terceiro estágio são especificadas distribuições multinomiais para as probabilidades *a priori* das atribuições de paternidade:

$$p(s_{i^*} = s_{ij}) = \pi_{ij}, \quad i=1,2,\dots,p, \quad j=1,2,\dots,v_i. \quad [7]$$

Se  $\pi = \{\pi_{ij}\}_{i=1,2,\dots,p, j=1,2,\dots,v_i}$  é o conjunto de todas essas probabilidades incluindo  $v_i$  pais para cada um dos  $p$  animais, podemos marginalizar *a priori* na Equação [6] com respeito às possíveis atribuições de parentes usando as probabilidades *a priori* definidas na Equação [7], obtendo o seguinte resultado:

$$p(\mathbf{u} | \sigma_u^2) \propto \prod_{i=1}^p \prod_{j=1}^{v_i} \pi_{ij} d_{ij}^{-1} \exp \left( - \frac{(u_i - 0,5(u_{s_{ij}} + u_{d_i}))^2}{2d_{ij}\sigma_u^2} \right). \quad [8]$$

Na Equação [8] em vez de condicionar ao conhecimento da atribuição de paternidade, a contribuição de todos os touros candidatos é ponderada por suas probabilidades *a priori*.

A distribuição *a posteriori* de  $\boldsymbol{\beta}$  e  $\mathbf{u}$ , assumindo que os parâmetros de dispersão são conhecidos e que [2], [4], [6] e [7] são condicionalmente independentes e ainda usando a marginalização em [8] é dada por:

$$\begin{aligned}
p(\boldsymbol{\beta}, \mathbf{u} \mid \mathbf{y}, \sigma_u^2, \sigma_e^2) &= p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2) p(\boldsymbol{\beta}) p(\mathbf{u} \mid \mathbf{s}_* = \mathbf{s}_j, \sigma_u^2) p(\mathbf{s}_* = \mathbf{s}_j) \\
&\propto \exp\left(-\frac{1}{2\sigma_e^2} \sum_{i=1}^n (y_i - \mathbf{x}'_i \boldsymbol{\beta} - \mathbf{z}'_i \mathbf{u})^2\right) \prod_{i=1}^p \prod_{j=1}^{v_i} \pi_{ij} d_{ij}^{-1} \exp\left(-\frac{(u_i - 0,5(u_{s_{ij}} + u_{d_i}))^2}{2d_{ij}\sigma_u^2}\right) [9]
\end{aligned}$$

### 2.2.2 Consideração das atribuições de paternidade

Se  $\pi$  fosse tratado como conhecido em [9], teríamos então resultados equivalentes à especificação da matriz do parentesco média de HENDERSON (1988).

Entretanto, expandindo a estratégia de FOULLEY et al., (1987) para maximizar [9] e obter estimativas modais de  $\boldsymbol{\beta}$  e  $\mathbf{u}$  no caso de modelo animal, em vez de fixarmos as probabilidades de paternidade nos seus valores *a priori* em  $\pi$  atualizamos esses valores utilizando as informações dos dados, conforme descrito por CARDOSO & TEMPELMAN (2003), gerando as seguintes probabilidades de paternidade *a posteriori*:

$$q_{ij} = \frac{\pi_{ij} d_{ij}^{-1} \exp\left(-\left(u_i - 0,5(u_{s_{ij}} + u_{d_i})\right)^2 / 2d_{ij}\sigma_u^2\right)}{\sum_{j=1}^{v_i} \pi_{ij} d_{ij}^{-1} \exp\left(-\left(u_i - 0,5(u_{s_{ij}} + u_{d_i})\right)^2 / 2d_{ij}\sigma_u^2\right)}. \quad [10]$$

Essas probabilidades são usadas para definir um sistema de equações de modelos mistos no qual a inversa da matriz de parentesco médio  $\bar{\mathbf{A}}^{-1}$  é baseada  $q_{ij}$ 's (probabilidades posteriores) e não em prioris fixas ( $\pi_{ij}$ ) como originalmente proposto por HENDERSON (1988). Entretanto, como as probabilidades posteriores dependem dos valores de  $\mathbf{u}$ , mesmo quando as variâncias são conhecidas, o sistema resultante é não-linear e precisa ser solucionado por iterações, como apresentado a seguir:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \bar{\mathbf{A}}^{-1[t-1]}\sigma_u^{-2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}^{[t]} \\ \hat{\mathbf{u}}^{[t]} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad [11]$$

Em que  $[t]$  representa a iteração atual e  $[t-1]$  em  $\bar{\mathbf{A}}^{-1[t-1]}$  significa que as probabilidades *a posteriori* usadas no computo foram obtidas a partir dos valores genéticos aditivos do ciclo anterior  $\hat{\mathbf{u}}^{[t-1]}$ .

### 2.2.3 Implementação do processo iterativo

Os valores de  $\bar{d}_i = 1 - 0,25 \sum_{j=1}^{v_i} \sum_{j'=1}^{v_i} q_{ij} q_{ij'} a_{s_{ij} s_{ij'}} - 0,25 a_{dd}$  necessários para compor

$\bar{\mathbf{A}}^{-1}$  (HENDERSON, 1988; PEREZ-ENCISO & FERNANDO, 1992) foram obtidos diretamente sem a necessidade de calcular primeiro  $\bar{\mathbf{A}}$  pelo algoritmo eficiente desenvolvido por FAMULA (1992).

O processo iterativo utilizado foi o seguinte:

a. Para iniciar o processo, antes de termos uma solução para  $\mathbf{u}$ , inicializamos os valores das probabilidades de paternidade por meio de suas priores, isto é em  $t = 0$   $q_{ij} = \pi_{ij}$  para todo  $i$  e  $j$ ;

b. Computa-se  $\bar{\mathbf{A}}^{-1[t-1]}$  usando se os valores de  $\hat{q}_{ij}^{[t-1]}$  usando o algoritmo de FAMULA (1992) e obtêm as soluções  $\hat{\boldsymbol{\beta}}^{[t]}$  e  $\hat{\mathbf{u}}^{[t]}$  a partir das Equações em [11];

c. Computa-se  $\hat{q}_{ij}^{[t]}$  por meio das Equações em [10] usando  $\hat{\mathbf{u}}^{[t]}$ ;

d. Confere-se convergência verificando se  $(\hat{\mathbf{u}}^{[t]} - \hat{\mathbf{u}}^{[t-1]})' (\hat{\mathbf{u}}^{[t]} - \hat{\mathbf{u}}^{[t-1]}) < 1E - 10$ .

Caso verdadeiro grava-se  $\hat{q}_{ij}^{[t]}$ ,  $\hat{\boldsymbol{\beta}}^{[t]}$  e  $\hat{\mathbf{u}}^{[t]}$  e finaliza o algoritmo. Se por outro lado o critério não foi atingindo então retorna-se ao passo 2.

### 2.3. Análises dos dados

Três abordagens foram adotadas no tratamento de paternidade desconhecida para os animais filhos de RM:

- ITER: modelo bayesiano empírico iterativo para calcular as probabilidades da paternidade *a posteriori* via máxima verossimilhança, seguindo os conceitos propostos por FOULLEY et al. (1987), como proposto na sessão 2.2;

- HIER: modelo hierárquico bayesiano que infere sobre a incerteza da paternidade proposto por CARDOSO & TEMPELMAN (2003);

- ANRM: modelo baseado na matriz de parentesco médio proposto por Henderson (1988), em que probabilidades iguais são atribuídas para cada um dos possíveis pais dos animais com paternidade incerta.

O modelo ITER foi comparado ao modelo ANRM e ao HIER para a estimação do efeito genético do animal com paternidade desconhecida.

Foram calculadas correlações de Spearman e Pearson entre a média *a posteriori* dos efeitos genéticos dos animais, a fim de se comparar o ordenamento dos animais para seleção. As correlações foram obtidos pelos modelos ANRM e HIER, usando o software Intergen (CARDOSO, 2008), e para a moda estimada pelo modelo ITER, utilizou-se um software desenvolvido para esse propósito, denominado Intersires. Essas correlações foram obtidas para todos os animais do pedigree e para os animais classificados como os 10, 30 e 50% melhores pelo seu valor genético. Foi verificado o tempo computacional aproximado que os procedimentos estatísticos despenderam.

### 3. RESULTADOS E DISCUSSÃO

Na Tabela 1, observam-se que as correlações de Spearman e Pearson entre o ITER com os demais modelos (ANRM, HIER) foram semelhantes para todos os arquivos. Quando todos os animais foram considerados, altas correlações entre o efeito genético animal foram observadas. Entretanto, quando somente animais no extremo superior (os animais 10% superiores pelo valor genético) foram considerados, as correlações decresceram para valores ao redor de 80%, que indicam reordenações de moderada importância biológica. Esses resultados sugerem que as estimativas dos valores genéticos dos animais são relativamente semelhantes pelos modelos avaliados, mas a escolha para utilização (entre o ITER e ANRM; ou ITER e HIER) implicará em algumas alterações na lista de animais a serem selecionados.

As médias e desvios-padrão dos valores genéticos estimados pelo modelo proposto ( $1,18 \pm 5,74$  kg) foram levemente superiores em relação às estimadas com o modelo HIER ( $0,99 \pm 3,81$ ). Além disso, o intervalo de credibilidade de 95%, para os

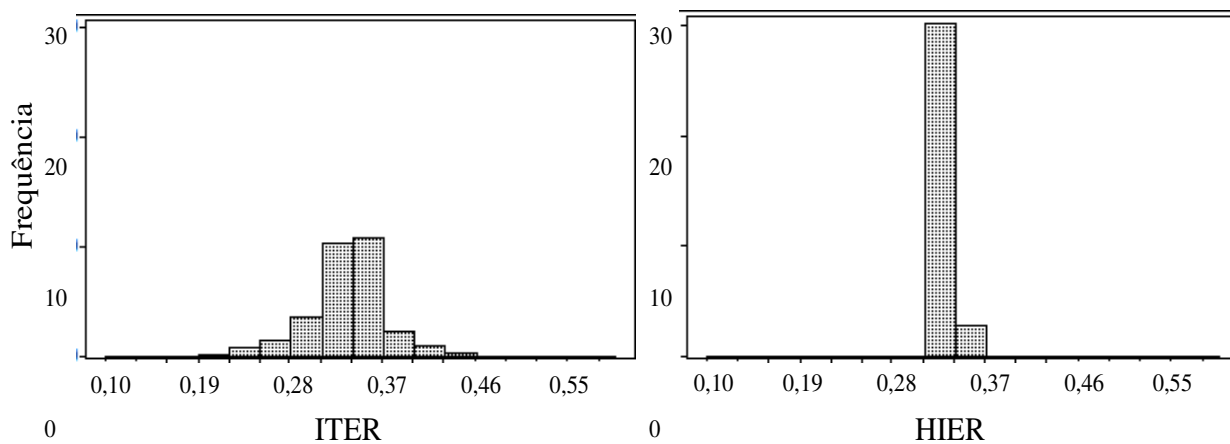
valores genéticos apresentou uma amplitude um pouco maior quando utilizado o modelo ITER (-10,27 à 12,72) comparado ao HIER (-6,45 à 8,87). Este resultado não seria necessariamente esperado, uma vez que o modelo ITER assume como conhecidos os componentes de variância que, no presente trabalho, foram assumidos como sendo as médias *a posteriori* obtidas por HIER e, portanto, dispersões similares seriam mais plausíveis.

**Tabela 1-** Correlação de Spearman e Pearson entre os efeitos genéticos do animal obtidos entre o modelo proposto e o modelo hierárquico bayesiano

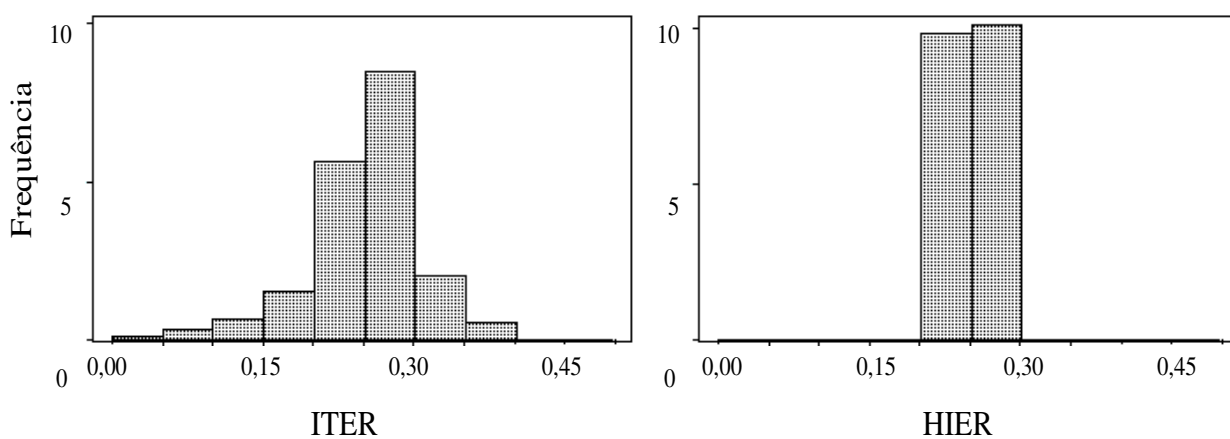
| Correlação entre | Correlação   |                 |                |
|------------------|--------------|-----------------|----------------|
|                  | Arquivo      | <i>Spearman</i> | <i>Pearson</i> |
| ITER vs HIER     | Completo     | 0,973           | 0,970          |
|                  | 50%          | 0,905           | 0,922          |
|                  | 30%          | 0,804           | 0,886          |
|                  | 10%          | 0,768           | 0,821          |
|                  | ITER vs ANRM | Completo        | 0,972          |
|                  | 50%          | 0,901           | 0,918          |
|                  | 30%          | 0,842           | 0,881          |
|                  | 10%          | 0,763           | 0,817          |

Verificou-se que as probabilidades dos pais candidatos, estimadas pelo modelo proposto no trabalho, tiveram amplitudes maiores comparada ao modelo HIER (Figura 1). CARDOSO & TEMPELMAN (2003) compararam o modelo HIER ao modelo com base na matriz de parentesco médio (ANRM) proposto por HENDERSON (1988), e verificaram que as probabilidades dos candidatos *a posteriori* foram um pouco maiores calculadas pelo HIER, indicando tendência de inferir o touro corretamente.

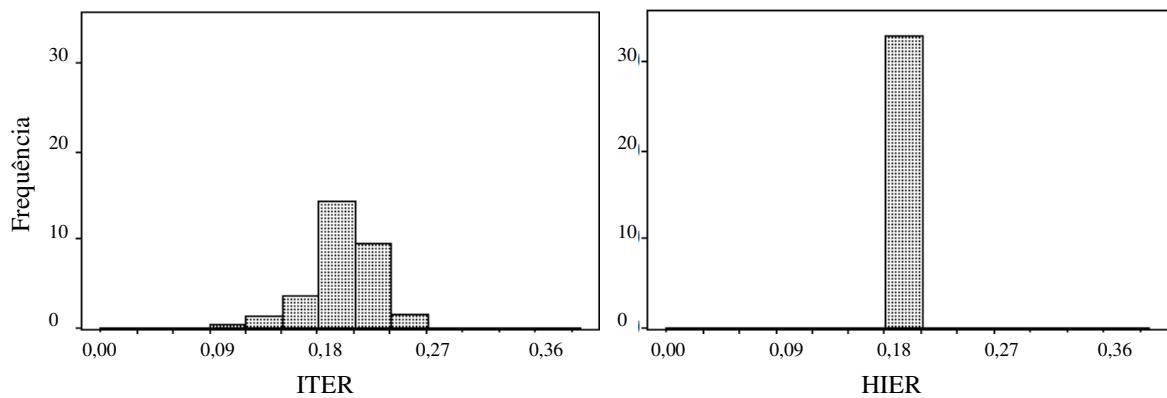
Tamanho do grupo de RM = 3



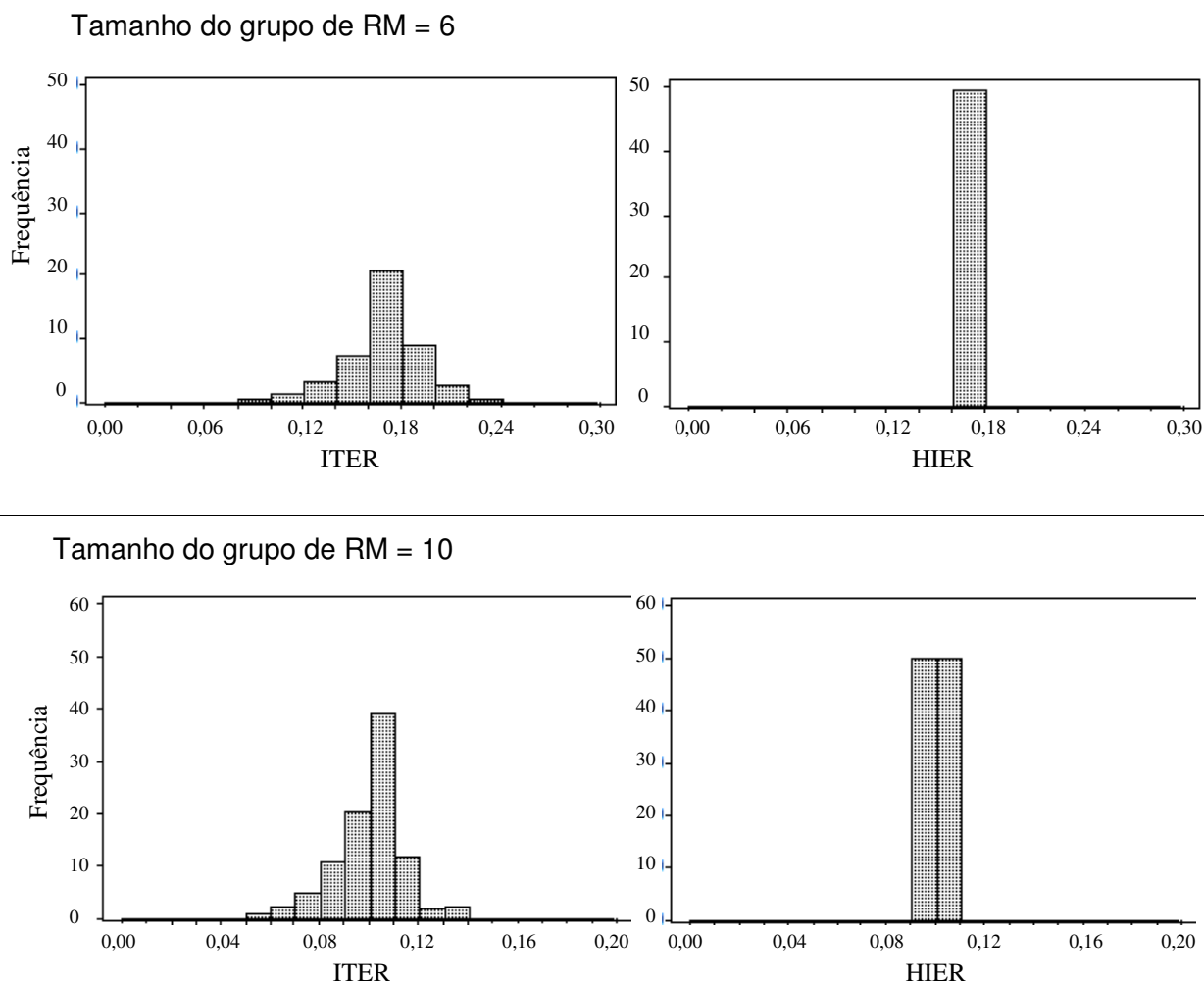
Tamanho do grupo de RM = 4



Tamanho do grupo de RM = 5







**Figura 1-** Distribuição das probabilidades *a posteriori* dos candidatos calculadas pelo modelo proposto no trabalho (Intersires) e o modelo hierárquico bayesiano (HIER)

A probabilidade *a posteriori* média dos pais candidatos foram as mesmas atribuídas *a priori* (considerando a razão de 1/número de candidatos, tamanho respectivo ao grupo de reprodutores múltiplos) para o modelo proposto no presente trabalho. Na Tabela 2, podem ser observados os valores de moda da probabilidade *a posteriori* estimada pelo modelo HIER e o Intersires. O valor da moda para a probabilidade *a posteriori* foi entre 1 a 4% maior que as respectivas probabilidades *a priori* atribuídas aos pais candidatos. CARDOSO & TEMPELMAN (2003) trabalhando com dados simulados e conhecendo as probabilidades do pai verdadeiro, verificaram que o modelo HIER atribuiu probabilidades *a posteriori* para os candidatos de 1 a 10%

maiores comparadas às probabilidades *a priori* (definida como no presente trabalho, 1 pela quantidade de pais candidatos) para característica de média herdabilidade e probabilidades *a posteriori* de 4 a 13% maiores para característica com herdabilidade alta.

CARDOSO & TEMPELMAN (2003) estimaram probabilidades *a posteriori* de pais candidatos quando animais com incerteza de paternidade estavam presentes nos dados, e assim como no presente trabalho, definiram as probabilidades *a priori* de cada touro candidato por um sobre o número de reprodutores múltiplos do grupo de reprodutor múltiplo ( $1/n$ ), e encontraram probabilidades *a posteriori* média levemente maiores que as descritas no presente trabalho (0,352 e 0,28 para grupo de RM com 3 e 4 touros, respectivamente). SAPP et al. (2007), também descreveram probabilidades dos touros candidatos um pouco maiores comparadas às apresentadas no presente trabalho (0,359 e 0,267 para grupo de RM com 3 e 4 touros), trabalhando com dados simulados para uma análise unicaracterística considerando herdabilidade alta (0,50).

As correlações entre a moda das probabilidades *a posteriori* obtidas pelos dois modelos também foram apresentadas na Tabela 2. Verificou-se que as maiores probabilidades foram encontradas nos grupos de reprodutores múltiplos com menor número de candidatos. As correlações positivas e moderadamente altas indicam que as probabilidades *a posteriori* calculadas pelo modelo HIER são semelhantes às probabilidades calculadas pelo modelo ITER proposto.

Maiores rebanhos proporcionam um maior potencial para seleção, conseqüentemente, causam maior impacto no progresso genético. A utilização exclusiva de monta controlada e/ou inseminação artificial representa alto custo aos produtores, possivelmente impraticável, devido à mão-de-obra e distribuição de áreas requerida. Muitos programas de avaliações genéticas não consideram animais provenientes de acasalamento com reprodutores múltiplos, em que não se sabe corretamente qual o verdadeiro pai do animal, ou seja, animais com incerteza da paternidade. No entanto, a exclusão de informações de animais com paternidade incerta causa redução no progresso genético (SULLIVAN, 1995), além da impossibilidade de se comparar o desempenho reprodutivo e produtivo dos diferentes touros.

**Tabela 2-** Amplitude do valor modal de probabilidade *a posteriori* do touro candidato a ser o pai verdadeiro pelo modelo hierárquico bayesiano (HIER) e pelo modelo bayesiano empírico iterativo (ITER) para diferentes tamanhos de grupos de reprodutores múltiplos (RM) e suas correlações

| Tamanho do grupo<br>de RM | Programa    |             | Correlação |
|---------------------------|-------------|-------------|------------|
|                           | <i>HIER</i> | <i>ITER</i> |            |
| 3                         | 0,30 à 0,37 | 0,10 à 0,63 | 0,84       |
| 4                         | 0,22 à 0,28 | 0,01 à 0,65 | 0,78       |
| 5                         | 0,18 à 0,23 | 0,02 à 0,61 | 0,75       |
| 6                         | 0,15 à 0,19 | 0,00 à 0,66 | 0,76       |
| 10                        | 0,09 à 0,10 | 0,04 à 0,17 | 0,67       |

O modelo HIER foi recomendado por trabalhos na literatura para predizer o valor genético de animais com incerteza de paternidade comparado ao modelo ANRM (CARDOSO & TEMPELMAN, 2003; 2004). O HIER utiliza informações *a priori* e fenotípicas para calcular probabilidades *a posteriori*, utilizando o método de Monte Carlo via cadeias de Markov. Os métodos HIER e ANRM despenderam quase o mesmo tempo computacional, no entanto, os critérios de escolha do modelo confirmaram que o HIER pode melhor predizer o mérito genético de animais com paternidade desconhecida.

O modelo proposto no presente trabalho via soluções iterativas (ITER) pode resultar em estimativas de efeito genético animal e probabilidades dos pais candidatos próximas às encontradas com o modelo hierárquico bayesiano. Uma vantagem do ITER é o tempo computacional despendido em cada análise que foi menor que com o HIER. Portanto, para avaliações genéticas utilizando grandes bancos de dados, o modelo proposto pode ser recomendado como uma alternativa computacionalmente viável.

Por outro lado, as diferenças nas estimativas de valores genéticos dos animais e de probabilidades dos touros candidatos, embora moderadas, apontam para a necessidade de estudos adicionais comparando esses modelos utilizando dados simulados. Com o conhecimento da situação real de paternidade e valores genéticos,

será possível determinar a precisão relativa dos modelos HIER e ITER na estimação dos valores genéticos de animais com incerteza de paternidade e das atribuições de probabilidade aos pais verdadeiros.

O uso conjunto de informações *a priori* e dos dados fenotípicos para as atribuições de paternidade em programas de melhoramento genético de larga escala, pode melhorar a predição dos valores genéticos dos animais em situações de incerteza de paternidade devido ao uso de acasalamento com reprodutores múltiplos.

## **CONCLUSÕES**

O modelo proposto no presente trabalho representa uma alternativa computacionalmente viável para calcular as probabilidades dos touros candidatos e os valores genéticos dos animais em avaliações genéticas de grandes bases de dados, quando existe incerteza nas atribuições de parentesco de alguns animais, sem a necessidade de usar Métodos de Monte Carlo via Cadeias de Markov. Embora as diferenças sejam moderadas, estudos adicionais são necessários para comparar esses modelos utilizando dados simulados.

## REFERÊNCIAS BIBLIOGRÁFICAS

CARDOSO, F. F.; R. J. TEMPELMAN. Bayesian inference on genetic merit under uncertain paternity. **Gen. Sel. Evol.** v.35, p. 469-487. 2003.

CARDOSO, F. F.; R. J. TEMPELMAN. Genetic evaluation of beef cattle accounting for uncertain paternity. **Liv. Prod. Sci.**, v. 89, p. 109-119. 2004.

FAMULA, T.R. Simple and rapid inversion of additive relationship matrices incorporating parental uncertainty. **J. Anim. Sci.**, v. 70, p. 1045-1048. 1992.

FOULLEY, J.L.; GIANOLA, D.; PLANCHENAULT, D. Sire evaluation with uncertain paternity. **Gen. Sel. Evol.** 19: 83-102. 1987.

HENDERSON C.R. A simple way for calculating 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, **J. Anim. Sci.**, v. 66, p. 1614-1621. 1988.

PEREZ-ENCISO M.; FERNANDO R.L. Genetic evaluation with uncertain parentage- a comparison of methods, **Theor. Appl. Genet.**, v. 84 p. 173-179. 1992.

QUAAS, R. L. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* v. 32, p. 949-953. 1976.

QUAAS, R. L. Additive genetic model with groups and relationships. **J. Dairy Sci.**, v. 71, p. 1338-1345. 1988.

SAPP, R.L.; ZHANG, W.; BERTRAND, J.K. et al. Genetic evaluation in the presence of uncertain additive relationships. I. Use of phenotypic information to ascertain paternity. **J. Anim. Sci.**, v. 85, p. 2391-2400. 2007.

SULLIVAN P.G. Alternatives for genetic evaluation with uncertain parentage, **Can. J. Anim. Sci.** v. 75, p. 31-36. 1995.

## CAPÍTULO 5 – IMPLICAÇÕES

O sistema de acasalamentos empregando reprodutores múltiplos é amplamente utilizado na pecuária de corte extensiva no Brasil, o mesmo é caracterizado pela exposição de um grupo de vacas a diversos touros durante a mesma estação de monta. Este tipo de sistema causa a formação de pedigrees incompletos, os quais podem afetar diretamente a acurácia das avaliações genéticas e a intensidade de seleção.

No presente trabalho, diversos procedimentos estatísticos, que levam em conta a paternidade incerta, foram comparados para predizer o mérito genético dos animais com paternidade desconhecida. Com exceção do modelo que incluiu grupo genético, formado com animais pertencentes ao grupo de reprodutores múltiplos, todos os demais procedimentos avaliados que levaram em conta a paternidade desconhecida foram mais adequados. Entretanto, poucos programas de melhoramento empregam qualquer procedimento para corrigir para incerteza da paternidade no Brasil.

Embora o trabalho aponte que a inclusão de grupos genéticos no modelo seja apropriada para predição do mérito genético de animais, quando pedigrees incompletos estão presentes, verificou-se que o grupo genético, definido como grupos de reprodutores múltiplos, não foi tão eficiente quanto ao grupo genético formado de acordo com a geração. Assim, a melhor definição do grupo genético pode variar, provavelmente, dependendo da estrutura dos dados e da característica. Desta forma, uma definição única de grupos genéticos não existe.

Além disso, o modelo hierárquico necessita de Métodos de Monte Carlo via Cadeias de Markov para calcular as probabilidades dos touros candidatos e os valores genéticos dos animais. Isso pode representar um aumento significativo no tempo computacional para análise de grandes bases de dados. No presente trabalho foi proposto e implementado o modelo bayesiano empírico iterativo, o qual calcula as probabilidades da paternidade *a posteriori* via máxima verossimilhança. Com este modelo é possível diminuir o tempo de processamento em relação ao modelo hierárquico, sendo uma alternativa computacionalmente viável para calcular as probabilidades dos touros candidatos e os valores genéticos dos animais em

avaliações genéticas quando existe incerteza nas atribuições de parentesco de alguns animais e grandes bancos de dados são avaliados.

Para a eficiente utilização de modelos que consideram a incerteza de paternidade e que esses modelos forneçam maior precisão, é aconselhável que arquivos zootécnicos incluam o maior número possível de identificação dos touros participantes de cada grupo de reprodutores múltiplos.