

FERNANDA CRISTINA BREDÁ

**MODELOS DE DIMENSÃO INFINITA PARA ESTIMAÇÃO DE PARÂMETROS  
GENÉTICOS DA PRODUÇÃO DE LEITE DE BÚFALOS DA RAÇA MURRAH**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

VIÇOSA  
MINAS GERAIS - BRASIL  
2008

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

B831m  
2008

Breda, Fernanda Cristina, 1978-

Modelos de dimensão infinita para estimação de parâmetros genéticos da produção de leite de búfalos da raça Murrah / Fernanda Cristina Breda. – Viçosa, MG, 2008.

viii, 60f. : il. ; 29cm.

Orientador: Ricardo Frederico Euclides.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Búfalos - Melhoramento genético. 2. Leite - Produção. 3. Análise de regressão. 4. Teoria bayesiana.  
I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 636.293

FERNANDA CRISTINA BREDÁ

**MODELOS DE DIMENSÃO INFINITA PARA ESTIMAÇÃO DE PARÂMETROS  
GENÉTICOS DA PRODUÇÃO DE LEITE DE BÚFALOS DA RAÇA MURRAH**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 18 de janeiro de 2008.

---

Prof. Paulo Sávio Lopes

---

Prof. Robledo de Almeida Torres

---

Prof<sup>a</sup>. Lúcia Galvão de Albuquerque  
(Co-orientador)

---

Carmen Silva Pereira  
(Membro)

---

Prof. Ricardo Frederico Euclides  
(Orientador)

*Aos meus pais, Roberto Luiz Breda e Marli Breda por terem me dado a base para que a realização de todos os meus sonhos pudesse se tornar realidade.*

*Aos meus irmãos Carlos Mario Breda Neto e Otilia Batista.*

*Ao meu querido e amado Renius Mello por me fazer acreditar, todos os dias, que é muito bom viver.*

*Ao meu filho Heitor por ter tornado a vida ainda mais bela e doce.*

*Dedico e Ofereço*

## AGRADECIMENTOS

A Deus

À Universidade Federal de Viçosa, Universidade Estadual Paulista (UNESP de Jaboticabal) e a Universidade Federal de Roraima.

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

Ao meu orientador e amigo Ricardo Frederico Euclides, pela oportunidade concedida, confiança e amizade.

Aos meus co-orientadores Lucia Galvão de Albuquerque e Humberto Tonhati, pelos ensinamentos e por terem me recebido tão bem na UNESP de Jaboticabal.

Aos professores Paulo Sávio Lopes, Robledo de Almeida Torres, Carmen Silva Pereira, pelos ensinamentos, apoio e amizade.

Aos amigos José Lindenberg Sarmiento, Marcelo Freitas, Leandro Barbosa, Herluce Cavalcante, Vicente Gontijo, Dilcéia, Caroline Pelegrina Teixeira, Rosana Colatino Soares Reis e Annaíza Braga Bignardi, pelo companheirismo nos momentos bons e ruins.

Aos colegas do curso de pós-graduação: Antônio Policarpo, Cláudio, Giselle, Jaime, Guilherme, Gustavo, Herluce, José Lindenberg, José Marques, Leandro, Marcelo, Paulo Carneiro, Rachel, Ricardo, Rodolpho, Urbano, Vicente, Kecya, Fred, Alex, Luciara, Joãozinho, Annaíza, Marcos, Mônia, Severino, Denise, Márcio, Arione, Fábio, Luciana, Roberta e Leonardo.

Aos professores, funcionários e alunos do Departamento de Zootecnia e dos demais Departamentos da Universidade, pelo agradável convívio.

Enfim, a todos que, diretamente ou indiretamente, contribuíram para que este trabalho fosse concretizado.

## BIOGRAFIA

**FERNANDA CRISTINA BREDA**, casada, nascida em 30 de agosto de 1978, na cidade de São Paulo-SP, filha de Roberto Luiz Breda e Marli Breda. Iniciou em janeiro de 1996 o curso de Zootecnia na Universidade Marília - SP, graduando-se em dezembro de 2000. Em Agosto 2001, ingressou no Programa de pós graduação em Zootecnia, na área de Melhoramento Genético Animal, na Universidade Federal de Viçosa, Minas Gerais, obtendo o grau de mestre em 29 de julho de 2003, sob a orientação do Prof<sup>o</sup>. Dro. Ricardo Frederico Euclides. Em fevereiro de 2004, iniciou o curso de Doutorado em Genética e Melhoramento na Universidade Federal de Viçosa, Minas Gerais. Em 09 de dezembro de 2006 foi aprovada no concurso publico para professor magistério do 3<sup>o</sup> grau na Universidade Federal de Roraima e, em 18 de Janeiro de 2008, submeteu-se aos exames finais de defesa de tese para obtenção do título de *Doctor Scientiae* em genética e Melhoramento sob a orientação do Prof<sup>o</sup>. Dro. Ricardo Frederico Euclides.

## SUMÁRIO

<b>RESUMO .....</b>	<b>VI</b>
<b>ABSTRACT .....</b>	<b>VIII</b>
<b>INTRODUÇÃO GERAL.....</b>	<b>1</b>
<b>REVISÃO DE LITERATURA.....</b>	<b>3</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>8</b>
<b>CAPÍTULO 1 - ESTIMAÇÃO DE PARÂMETROS GENÉTICOS PARA PRODUÇÃO DE LEITE DE BÚFALOS DA RAÇA MURRAH UTILIZANDO FUNÇÕES PARAMÉTRICAS .....</b>	<b>11</b>
RESUMO.....	11
ABSTRACT.....	12
INTRODUÇÃO.....	13
MATERIAL E MÉTODOS .....	14
RESULTADOS E DISCUSSÃO .....	20
CONCLUSÕES.....	29
REFERÊNCIAS BIBLIOGRÁFICAS .....	30
<b>CAPÍTULO 2 - ESTIMAÇÃO DE PARÂMETROS GENÉTICOS PARA PRODUÇÃO DE LEITE DE BÚFALOS DA RAÇA MURRAH USANDO INFERÊNCIA BAYESIANA.....</b>	<b>34</b>
RESUMO.....	34
ABSTRACT.....	35
INTRODUÇÃO.....	36
MATERIAL E MÉTODOS .....	37
RESULTADOS E DISCUSSÃO .....	41
CONCLUSÕES.....	56
REFERÊNCIAS BIBLIOGRÁFICAS .....	56
<b>CONCLUSÕES GERAIS.....</b>	<b>60</b>

## RESUMO

BREDA, Fernanda Cristina, D.Sc., Universidade Federal de Viçosa, janeiro de 2008. **Modelos de dimensão infinita para estimação de parâmetros genéticos da produção de leite de búfalos da raça Murrah.** Orientador: Ricardo Frederico Euclides. Co-orientadores: Lúcia Galvão de Albuquerque e Humberto Tonhati.

Parâmetros genéticos para a produção de leite no dia do controle de primeiras lactações de búfalas da raça Murrah foram estimados por meio de modelos de regressão aleatória, utilizando a metodologia da máxima verossimilhança restrita e Bayesiana. Para descrever as mudanças nas variâncias com o decorrer dos dias em lactação foram utilizados 17.935 registros da produção de leite no dia do controle, provenientes da primeira lactação de 1.433 búfalas. Os modelos incluíram os efeitos fixos de grupo contemporâneo, número de ordenhas e a covariável idade da vaca ao parto. Primeiramente foram analisados 27 modelos de regressão aleatória e os efeitos genético aditivo e de ambiente permanente foram modelados usando a função de Wilink, Ali e Schaeffer e polinômios ortogonais de Legendre. Também foram avaliadas diferentes formas de ajuste do efeito residual. Os modelos foram comparados usando o critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC), Log da função de máxima verossimilhança (LMV), medida da informação teórica da complexidade do modelo (ICOMP), percentagem relativa da redução da complexidade (PRRC) e variância residual ponderada (VRP). Na segunda parte do trabalho, foram testados 12 modelos que usaram diferentes combinações dos polinômios ortogonais de ordens três, quatro, cinco, seis e sete para os efeitos genético aditivo e de ambiente permanente. Os modelos foram comparados pelas estimativas do quadrado médio do resíduo, componentes de variâncias, herdabilidades, correlações, autovalores e autofunções. Os resultados obtidos indicam que: os modelos de regressão aleatória foram adequados em descrever a variação genética da produção de leite, com exceção do início e final da lactação, sendo que os seguintes modelos de regressão aleatória podem ser indicados para modelar a produção de leite no dia do controle de búfalas, em programas de avaliação genética: modelo que usou a função de Ali

e Schaeffer com variância residual homogênea (AS\_HO); e o modelo que utilizou a ordem de ajuste do polinômio para o efeito genético aditivo e de ambiente permanente igual quatro e seis, respectivamente, com seis classes de variâncias residuais (PL466).

## ABSTRACT

BREDA, Fernanda Cristina, D.Sc., Universidade Federal de Viçosa, January 2008. **Infinite dimension models for estimation of genetic parameters of milk production from buffaloes Murrah breed.** Adviser: Ricardo Frederico Euclides. Co-advisers: Lúcia Galvão de Albuquerque and Humberto Tonhati.

Genetic parameters for first lactation test-day milk yields of buffaloes Murrah breed were estimated using random regression models, using restricted maximum likelihood and Bayesian methods. A total of 17,935 test-day milk records of 1,433 buffaloes were used to describe the changes that happen in the variances along lactation curve. The models included fixed effects of contemporary group, milking number and age of buffalo as covariable. Firstly were analyzed twenty seven random regression models using Wilmlink, Ali & Schaeffer function or Legendre orthogonal polynomial to fit additive genetic and permanent environmental effects. The models were compared by Akaike's information criterion (AIC), Bayesian information criterion (BIC), Log likelihood function, information-theoretic measure of model complexity (ICOMP), percentage relative reduction of complexity (PRRC) and weighted residual variance (VRP). In second part of this work, were tested 12 models, varying the orders of the orthogonal polynomial (third, fourth, fifth, sixth and seventh orders) for the additive and permanent environmental effects. The models were compared by residual mean square errors, variances components, heritabilities, correlations, eigenvalues and eigenfunctions estimates. The results indicated that random regression models were adequate to describe the genetic variation for test-day milk yield, except for the beginning and at the end of the lactation, and the following random regression models can be recommended to fit test-day milk yields of buffaloes in evaluation genetic programs: model using Ali and Schaeffer function with homogeneous residuals variances (AS\_HO) and models using fourth and sixth order for additive genetic and permanent environmental effect, respectively, with six residuals variances classes (PL466).

## INTRODUÇÃO GERAL

A população mundial de búfalos (*Bubalus bubalis*) estimada é de 177 milhões de cabeças (13% do rebanho bovino mundial), em que 96,94% deste efetivo encontra-se na Ásia, 2,25% na África, 0,63% na América, 0,18% na Europa e 0,004% na Oceania. Mundialmente o principal produto do búfalo é o leite, representando cerca de 12% de todos os tipos de leite explorados. No período de 1996 a 2006 a produção mundial de leite de búfalas aumentou 39%, sendo esta taxa bem superior ao aumento verificado na produção de leite de vaca de 17%, no mesmo período (FAO, 2008).

O Brasil possui um efetivo de 1,2 milhões de búfalos, os quais 50% encontram-se na região Norte, onde o enfoque principal é a produção de carne. Já na região sudeste a produção leiteira vem aumentando continuamente nos últimos anos, em razão dos bons preços pagos pelos laticínios especializados na produção da "*mozzarella*", que foi introduzida no Brasil em 1922, pelo conde Matarazzo. A verdadeira "*mozzarella*" é de origem Italiana, onde somente é permitido o uso deste nome para produtos elaborados exclusivamente com leite de búfala.

Embora o Brasil tenha a maior população de búfalos da América, a produtividade do rebanho brasileiro ainda é baixa. A Índia possui as melhores médias de produção de leite das raças Murrah (2.226kg) e Jafarabadi (2.100kg) e a Itália da raça Mediterrâneo (2.169kg) (Sethi, 2003; ANASB, 2007), enquanto no Brasil a média da produção de leite da raça Murrah, considerada a maior produtora de leite do país, é de aproximadamente 1.500kg por lactação (Tonhati et al., 2000). As diferenças de produção de leite, entre os países são atribuídas às divergências nas condições ambientais e na constituição genética do rebanho, já que na Índia e na Itália a avaliação genética é realizada regularmente (Moioli & Borghese, 2005). No Brasil, apenas em 2005, o Ministério da Agricultura, Pecuária e Abastecimento (MAPA) e a Associação Brasileira de Criadores de Búfalos (ABCB) implementaram o "Programa de Melhoramento Genético de Búfalos leiteiros", nos moldes do teste de progênie da raça Gir. No entanto, os custos elevados do controle leiteiro têm se tornado um grande obstáculo para a obtenção de uma escrituração zootécnica de

qualidade e em quantidade suficiente para boas avaliações genéticas dos animais. Além disso, gastam-se, aproximadamente, seis anos para se avaliar um touro pela progênie, utilizando-se informações de lactações completas das filhas dos touros. O ajuste da produção de leite no dia do controle permitiria acelerar esse processo, pela diminuição do intervalo de geração, ao se concluir a prova dos touros em menor tempo.

Na avaliação quantitativa de medidas repetidas, os modelos usualmente empregados (unicaracterística, de repetibilidade e multicaracterísticas) são questionados quanto à sua praticidade e adequabilidade às situações encontradas. Por outro lado, os modelos de regressão aleatória vêm sendo indicados como alternativa na avaliação genética de dados longitudinais, por especificar os efeitos ambientais relacionados a um determinado estágio da lactação; utilizar mais informações de um mesmo animal, melhorando a acurácia nas estimativas de parâmetros genéticos, principalmente de espécies cuja quantidade de informações é limitada, como é o caso dos búfalos; possibilitar a estimação de parâmetros genéticos em qualquer ponto da curva de lactação, no intervalo em que as medidas foram tomadas, mesmo de animais com lactações incompletas, possibilitando avaliações genéticas mais freqüentes e redução do intervalo de geração.

### Objetivos

- Avaliar, por meio de diferentes critérios estatísticos, modelos de regressão aleatória, utilizando diferentes funções paramétricas e classes de variâncias residuais para descrever as mudanças nas variâncias que ocorrem no decorrer da lactação de búfalas; e apontar qual destes modelos é o mais adequado para ser utilizado em programas de avaliação genética.

- Estimar, por meio de modelos de regressão aleatória, funções de covariância para os efeitos genético aditivo e de ambiente permanente, descritas por polinômios ortogonais de Legendre, e parâmetros genéticos para as produções semanais de leite de búfalas da raça Murrah, utilizando inferência Bayesiana.

## REVISÃO DE LITERATURA

A produção de leite é uma característica medida no mesmo animal, em diferentes estágios fisiológicos (dados longitudinais), o que merece especial atenção, visto que diferentes genes podem se expressar em diferentes idades, causando alterações fisiológicas e no desempenho do indivíduo. Kirkpatrick & Heckman (1989) definem como caracteres de dimensão infinita, ou seja, caracteres oriundos de dados longitudinais, aqueles o qual o fenótipo de um indivíduo é descrito por uma função, ao invés de um número finito de mensurações. A produção de leite é um exemplo de característica de dimensão infinita, a qual pode ser considerada como uma função que relaciona os dias em lactação de um animal com algumas mensurações de sua produção. Assim, a produção de leite para cada idade diferente, pode ser considerada uma diferente característica, existindo, portanto, um número infinito de idades.

Um modelo que considera a produção no dia de controle para a avaliação genética dos animais, pode quantificar fatores que são específicos para cada dia de controle, fatores estes que mudam de animal para animal e de um registro para o outro. Um modelo completo multicaracterístico assume que cada controle diário dentro de uma lactação é uma característica distinta. Esta pressuposição resulta numa análise superparametrizada, devido à alta correlação entre os controles diários. Nestas situações, a utilização de funções de covariância pode ser uma alternativa mais adequada para descrever o espaço paramétrico completo com a estimação de um número menor de parâmetros.

Kirkpatrick et al. (1990) desenvolveram dois métodos de estimação de funções de covariância aditiva. O primeiro método, denominado de método de ajuste completo, procura ajustar o grau do polinômio ortogonal usando  $k$  funções polinomiais igual ao número de idades mensuradas ( $t$ ). No segundo método, denominado de método de ajuste reduzido, produz uma estimativa reduzida, ou seja, o ajuste da função de covariância é baseado em  $k$  funções ortogonais, em que  $k$  é menor que  $t$ , e  $t$  é a dimensão de  $\hat{G}$  (matriz de (co)variância genética aditiva), isto é, o número de idades mensuradas. O método consiste de dois passos. Primeiro, a função de covariância candidata é

construída por meio de quadrados mínimos ponderados para ajustar uma função ortogonal mais simples. Segundo, a função candidata é testada (usando-se a distribuição de  $X^2$ ) para a verificação de sua consistência estatística com a matriz  $\hat{G}$ . Se a função diferir estatisticamente de  $\hat{G}$ , então considera-se uma estimativa de função de covariância reduzida mais complexa em relação à anterior. A nova função estimada é então testada. Se o teste for não significativo, a função é considerada consistente com  $\hat{G}$ , caso contrário, o processo se repete iterativamente.

Uma maneira para estimar a função de covariância por meio do método de Máxima Verossimilhança Restrita (REML) foi proposta por Meyer & Hill (1997). O maior benefício deste método, em relação ao anterior, está no fato de que a matriz estimada de coeficientes  $k$  é positiva definida, o qual não é o caso de se usar o método de quadrados mínimos ponderados de Kirkpatrick et al. (1990).

A vantagem deste procedimento sobre o modelo multicaracterístico com  $t$  características é que a dimensão do espaço paramétrico é reduzida e há somente  $\frac{k(k+1)}{2}$  parâmetros para serem estimados, ao invés de  $\frac{t(t+1)}{2}$ . O máximo da verossimilhança pode ser encontrado com menos esforço computacional.

O modelo de função de covariância é equivalente ao modelo de regressão aleatória, quando se ajusta as funções que descrevem as idades como covariáveis (Meyer, 1998). Quando se ajusta um modelo de regressão aleatória, assume-se, implicitamente, certa estrutura de covariância entre as observações, determinada pelas covariâncias entre os coeficientes de regressão e caracterizada como uma função de covariância. A função de covariância é uma função bivariada (Função de variáveis contínuas), sendo análoga a matriz de covariâncias amplamente empregada na estatística multivariada e na genética quantitativa. Desta forma, uma função de covariância pode ser decomposta em autofunções e autovalores, da mesma forma que uma matriz de covariâncias pode ser escrita em termos de seus autovalores e autovetores. Os Polinômios ortogonais constituem uma família satisfatória para descrever as funções de covariância. Todavia, funções

paramétricas, como as de Wilmink (Wilmink, 1987) e a de Ali e Schaeffer (Ali & Schaeffer, 1987) também vem sendo aplicadas na modelagem de curvas de lactação (López-Romero & Carabaño, 2003; Cobuci et al., 2005; Breda et al., 2006). Modelos de regressão aleatória podem ser utilizados para descrever a variabilidade genética da produção de leite de búfalos, utilizando a mesma estrutura de produção no dia do controle. As diferenças estariam na definição de subclasses para as curvas fixas, intervalos de tempo e a ordem das regressões que deverão ser adotadas (Schaeffer, 2004).

Outro método para se estimar os coeficientes de funções de covariância além do REML, é o Bayesiano, com base na Cadeia de Markov Monte Carlo (MCMC). Esses métodos apresentam várias diferenças nas suas metodologias para fazer inferências. Primeiramente, a expressão da incerteza sobre os parâmetros desconhecidos. A escola freqüentista estuda a distribuição do estimador, enquanto a metodologia Bayesiana deseja obter a função densidade de probabilidade do parâmetro dado um conjunto de observações. A partir dessa distribuição, pode-se obter o valor mais provável do parâmetro, ou ainda, a probabilidade do parâmetro residir dentro de certos limites. A inferência freqüentista baseia-se em como um grande número de estimativas estaria distribuído em torno do valor verdadeiro caso um número infinito de repetições do experimento fosse realizado, enquanto a Bayesiana examina a probabilidade da distribuição do valor verdadeiro, conforme a ocorrência dos dados observados. Para a frequentista, usualmente, o valor verdadeiro é fixo e a amostra é aleatória, enquanto para a Bayesiana, a amostra é fixa e o parâmetro de interesse é uma variável aleatória, ou seja, como o parâmetro é desconhecido, fala-se sobre a probabilidade do parâmetro assumir um determinado valor. Além disso, os métodos com base na estatística Bayesiana possuem a vantagem de permitir a inclusão na análise do conhecimento *a priori* existente sobre os parâmetros desconhecidos.

Um segundo ponto, refere-se ao fato de que alguns conceitos usualmente empregados no melhoramento animal não possuem uma interpretação Bayesiana, por exemplo: viés e diferença entre efeitos fixos e aleatórios. No contexto Bayesiano, o viés não existe, uma vez que repetições conceituais do experimento não são consideradas. Assumem-se também todos os efeitos como aleatórios porque a forma Bayesiana de expressar incerteza sobre os

parâmetros desconhecido é obtendo suas funções densidade de probabilidade. Em terceiro lugar, as inferências obtidas em ambas as escolas não são sempre coincidentes, particularmente para amostras pequenas (Wang et al., 1994).

Na inferência freqüentista, repetindo-se conceitualmente um experimento um infinito número de vezes, pode-se chegar a um número infinito de intervalos de confiança, os quais incluiriam o verdadeiro valor do parâmetro em, por exemplo, 95% dos casos. O objetivo da inferência Bayesiana é encontrar a densidade de probabilidade do parâmetro ( $\theta$ ), fornecidos os dados ( $f(\theta | y)$ ), em que:  $y$  é o vetor de observações. Quando essa distribuição é obtida pode-se determinar o menor intervalo no qual a probabilidade de encontrar  $\theta$  é maior que 95%. Caso o interesse seja na estimação pontual do parâmetro, a moda, a média e a mediana podem ser calculados a partir da distribuição ( $f(\theta | y)$ ).

### Inferência Bayesiana

Se  $\theta$  é um vetor de parâmetros e  $y$  é um vetor de observações, então a função densidade de probabilidade conjunta de  $\theta$  e  $y$  ( $f(\theta, y)$ ) pode ser definida como:

$$f(\theta, y) = f(\theta) * f(y | \theta) = f(y) * f(\theta | y)$$

Rearranjando

$$f(\theta | y) = \frac{f(\theta) f(y | \theta)}{f(y)}$$

em que a última expressão representa o teorema de Bayes. Como  $f(y)$  não é uma função  $\theta$ , tem-se uma forma usual do teorema de Bayes (Gianola & Fernando, 1986):

$$f(\theta | y) \propto f(\theta) * f(y | \theta),$$

sendo,  $f(\theta)$  a função densidade de probabilidade, a *priori* de  $\theta$ , que reflete a relativa incerteza sobre os possíveis valores de  $\theta$ , antes do vetor  $y$  ter sido observado;  $f(y | \theta)$  é a função densidade de probabilidade de  $y$  condicional a  $\theta$ , que representa a verossimilhança, ou seja, a quantidade de informação contida em  $y$  para o conhecimento de  $\theta$ ;  $f(\theta | y)$  é a função de densidade de

probabilidade *a posteriori*, que reúne em si, o estado de incerteza sobre  $\theta$ , dado o conhecimento prévio sobre os parâmetros e a informação contida no vetor de observações; e  $\propto$  é o operador que indica proporcionalidade.

Quando se analisa um conjunto de dados de grande volume, a informação *a priori* tende a ser subjugada pela função de verossimilhança no estabelecimento da *posteriori*. Neste caso, as estimativas a respeito dos parâmetros em  $\theta$ , são próximas àquelas obtidas por meio de métodos freqüentistas. A distribuição *posteriori* de  $\theta$  contém todas as informações necessárias para se fazer inferência a respeito de  $\theta$ . No entanto, é necessário integrar a função  $f(\theta | y)$ , o que inviabilizou por muito tempo a aplicação da inferência Bayesiana.

A amostragem de Gibbs é uma técnica de integração numérica por simulação, muito usual em situações em que a integração analítica completa é impossível. Casella & George (1992) definem o amostrador de Gibbs como um procedimento numérico para obter amostra da distribuição conjunta ou marginal de todos os parâmetros do modelo, através de repetidas amostragens de todas as distribuições condicionais completas formando uma cadeia de Markov. Os métodos da Cadeia de Markov Monte Carlo (MCMC) formam um conjunto de métodos iterativos, com base em simulação estocástica, que produzem uma seqüência de amostras aleatórias (cadeia de Markov), que após grande número de rodadas convergem para uma distribuição estacionária única de equilíbrio. A aplicação de métodos da MCMC, dentre os quais se destaca o amostrador de Gibbs, é relativamente recente e tem sido sugerido por alguns autores para uso em contexto que envolve o melhoramento animal (Wang et al., 1994; Sorensen et al., 1994; Jamrozik & Schaeffer, 1997; Jamrozik et al., 2001; López-Romero et al., 2003; Meyer, 2005).

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALI, E.; SCHAEFFER, L.R. Accounting for covariances among test day milk yields in dairy cows. **Canadian Journal Animal Science**. v.67, n.3, p.637-644, 1987.
- ANASB – Associazione Nazionale Allevatori specie bufalina. Statistiche. [www.anasb.it](http://www.anasb.it). Acessado em dezembro de 2007.
- BREDA, F.C.; ALBUQUERQUE, L.G.; YAMAKI, M. et al. Estimação de parâmetros genéticos para a produção de leite de cabras da raça Alpina. **Revista Brasileira de Zootecnia**, v.35, n.2, p.396-404, 2006.
- CASELLA, G.; GEORGE, E.I. Explain the Gibbs Sampler. **The American Statistician**, v.46, n.3, p.167-174, 1992.
- COBUCI, J.A.; EUCLYDES, R.F.E.; LOPES, P.S. et al. Estimation of genetic parameters for test-day milk yield in Holstein cows using a random regression model. **Genetics and Molecular Biology**, v.28, n.1, p.75-83, 2005.
- FAO - Food and Agriculture Organization. <http://faostat.fao.org>. Acessado em janeiro de 2008.
- GIANOLA, D.; FERNANDO, R.L. Bayesian Methods in Animal Breeding theory. **Journal Animal Science**, v.63, p. 217-244, 1986.
- JAMROZIK, J.; SCHAEFFER, L.R. Estimates of genetic parameters for a test day model with random regression for yield traits of first lactation Holstein. **Journal of Dairy Science**, v.80, n.4, p.762-770, 1997.
- JAMROZIK, J.; GIANOLA, L.R.; SCHAEFFER, L.R. Bayesian estimation of genetic parameters for test day records in dairy cattle using linear hierarchical models. **Livestock Production Science**, v.71, p.223-240, 2001.

- KIRKPATRICK, M.; HECKMAN, N. A. Quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. **Journal of Mathematical Biology**, v.27, n.2, p.429-450, 1989.
- KIRKPATRICK, M.; LOFSVOLD, D.; BULMER, M. Analysis of the inheritance, selection and evolution of growth trajectories. **Genetics**, v.124, n.3, p.979-993, 1990.
- LÓPEZ-ROMERO, P.; REKAYA, R.; CARABAÑO, M.J. Assessment of homogeneity vs heterogeneity of residual variance in random regression test-day models in a Bayesian analysis. **Journal Dairy Science**, v.83, p.3374-3385, 2003.
- LÓPEZ-ROMERO, P.; CARABAÑO, M.J. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein Friesian cattle. **Livestock Production Science**, v.82, p.81-96, 2003.
- MEYER, K.; HILL, W. G. Estimation of genetic and phenotypic covariance functions for longitudinal or "repeated" records by restricted maximum likelihood. *Livestock Production Science*, 47(3):185-200, 1997.
- MEYER, K. Modeling "repeated" records: covariance functions and random regression models to analyse animal breeding data, In: WORLD CONGRESS GENETIC APPLIED LIVESTOCK PRODUCTION, 6., 1998, Armidale, Australia, **Proceedings...** Armidale: University of New England, 1998. CD-ROM.
- MEYER, K. Estimates of genetic covariance function for growth of Angus cattle. **Journal Animal Breeding Genetics**, v.122, p. 73-85, 2005.
- MOIOLI, B; BORGHESE, A. Buffalo breeds and management systems. In: **Buffalo Production Research**. Istituto Sperimentale per la Zootecnia. Animal. Monterotondo: Rome, Italy p.51-76, 2005.
- SETHI, R.K. Búfalo Breeds of Índia. In: ASIAN BUFFALO CONGRESS, 4., 2003, New Delhi. **Proceedings...** New Delhi: Índia, 2003.

- SCHAEFFER, L.R. Application of regression models in animal breeding. **Livestock Production Science**. v.86, p. 35-45, 2004.
- SORENSEN, D.A.; WANG, C.S.; JENSEN, J.; GIANOLA, D. Bayesian analysis of genetic change due to selection using Gibbs sampling. **Genetics Selection Evolution**, v.26, p.333-360, 1994.
- TONHATI, H.; et al. Parâmetros genéticos para a produção de leite, gordura e proteína em bubalinos. **Revista Brasileira de Zootecnia.**, v.29, n.6, p.2051-2056, 2000.
- WANG, C.S.; RUTLEDGE, J.J; GIANOLA, D. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. **Genetic Selection Evolution**, v.26, p.91-115, 1994.
- WILMINK, J.B.M. Efficiency of selection for different cumulative milk, fat and protein yields in first lactation. **Livestock Production Science**, v.17, n.3, p.211-224, 1987.

## Capítulo 1

### Estimação de Parâmetros Genéticos para Produção de Leite de Búfalos da Raça Murrah Utilizando Funções Paramétricas

**Resumo** – Foram utilizados 27 modelos de regressão aleatória para ajustar 17.935 registros da produção de leite no dia do controle, provenientes da primeira lactação de 1.433 búfalas. Os modelos incluíram os efeitos fixos de grupo contemporâneo, número de ordenhas e a covariável idade da búfala ao parto. Os efeitos genético aditivo e de ambiente permanente foram modelados por meio de regressão aleatória, usando a função de Wilmink, Ali e Schaeffer e polinômios ortogonais de Legendre. Também foram avaliadas diferentes formas de ajuste do efeito residual. Os modelos foram comparados usando o critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC), Logaritmo da função de máxima verossimilhança (LMV), medida da informação teórica da complexidade do modelo (ICOMP), porcentagem relativa da redução da complexidade (PRRC) e variância residual ponderada (VRP). A função de Ali e Schaeffer apresentou melhores valores de LMV, VRP e ICOMP para o AS\_HE42, AIC, VRP e índice para o AS\_HE17 e de BIC para o AS\_HO, enquanto a função de Wilmink resultou apenas em melhor valor de PRRC (W\_HE42). Da 2<sup>a</sup> a 38<sup>a</sup> semanas o AS\_HE17, AS\_HO e W\_HE42 apresentaram comportamento semelhante, principalmente o AS\_HE17 e AS\_HO. Estes modelos também foram parecidos nas estimativas produzidas nos períodos com menor quantidade de registros (início e final da lactação), sendo as maiores diferenças visualizadas nas estimativas de variância residual e de ambiente permanente. Os modelos de regressão aleatória foram adequados em descrever a variação genética da produção de leite, com exceção do início e final da lactação, sendo o AS\_HO o que melhor ajustou a curva de lactação de búfalas da raça Murrah, segundo o critério de informação Bayesiano ou BIC.

Palavras-chave: produção de leite, variâncias, funções paramétricas, regressão aleatória

## Chapter 1

### Genetic Parameters Estimation for Milk Yield of Buffaloes Murrah Breed Using Parametric Functions

**Abstract** – Twenty seven random regression models were used to adjust 17,935 test-day milk records of 1,433 buffaloes. The models include fixed effects of contemporary group, milking number and age of buffalo as covariable. The additive genetic and permanent environmental effects were modeled through random regression models, using Wilmink and Ali & Schaeffer functions and Legendre orthogonal polynomial. Different structures of residual variances were fitted. Models were compared by Akaike's information criterion (AIC), Bayesian information criterion (BIC), Log likelihood function, information-theoretic measure of model complexity (ICOMP), percentage relative reduction of complexity (PRRC) and weighted residual variance (VRP). The Ali and Schaeffer function presented better values of LMV, VRP and ICOMP to AS\_HE42, AIC, VRP and index to AS\_HE17 and BIC to AS\_HO, while the Wilmink function resulted in better value of PRRC to W\_HE42. The AS\_HE17, AS\_HO and W\_HE42 presented estimate similar behavior in 2nd to 38th weeks, mainly AS\_HE17 and AS\_HO. These models were also similar in the estimates produced in the periods with smaller amount of records (beginning and at the end of the lactation), being the largest differences visualized in the residual and permanent environmental variances estimates. Random regression models were efficient in describing the milk production genetic variation, except at the beginning and at the end of the lactation, being AS\_HO presented the best fit lactation trajectory of Murrah buffaloes breed, according with the Bayesian information criterion or BIC.

Key Words: milk yield, parametric functions, random regression, genetic variances.

## Introdução

O rebanho nacional de búfalos é composto por 1,2 milhões de cabeças, constituindo-se o maior rebanho de búfalos das Américas (FAO, 2008). A região norte, com 50% do rebanho brasileiro, e sul se destacam na criação de búfalos destinados à produção de carne. Na região sudeste e centro oeste a criação de búfalos leiteiros tem se tornado uma atividade potencialmente rentável, em virtude dos bons preços pagos pelos laticínios especializados na produção da "*mozzarella*". Com o intuito de aproveitar o potencial do país na fabricação de derivados lácteos, os produtores nacionais estão intensificando ações para melhorar a qualidade genética do rebanho. A variabilidade da produção leiteira das búfalas é ainda muito ampla, podendo ser identificados animais produzindo de 1.308 a 4.619kg de leite por lactação (Tonhati et al., 2006), dependendo da qualidade genética e do sistema de criação do rebanho. Portanto, a intensificação do processo de melhoramento genético encontra nos búfalos um excelente potencial de desenvolvimento.

Atualmente, a avaliação genética de bovinos leiteiros usando modelos que consideram a produção de leite no dia do controle, em substituição àqueles que utilizam a produção total na lactação, tem sido adotada em muitos países. A produção de leite total é obtida a partir dos registros de produção no dia do controle e muitas vezes é predita a partir de poucas observações ou de lactações não finalizadas, levando a estimativas tendenciosas. Os fatores de projeção assumem uma forma padrão da curva de lactação dependendo da raça e número de lactações, dessa forma, vacas que possuem maior persistência de lactação, geralmente, tem sua produção subestimada e vice-versa (Jamrozik & Schaeffer, 1997). Além disso, produções acumuladas podem ser influenciadas por diferentes efeitos de ambiente temporário, sendo estes difíceis de serem quantificados com este tipo de modelo (Jakobsen et al., 2002).

Na avaliação quantitativa de medidas repetidas, os modelos usualmente empregados (unicaracterística, de repetibilidade e multicaracterísticas) são questionados quanto à sua praticidade e adequabilidade às situações encontradas. Por outro lado, os modelos de regressão aleatória vêm sendo indicados como alternativa na avaliação genética de dados longitudinais, por permitir a especificação dos efeitos ambientais relacionados a um determinado

estágio da lactação; utilização de maior número de informações de um mesmo animal, melhorando a acurácia das estimativas de parâmetros genéticos, principalmente de espécies cuja quantidade de informações é limitada, como é o caso dos búfalos. Além disso, possibilita a estimação de parâmetros genéticos em qualquer ponto da curva de lactação, no intervalo em que as medidas foram tomadas, mesmo de animais com lactações incompletas, permitindo avaliações genéticas mais freqüentes e redução do intervalo de gerações. No entanto, dos poucos trabalhos encontrados na literatura sobre melhoramento genético da produção de leite de búfalos, a maioria restringe-se a produção acumulada aos 270 dias de lactação e produção no dia do controle usando modelos uni e multicaracterísticos.

Dessa forma, o objetivo neste trabalho foi avaliar, por meio de diferentes critérios estatísticos, modelos de regressão aleatória, utilizando diferentes funções paramétricas e classes de variâncias residuais para descrever as mudanças nas variâncias que ocorrem no decorrer da lactação de búfalas; e apontar qual destes modelos é o mais adequado para ser utilizado em programas de avaliação genética.

### **Material e Métodos**

O banco de dados inicial constituía-se de 65.101 controles leiteiros provenientes de 2.074 lactações de búfalas da raça Murrah, pertencentes a rebanhos supervisionados pelos técnicos do programa de controle leiteiro de búfalos do Departamento de Zootecnia da Faculdade de Ciências Agrárias e Veterinária (UNESP, Jaboticabal).

Para o ajuste dos modelos de regressão aleatória, as produções de leite no dia do controle foram consideradas em classes semanais, sendo utilizados os registros obtidos entre a 2<sup>a</sup> e a 43<sup>a</sup> semana após o parto de búfalas com idades entre 24 a 48 meses com, no mínimo, quatro controles de produção durante a lactação, sendo o primeiro registro realizado até 75 dias após o parto. Adicionalmente, foram excluídos animais com três desvios-padrão acima ou abaixo da média de produção de leite de animais da mesma semana de lactação. Após estas restrições, foram utilizadas nas análises estatísticas: 17.935 registros da produção de leite no dia do controle, provenientes da primeira lactação de 1.433 búfalas, controladas de 1985 a 2006, em 11

rebanhos localizados no estado de São Paulo. A média do número de animais dentro dos 708 grupos contemporâneos formados pela fazenda, ano e mês do controle foi igual a 27, sendo o número mínimo igual a cinco e o máximo igual a 66 animais. Na Figura 1 são apresentados o número de registros, a média e o desvio-padrão para produção de leite referente à semana do controle.

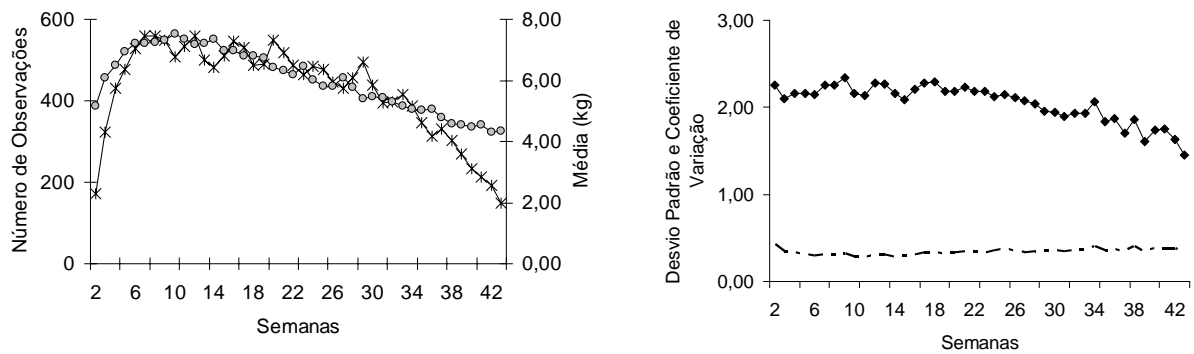


Figura 1 - Número de registros (\*), média (°), desvio-padrão(◆) e coeficiente de variação dividido por 100 ( ---) da produção de leite referente à semana do controle de primeiras lactações de búfalas da raça Murrah.

A produção de leite no dia do controle foi analisada por meio de um modelo animal, unicaracterístico, de regressão aleatória, considerando como efeitos fixos o grupo contemporâneo (708 níveis), número de ordenhas (dois níveis) e os efeitos linear e quadrático da covariável idade da búfala ao parto, em meses. Considerou-se também a regressão sobre o tempo (semana da lactação), o que gerou a curva média da lactação ou a trajetória média da população (fixa). Em todos os modelos a trajetória média da população foi modelada por polinômios ortogonais de Legendre de ordem três.

Na forma matricial o modelo é dado por:

$$y = Xb + Za + Wc + e,$$

em que:  $y$  é o vetor das  $N$  observações, medidas em  $N_d$  animais;  $b$  é o vetor de efeitos fixos;  $a$  é o vetor de coeficientes de regressão aleatória genético aditivo;  $c$  é o vetor dos coeficientes de regressão aleatória de ambiente permanente;  $e$  é o vetor de efeito aleatório residual;  $X$ ,  $Z$  e  $W$  são as matrizes de incidência para os efeitos fixos, aleatórios genético aditivo e de ambiente permanente, respectivamente, para os quais assumem-se:

$$E \begin{bmatrix} y \\ a \\ c \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad e \quad \begin{aligned} V(a) &= K_a \otimes A; \\ V(c) &= K_c \otimes I_{Nd}; \\ V(e) &= R, \end{aligned}$$

sendo,  $K_a$  e  $K_c$  as matrizes de (co)variâncias entre os coeficientes de regressão aleatória genético aditivo e de ambiente permanente, respectivamente;  $A$  a matriz dos numeradores dos coeficientes de parentesco entre animais;  $I_{Nd}$  é a matriz identidade de dimensão  $N_d$ ;  $\otimes$  é o operador de produto direto entre matrizes; e  $R$  é uma matriz diagonal de variâncias residuais.

Os modelos empregados diferiram entre si pelo uso de diferentes funções para descrever a trajetória da lactação. Foram estimadas funções de covariâncias para os efeitos genético aditivo e de ambiente permanente, modeladas pelas seguintes funções:

- a) Função exponencial de Wilmink (Wilmink, 1987), caracterizada por quatro parâmetros e descrita por:  $W(t) = a_0 + a_1t + a_2 \exp^{-a_3t}$ ;
- b) Função logarítmica de Ali e Schaeffer (Ali & Schaeffer, 1987), caracterizada por cinco parâmetros e representada por:  $R(t) = a_0 + a_1u + a_2u^2 + a_3v + a_4v^2$ , em que:  $u = t/305$ ,  $v = \ln(305/t)$ ;  $t$  = dias em lactação; e  $a_i$  = coeficientes de regressão;
- c) Polinômios ortogonais, numa escala de Legendre, em que o  $j$ -ésimo polinômio é dado pela seguinte expressão:
$$\phi_j(t_i^*) = \frac{1}{2^j} \sqrt{\frac{2j+1}{2}} \sum_{m=0}^{[j/2]} (-1)^m \binom{j}{m} \binom{2j-2m}{j} a_i^{*j-2m}$$
, sendo que no resultado da divisão  $[j/2]$  os valores são arredondados para baixo, assumindo o valor inteiro mais próximo; e  $t^*$  = dias em lactação padronizado. Utilizaram-se os polinômios ortogonais de mesma ordem de ajuste das funções paramétricas (cúbica e quártica) e classes de variâncias residuais com o propósito de comparação.

Na curva de Wilmlink, o quarto parâmetro ( $a_3$ ) é geralmente considerado uma constante, reduzindo o número de parâmetros a serem estimados de quatro para três, simplificando o ajuste da curva. Neste estudo, foram feitas modificações no termo exponencial, em que considerou-se  $a_3$  igual a -0,05, padrão da função proposta por Wilmlink; e -0,10 valor proposto por Brotherstone et al. (2000). Além disso, no segundo termo ( $a_1t$ ), dividiu-se  $t$  por 100, no intuito de melhorar as propriedades numéricas do modelo (Jakobsen et al., 2002).

Também foram avaliadas diferentes formas de ajuste do efeito residual, são elas: variância residual homogênea; e variância residual heterogênea, considerando classes ou funções de variâncias residuais de ordem quatro, cinco e seis, modeladas por polinômios ordinários. Para definir as classes de variância residual, primeiramente processou-se a análise considerando 42 classes de resíduo (2ª a 43ª semana) e após agruparam-se as variâncias mais semelhantes. Esse procedimento foi feito para os modelos que usaram a função de Wilmlink e Ali e Schaeffer sendo esta a razão pela qual, as classes de variâncias residuais obtidas por esses modelos não coincidiram, como pode ser visualizado na Tabela 1.

Tabela 1 - Classes de variâncias residuais para os modelos que aplicaram a função de Wilmlink e Ali e Schaeffer

<b>Número de Classes</b>	<b>Classes Residuais</b>
<b>Wilmlink</b>	
13	2ª; 3ª; 4ª e 5ª; 6ª; 7ª; 8ª a 10ª; 11ª a 14ª; 15ª a 17ª; 18ª a 28ª; 29ª a 33ª; 34ª a 36ª; 37ª a 39ª; 40ª a 43ª
08	2ª; 3ª; 4ª e 5ª; 6ª; 7ª; 8ª a 14ª; 15ª a 17ª; 18ª a 43ª
02	2ª a 7ª; 8ª a 43ª
<b>Ali e Schaeffer</b>	
17	2ª; 3ª a 5ª; 6ª; 7ª; 8ª; 9ª e 10ª; 11ª; 12ª a 14ª; 15ª a 17ª; 18ª e 19ª; 20ª a 25ª; 26ª a 28ª; 29ª a 33ª; 34ª a 36ª; 37ª a 39ª; 40ª e 41ª; 42ª e 43ª
13	2ª; 3ª a 5ª; 6ª; 7ª; 8ª; 9ª e 10ª; 11ª a 14ª; 15ª a 17ª; 18ª a 25ª; 26ª a 28ª; 29ª a 33ª; 34ª a 36ª; 37ª a 43ª
06	2ª; 3ª a 5ª; 6ª; 7ª; 8ª a 36ª; 37ª a 43ª
02	2ª; 3ª a 43ª

No total, 27 modelos foram testados, os quais foram identificados por LEG (Polinômios Ortogonais de Legendre) seguido pela ordem de ajuste, W (função de Wilmink com  $a_3 = -0,05$ ), W\* (função de Wilmink com  $a_3 = -0,10$ ), WJ (função de Wilmink com  $a_3 = -0,05$  e segundo termo da função dividido por 100), W\*J (função de Wilmink com  $a_3 = -0,10$  e segundo termo da função dividido por 100) e AS (Função de Ali e Schaeffer). A identificação da forma de ajuste do efeito residual foi: HO (homogêneo), HE (heterogêneo), sendo apresentados na seqüência o número de classes do resíduo e F (função de variância residual) seguido da ordem de ajuste. Assim, LEG5\_HE17 indica um modelo de regressão aleatória usando polinômios de Legendre de ordem cinco com ajuste de 17 classes residuais e W\_F4 indica um modelo de regressão aleatória usando a função de Wilmink com  $a_3 = -0,05$  com ajuste de função de variância residual de ordem quatro.

Os componentes de (co)variâncias foram estimados pelo método da máxima verossimilhança restrita (REML), utilizando-se a opção DXMRR do pacote estatístico DFREML (Meyer, 1998a). A escolha do melhor modelo de regressão aleatória foi feita com base em diversos critérios estatísticos, como segue:

- Variância Residual Ponderada (Jamrozik & Schaeffer, 2002):

$$VRP = \frac{1}{TS} * \sum_{l=1}^f \sigma_{ef}^2 NS_l ;$$

- Logaritmo da Função de Máxima Verossimilhança:  $LMV = \ln L$ ;
- Critério de Informação de Akaike (Akaike, 1973):  
 $AIC = -2 \ln L + 2p$ ;
- Critério de Informação Bayesiano (Schwarz, 1998):  
 $BIC = -2 \ln L + p \ln (N - r)$ ;
- Medida da Informação Teórica da Complexidade do Modelo (Bozdogan, 2000):  $ICOMP = -2 \ln L + 2 (C_{Modelo})$ ; e
- Percentagem Relativa da Redução da Complexidade (Bozdogan, 2000):  $PRRC = \frac{(C_{Modelo} - C_{RModelo})}{C_{Modelo}}$ ,

em que:  $TS$  é o número total de semanas em lactação;  $f$  é o número de classes residuais;  $NS_l$  é o número de semanas existentes na  $l$ -ésima classe;  $p$  é o número de parâmetros do modelo;  $N$  é o número total de observações;  $r$  é o posto da matriz de incidência para os efeitos fixos;  $C_{Modelo} = C_{K_a} + C_{K_c}$ ,

sendo:  $C_{K_a}$  ou  $C_{K_c} = \frac{p(K)}{2} \ln \left[ \frac{\text{traço}(K)}{p(K)} \right] - \frac{1}{2} \ln|K|$ ;  $p(K)$  é o posto da matriz

$K$ ; e  $C_{RModelo}$  é calculado da mesma forma que  $C_{Modelo}$ , porém ao invés das matrizes de covariâncias utilizam-se as matrizes de correlações entre os coeficientes de regressão aleatória genético aditivo e de ambiente permanente.

A medida da informação teórica da complexidade (ICOMP) é um critério de seleção de modelos que informa o grau de interdependência entre os parâmetros do modelo. O primeiro componente do ICOMP ( $-2 \ln L$ ) mede a falta de ajuste e o segundo componente  $2(C_{Modelo})$  mede a interação entre os parâmetros do modelo, ou seja, a complexidade. Os resultados do ICOMP indicam o modelo mais simples que fornece as estimativas mais acuradas dos parâmetros. Além disso, pode-se observar que o  $\text{traço}(K)$  é a soma dos elementos da diagonal (soma das variâncias) e o determinante envolve também os elementos fora da diagonal (variâncias + covariâncias = variância geral). Portanto,  $C_{K_a}$  ou  $C_{K_c}$  equivale a relação entre a média geométrica ( $\bar{\lambda}_g$ ) e aritmética ( $\bar{\lambda}_a$ ) dos autovalores da matriz de variâncias e covariâncias

entre os coeficientes de regressão  $\left( C_K = \frac{p}{2} \ln \left( \frac{\bar{\lambda}_a}{\bar{\lambda}_g} \right) \right)$ . A maior simplicidade é

alcançada quando  $C_K$  é aproximadamente uma matriz identidade, sugerindo que os parâmetros são ortogonais e podem ser estimados com igual precisão.

A percentagem relativa da redução da complexidade do modelo (PRRC) nos fornece informação sobre a complexidade e o grau de multicolinearidade entre os parâmetros do modelo, não considerando em seu calculo a falta de ajuste do modelo, ou seja, o valor de LMV, como ocorre no ICOMP.

Pelo fato de cada critério estatístico eleger diferentes modelos, utilizou-se um índice, sendo este construído da seguinte forma:

$I = AIC + BIC + VRP + ICOMP + |LMV| - PRRC$  (adaptado de Liu et al., 2006). Os

resultados de cada critério de seleção de modelos são expressos por meio da comparação entre os mesmos, uma vez que menores valores de AIC, BIC, VRP, ICOMP e índice e maiores valores de LMV e PRRC são preferidos.

### Análises Univariadas

Para que as análises unicaracterísticas pudessem ser comparadas com as de regressão aleatória, os dados foram extraídos do mesmo conjunto de dados. Assim, foram realizadas nove análises unicaracterísticas por meio da máxima verossimilhança restrita, considerando as idades de 7 a 60; 30 a 90; 60 a 120; 90 a 150; 120 a 180; 150 a 210; 180 a 240; 210 a 270; e 240 a 300 dias de lactação.

### **Resultados e Discussão**

Na Tabela 2 são apresentados os resultados dos critérios estatísticos de seleção de modelos. Dentre os modelos de regressão aleatória que usaram a função de Wilmink sem modificações, o W\_HE42 foi o que proporcionou melhor ajuste dos dados segundo o AIC, LMV, VRP, ICOMP, PRRC e índice. O BIC apontou o W\_HE02 como mais adequado para descrever as mudanças nas variâncias que ocorrem ao longo da lactação. Assim, as análises com modificações na função de Wilmink foram realizadas apenas para o W\_HE42 e W\_HE02, os quais, de acordo com os critérios estatísticos de escolha de modelos, não melhoraram o ajuste da produção de leite no dia do controle.

A função de Wilmink quando comparada à de Ali e Schaeffer, resultou apenas em melhor valor de PRRC (W\_HE42), enquanto a função de Ali e Schaeffer forneceu os melhores resultados de LMV, VRP e ICOMP para o AS\_HE42; AIC, VRP e índice para o AS\_HE17; e de BIC para o AS\_HO. Porém, o AS\_HE42 não diferiu estatisticamente do AS\_HE17, pelo teste de razão de verossimilhança a 5% de probabilidade.

Vale ressaltar que o AIC e BIC penalizam modelos com maior número de parâmetros e o ICOMP penaliza modelos com maior multicolinearidade. Para o cálculo do AIC, BIC e ICOMP, utiliza-se o logaritmo da função de máxima verossimilhança (LMV). O que podemos observar, na Tabela 2, é que o AIC e ICOMP parecem ser mais afetados pelo LMV do que o BIC, tendendo a eleger

os modelos mais parametrizados. No BIC a penalidade, quanto ao número de parâmetros, foi mais rigorosa do que a verificada no AIC e ICOMP, tendendo a favorecer modelos mais parcimoniosos (Nunez- Antón & Zimmerman, 2000).

Tabela 2 – Critérios Estatísticos para escolha de modelos

Modelo	Critérios Estatísticos							
	p	AIC	BIC	LMV	VRP	ICOMP	PRRC	Índice
<i>Função de Wilmink</i>								
W_HE42	54	225	646	-558	0,98	133	84%	1479
W_HE13	25	395	590	-673	1,03	372	82%	1949
W_HE08	20	410	566	-685	1,03	397	83%	1976
W_HE02	14	430	539	-701	1,03	430	79%	2022
W_HO	13	464	565	-719	1,02	465	82%	2132
W_F4	16	436	561	-702	1,04	432	80%	2052
W_F5	17	435	568	-701	1,04	430	78%	2057
W_F6	18	433	573	-698	1,04	426	76%	2055
<i>Função de Wilmink Modificada</i>								
WJ_HE02	14	430	539	-701	1,03	422	60%	2033
W*_HE02	14	658	767	-815	1,06	660	82%	2819
W*_J_HE02	14	658	767	-815	1,06	650	69%	2822
WJ_HE42	54	427	848	-656	1,03	329	28%	2233
W*_HE42	54	650	1071	-771	1,04	573	83%	2983
W*_J_HE42	54	650	1071	-771	1,04	562	52%	3003
<i>Função de Ali e Schaeffer</i>								
AS_HE42	72	196	757	-526	0,93	103	30%	1553
AS_HE17	47	157	523	-531	0,93	109	24%	1297
AS_HE13	43	175	510	-545	0,96	139	54%	1316
AS_HE06	36	212	493	-570	0,95	182	45%	1413
AS_HE02	32	253	502	-594	0,96	227	57%	1520
AS_HO	31	242	484	-590	0,96	236	35%	1518
AS_F4	34	228	493	-580	0,99	215	45%	1472
AS_F5	35	229	502	-579	0,99	196	26%	1481
AS_F6	36	229	509	-578	0,99	197	54%	1460
<i>Polinômio de Legendre</i>								
LEG3_HE42	54	700	1121	-796	1,15	595	90%	3123
LEG5_HE17	47	175	541	-540	0,98	113	48%	1322
LEG5_HO	31	321	562	-629	0,96	286	39%	1760

p = número de parâmetros estimados; AIC = critério de informação de Akaike; BIC = critério de informação Bayesiano; LMV = função de verossimilhança; VRP = variância residual ponderada; ICOMP = informação teórica da complexidade do modelo; e PRRC = porcentagem da redução da complexidade.

Os modelos que empregaram a função de Wilmink e Ali e Schaeffer elegeram modelos diferentes, com relação ao agrupamento das classes de variância residuais (Tabela 1), sugerindo que diferentes modelos requerem diferentes estruturas de variância residual. Menores estimativas de VRP foram

verificadas quando se utilizou os modelos mais parametrizados e, segundo Meyer (1998b), esta tendência está relacionada ao aumento da capacidade destes modelos em ajustar a produção de leite no dia do controle no início da lactação. Resultados encontrados na literatura, em que a função de Ali Schaeffer foi superior à de Wilmink foram apresentados por Brotherstone et al. (2000), Costa et al. (2005), Breda et al. (2006) e Melo et al (2007). Jamrozik et al. (1997) relataram que estas funções foram semelhantes para classificar touros leiteiros aos 305 dias de lactação, todavia, a função de Ali e Schaeffer resultou em menor variância do erro de predição.

Ainda na Tabela 2, nota-se que os modelos W\_HE42, AS\_HE17 e AS\_HO foram mais adequados no ajuste dos dados do que os modelos que usaram polinômios de Legendre com a mesma ordem de ajuste para os efeitos aleatórios genético aditivo e de ambiente permanente e classes de variância residual, com exceção da PRRC, indicando que os modelos que aplicaram polinômios de Legendre possuem menor complexidade da heterocidade e multicolinearidade (Bozdogan, 2000), o que pode ser facilmente confirmado através das menores correlações entre os coeficientes de regressão aleatória apresentados por estes modelos quando comparados com os que usaram a função de Wilmink e Ali & Schaeffer.

Pelo motivo da discordância entre os resultados dos critérios adotados quanto ao modelo mais adequado, são apresentados, na Figura 3, as estimativas de (co)variâncias e correlações entre os coeficientes de regressão obtidas pelo AS\_HE17, AS\_HO e W\_HE42. As correlações entre os regressores mais elevadas resultaram do AS\_HE17 e AS\_HO, variando de -0,99 a 0,95 e de -1,00 a 0,99 para os efeitos genético aditivo e de ambiente permanente, respectivamente, sendo esta a razão da menor PRRC encontrada para o modelo que usou a função de Ali e Schaeffer quando comparado ao que usou a de Wilmink (-0,86 a 0,70 e -0,25 a 0,14 para os efeitos genético aditivo e de ambiente permanente, respectivamente), resultando em maior dificuldade de convergência do modelo mais parametrizado.

Na proporção da variância explicada pelos autovalores estimada pelos três modelos, observou-se, que o primeiro autovalor explicou mais de 96% da variação genética e de ambiente permanente, exceto o primeiro autovalor da função de covariância de ambiente permanente obtido pelo W\_HE42. Com

base nesse resultado, um modelo menos parametrizado poderia ser proposto, no entanto, a redução de dimensionalidade decorrente da eliminação de autovalores não é indicada em todos os casos, podendo resultar num ajuste

Tabela 3 – Estimativas de variâncias (diagonal), covariâncias (abaixo da diagonal) e correlações (acima da diagonal) entre os coeficientes de regressão aleatória, autovalores ( $\lambda$ ) e porção da variância explicada pelo autovalor da matriz de coeficientes

0	1	2	3	4	$\lambda$	% da Variância
<b><sup>1</sup>W_HE42</b>						
<i>Genético Aditivo</i>						
1,97	-0,86	-0,85	-	-	16,63	97,00
-0,006	0,01	0,70	-	-	0,51	3,00
-4,63	0,01	15,16	-	-	0,00	0,00
<i>Ambiente Permanente</i>						
1,84	-0,06	-0,25	-	-	1,86	81,22
-0,04	0,25	0,14	-	-	0,26	11,35
-0,15	0,03	0,20	-	-	0,17	7,42
<b><sup>2</sup>AS_HE17</b>						
<i>Genético Aditivo</i>						
118,45	-0,99	0,94	-0,98	0,85	569,59	98,46
-201,07	348,65	-0,98	0,95	-0,81	7,77	1,34
90,37	-161,17	78,03	-0,88	0,72	0,92	0,16
-60,87	101,80	-44,69	32,78	-0,94	0,23	0,04
7,21	-11,72	4,91	-4,20	0,60	0,00	0,00
<i>Ambiente Permanente</i>						
130,78	-0,97	0,79	-0,99	0,99	540,12	96,44
-201,25	327,73	-0,91	0,96	-0,95	19,22	3,43
68,00	-123,61	56,53	-0,76	0,74	0,44	0,08
-75,23	114,90	-37,72	43,91	-0,99	0,26	0,05
11,82	-17,92	5,78	-6,91	1,09	0,00	0,00
<b><sup>3</sup>AS_HO</b>						
<i>Genético-Aditivo</i>						
67,09	-0,98	0,92	-0,97	0,74	336,87	98,05
-116,44	208,46	-0,97	0,94	-0,68	5,74	1,67
55,04	-102,26	53,10	-0,86	0,58	0,72	0,21
-30,43	51,86	-23,93	14,72	-0,89	0,22	0,06
2,55	-4,12	1,77	-1,42	0,17	0,00	0,00
<i>Ambiente Permanente</i>						
320,04	-0,99	0,92	-0,99	0,99	1379,02	98,63
-510,39	831,96	-0,96	0,98	-0,98	18,42	1,32
192,34	-325,80	137,01	-0,90	0,89	0,46	0,03
-184,20	292,92	-109,36	106,70	-1,00	0,29	0,02
28,08	-44,54	16,52	-16,29	2,49	0,00	0,00

<sup>1</sup>Função de Wilmink com 42 classes de variância residual; <sup>2</sup>Função de Ali e Schaeffer com resíduo homogêneo; e <sup>3</sup>Função e Ali Schaeffer com 17 classes de variância residual

inadequado no início e no final da trajetória (Legarra et al., 2004). Além disso, modelos que usaram a função de Wilmink e polinômios de Legendre com menor ordem de ajuste para o efeito genético aditivo e de ambiente permanente não foram indicados pelos critérios estatísticos de escolha de modelos (Tabela 2), sugerindo que estes modelos não proporcionaram melhor ajuste do que o apresentado pelo mais parametrizado, ou seja, aquele que usou a função de Ali e Schaeffer.

As variâncias fenotípicas ( $\sigma_p^2$ ) e genéticas ( $\sigma_a^2$ ) para as produções de leite no dia do controle, estimadas pelo W\_HE42, AS\_HE17 e AS\_HO, juntamente com as estimadas pelas análises unicaracterísticas são apresentadas na Figura 2. As variâncias fenotípicas ( $\sigma_p^2$ ) geradas pelos modelos de regressão aleatória foram semelhantes, com exceção da 2ª semana, em que o W\_HE42 (5,80kg<sup>2</sup>) e o AS\_HO (5,70kg<sup>2</sup>) produziram estimativas levemente mais elevadas do que o AS\_HE17 (5,02kg<sup>2</sup>). De modo geral, as variâncias fenotípicas foram altas nas duas primeiras semanas, diminuíram até a semana seis e foram, praticamente, constantes nas demais semanas, tendendo a aumentar no final da lactação.

As variâncias genéticas aditivas estimadas pelos modelos de regressão aleatória foram superiores as estimadas pelas análises unicaracterísticas, todavia, apresentaram a mesma tendência. As variâncias genéticas ( $\sigma_a^2$ ) estimadas pelos três modelos foram muito semelhantes, sendo as maiores diferenças observadas nas semanas finais da lactação, em que o AS\_HE17 e AS\_HO produziram estimativas superiores. Comportamento similar para variância genética aditiva ao longo da lactação foi verificado por Jamrozik et al (1997), estudando modelagem da trajetória da lactação de bovinos, usando a função de Wilmink e Ali e Schaeffer.

As diferenças mais evidentes nas estimativas de variância de ambiente permanente (Tabela 2) foram verificadas após a 35ª semana, em que as fornecidas pelo W\_HE42 foram superiores às dos modelos que usaram a função de Ali e Schaeffer. De modo geral, as variâncias fenotípicas, genéticas e de ambiente permanente variaram em torno de 2,30 a 5,80kg<sup>2</sup>; de 0,50 a 3,14kg<sup>2</sup> e de 0,91 a 1,87kg<sup>2</sup>, respectivamente.

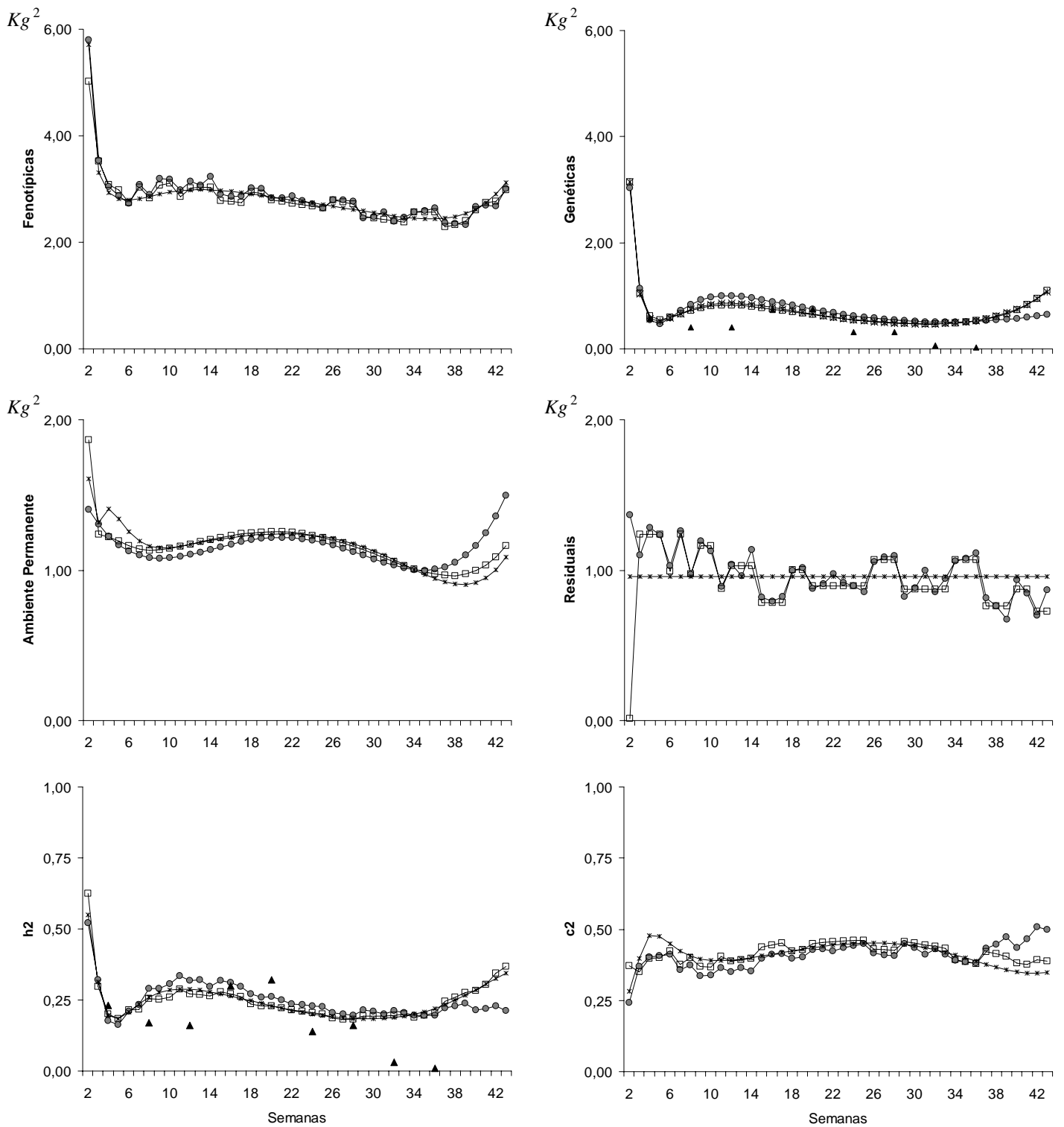


Figura 2 – Estimativas de variâncias fenotípicas, genéticas, de ambiente permanente, residuais, herdabilidades ( $h^2$ ) e das frações da variância fenotípica referente à variância de ambiente permanente ( $c^2$ ) obtidas com os modelos W\_HE42 ( $\square$ ), AS\_HO (\*), AS\_HE17 ( $\square$ ) e análise unicaracterística ( $\blacktriangle$ ).

As estimativas de variância residual foram superiores às variâncias genéticas aditivas, praticamente, ao longo de toda a lactação. Os modelos que consideraram a variância residual heterogênea ao longo da lactação (W\_HE42 e AS\_HE17) foram semelhantes, com exceção da 2ª semana, em que a estimativa do AS\_HE17 foi bastante inferior ( $0,01\text{kg}^2$ ) a do W\_HE42 ( $1,37\text{kg}^2$ ).

Ainda com relação a variância residual, ao comparar o W\_HE42 e AS\_HE17 com o AS\_HO, as diferenças mais expressivas foram visualizadas até a 11ª e a partir da 37ª semana, em que o W\_HE42 e AS\_HE17 resultaram em menores e maiores estimativas, respectivamente. Nos demais períodos as estimativas geradas pelo W\_HE42 e AS\_HE17 tenderam a ficar bem próximas da produzida pelo AS\_HO. Possivelmente, as elevadas variâncias residuais, obtidas pelo W\_HE42 e AS\_HE17, nas primeiras semanas, possam ser atribuídas a maior dificuldade de controle das influências ambientais nesta fase da lactação (Jensen, 2001). López-Romero et al. (2003) afirmaram que a heterogeneidade de variâncias residuais está relacionada com o estágio da lactação e é maior nos extremos da trajetória, devido a um conjunto de fatores não especificados no modelo, tais como: estágio de prenhez, características do período seco, condição corporal ao parto, entre outros, que fazem com que a variância residual seja altamente variável no início e no final da lactação.

De acordo com Olori et al. (1999) a suposição de homogeneidade da variância residual ao longo da lactação conduz a erros nas estimativas de variância residual no início da mesma, mas não tem significativa influência na estimativa dos demais componentes de variância, todavia, teria conseqüências nas estimativas de variância total e, conseqüentemente, na herdabilidade.

As herdabilidades estimadas pelos três modelos foram altas na 2ª semana, sendo iguais a 0,52 para o W\_HE42, 0,63 para o AS\_HE17 e 0,55 para o AS\_HO. Depois da segunda semana, as herdabilidades diminuíram de forma súbita, alcançando seu menor valor na 5ª semana ( $\cong 0,16$ ). Após a 6ª semana (início do pico de produção), as estimativas aumentaram gradativamente até alcançar, por volta da 11ª semana, o valor igual a 0,33 (W\_HE42) e 0,29 (AS\_HE17 e AS\_HO), sendo sempre levemente superiores para o W\_HE42. Nas semanas subseqüentes as herdabilidades estimadas se mantiveram praticamente constantes, voltando a diminuir após a semana 17.

Como observado nas variâncias genéticas, houve aumento nas herdabilidades a partir da 35ª semana, principalmente com o AS\_HE17 e AS\_HO.

De acordo com Costa et al. (2002) a superestimação da herdabilidade tem sido atribuída principalmente a problemas de ajuste da produção de leite no dia do controle, por modelos de regressão aleatória. Variação nas herdabilidades estimadas ao longo da lactação de bovinos utilizando modelos de regressão aleatória tem sido relatada por diversos autores, entre eles: Cobuci et al. (2005) que também verificaram maiores herdabilidades no início e no final da lactação; Jakobsen et al (2002) e Melo et al (2007) que, ao contrário do relatado no presente estudo, observaram maiores herdabilidades no período intermediário da lactação. Em estudos com búfalos relataram-se herdabilidades para a produção de leite no dia do controle usando modelos de dimensão finita oscilando de 0,01 a 0,20 (Hurtado-Lugo et al., 2006) e de 0,09 a 0,24 (Aspilcueta-Borquis et al., 2007) com comportamento similar ao verificado no presente estudo.

Com relação às frações da variância fenotípica referente à variância de ambiente permanente, estas foram superiores às herdabilidades durante toda a lactação, exceto na segunda semana, corroborando com os resultados encontrados por El Faro e Albuquerque (2003). Assim como nas estimativas de herdabilidades, maiores diferenças entre os modelos de regressão aleatória foram verificadas nas semanas iniciais e finais. De acordo com Meyer (2005) as diferenças entre os modelos são maiores nas idades com menos registros. Como pode ser observado na Figura 1, nas semanas 2 e da 39 a 43, a quantidade de registros foi inferior ao das demais semanas (<300 registros).

As estimativas de correlações fenotípicas, genéticas e de ambiente permanente obtidas pelo AS\_HE17, AS\_HO e W\_HE42, apresentadas na Figura 3, exibiram planos muito similares. Foram observadas estimativas de correlações fenotípicas e genéticas negativas, biologicamente inesperadas, quando envolviam controles realizados nas semanas iniciais da lactação (2ª a 4ª). Nas demais semanas as estimativas de correlações fenotípicas e genéticas foram menores quanto mais distantes os controles leiteiros, sugerindo que a seleção de animais, visando aumento na produção de leite em qualquer ponto da curva de lactação, resulta em resposta positiva nas demais semanas do controle.

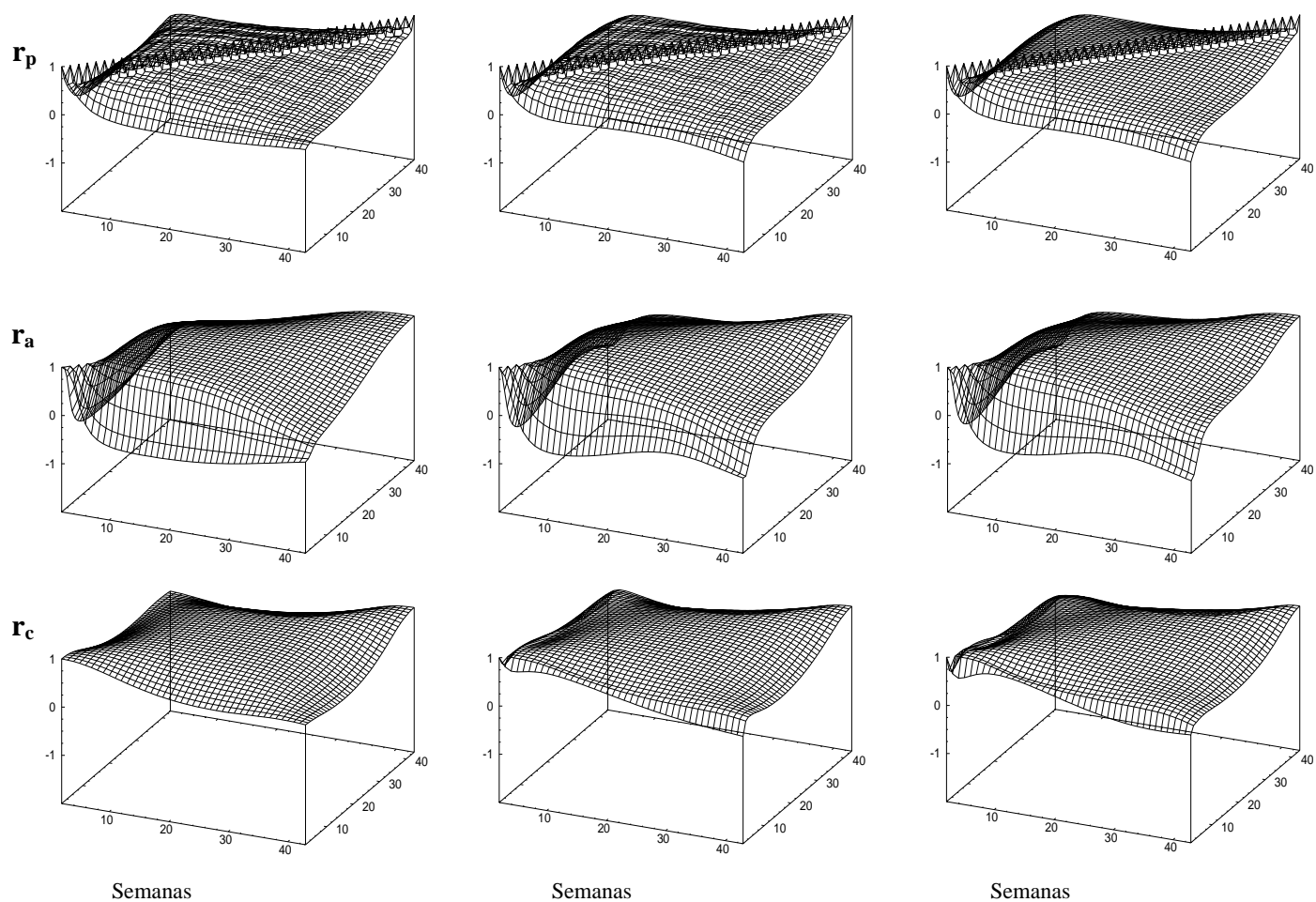


Figura 3 – Correlações fenotípicas ( $r_p$ ), genéticas ( $r_g$ ) e de ambiente permanente ( $r_c$ ), estimadas entre as produções de leite semanais, de acordo com o modelo que usou a função de Wilink com 42 classes de variância residual (esquerda) e a função de Ali e Schaeffer com resíduo homogêneo (a direita) e com dezessete classes de variância residual (meio).

Quanto às correlações de ambiente permanente produzidas pelos três modelos, as estimativas foram semelhantes, variando de 0,34 a 0,99, de 0,22 a 0,99 e de 0,21 a 0,99 para o W\_HE42, AS\_HE17 e AS\_HO, respectivamente. A partir da 3ª semana, estas foram maiores entre os controles adjacentes e menores quanto mais distantes estes controles foram realizados.

Resultados semelhantes, das estimativas de correlações, foram relatados por Cobuci et al. (2005), Costa et al. (2005) e Breda et al. (2006). Brotherstone et al. (2000) verificaram que as curvas paramétricas parecem ser mais eficientes em ajustar a trajetória da lactação que os polinômios ortogonais de Legendre com o mesmo número de parâmetros. Contudo, curvas paramétricas

exibem inabilidade em modelar a associação entre as produções iniciais e as finais, resultando em correlações genéticas negativas.

Segundo López-Romero & Carabaño (2003) a dificuldade de modelagem das variâncias no início e no final da lactação é esperado, pois, neste período, além da escassez de informações, ocorre maior influência ambiental do que as comuns à manutenção da produção na fase intermediária da lactação.

Enfim, como anteriormente discutido, o W\_HE42 tendeu a superestimar a variância de ambiente permanente no final da lactação, possivelmente a ordem de ajuste do efeito de ambiente permanente não foi grande o suficiente para adaptar as mudanças das variâncias que ocorrem ao longo da lactação. Adicionalmente, verifica-se que as estimativas de variâncias e herdabilidades produzidas pelos três modelos foram bastante semelhantes, não justificando o uso de modelos mais parametrizados (W\_HE42 e AS\_HE17) para descrever as mudanças nas variâncias que ocorrem ao longo da lactação. Dessa forma, o AS\_HO, apontado como o mais adequado pelo BIC (Tabela 2), é o indicado para modelar a produção de leite no dia do controle de búfalas em programas de avaliação genética.

### **Conclusões**

Os modelos que empregaram as funções paramétricas diferiram no agrupamento de classes de variâncias residuais, sugerindo que diferentes modelos requerem diferentes estruturas de variância residual.

O critério de informação Bayesiano é o mais indicado para escolha de modelos de regressão aleatória.

O modelo que considerou a função de Ali e Schaeffer, com resíduo homogêneo é o recomendado para modelar a produção de leite no dia do controle de búfalas, em programas de avaliação genética.

Os modelos de regressão aleatória foram adequados em descrever a variação genética da produção de leite de búfalas, com exceção das semanas iniciais e finais da lactação.

Para o rebanho de búfalos em estudo, detectou-se variabilidade genética para a produção de leite no dia do controle, com estimativas de herdabilidades de magnitude moderada ao longo de toda a trajetória da lactação, o que

viabiliza a predição de ganhos genéticos e o uso da seleção para indicar os animais geneticamente superiores.

### **Referências Bibliográficas**

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY, 2., 1973, Budapest. **Proceedings...**Budapest: Academiai Kiado, 1973, p.267-281.

ALI, E.; SCHAEFFER, L.R. Accounting for covariances among test day milk yields in dairy cows. **Canadian Journal Animal Science**, v.67, n.3, p.637-644, 1987.

ASPILCUETA-BORQUIS, R.; TANAKA, A.; ALBUQUERQUE, L.G. et al. Genetic parameters for milk, fat and protein yield analyzed by test day models for murrah buffaloes. In: WORLD BUFFALO CONGRESS, 8., 2007, Caserta:Itália. **Proceeding...**Caserta:Itália, 2007.

BOZDOGAN, H. Akaike's information criterion and recent developments in information complexity. **Journal of Matematical Psychology**, v.44, p.62-91, 2000.

BREDA, F.C.; ALBUQUERQUE, L.G.; YAMAKI, M. et al. Estimação de parâmetros genéticos para a produção de leite de cabras da raça Alpina. **Revista Brasileira de Zootecnia**, v.35, n.2, p.396-404, 2006.

BROTHERSTONE, S.; WHITE, I.M.S.; MEYER, K. Genetic modeling of daily milk yield using orthogonal polynomials and parametric curves. **Journal Animal Science**, v.70, p.407-415, 2000.

COBUCI, J.A.; EUCLYDES, R.F.E.; LOPES, P.S. et al. Estimation of genetic parameters for test-day milk yield in Holstein cows using a random regression model. **Genetics and Molecular Biology**, v.28, n.1, p.75-83, 2005.

- COSTA, C.N.; MELO, C.M.R.; MACHADO, C.H.C, et al. Avaliação de funções polinomiais para ajuste da produção de leite no dia do controle de primeiras lactações de vacas Gir com modelo de regressão aleatória. In: REUNIÃO DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 39., 2002, Recife. **Anais...Recife: 2002. CD-ROM.**
- COSTA, C.N.; MELO, C.M.R.; MACHADO, C.H.C. et al. Parâmetros genéticos para a produção de leite de controles individuais de vacas da raça Gir estimados com modelos de repetibilidade e regressão aleatória. **Revista Brasileira de Zootecnia**, v.34, n.5, p.1520-1531, 2005.
- EL FARO, L.; ALBUQUERQUE, L.G. Utilização de modelos de regressão aleatória para produção de leite no dia do controle, com diferentes estruturas de variâncias residuais. **Revista Brasileira de Zootecnia**, v.32, n.5, p.1104-1113, 2003.
- FAO – Food and Agriculture Organization, <http://faostat.fao.org>. Acessado em janeiro de 2008.
- HURTADO-LUGO, N.; CERÓN-MUÑOZ, M.; GUTIÉRREZ-VALENCIA, A. Estimación de parámetros genéticos para la producción de leche en el día del control en búfalos de la Costa Atlántica Colombiana. **Livestock Research for Rural Development**, v.18, n.3, 2006.
- JAKOBSEN, J.H.; MADSEN, P.; JENSEN, J.; et al. Genetic parameters for milk production and persistency for Danish Holsteins estimated in random regression models using REML. **Journal of Dairy Science**, v.85, n.6, p.1607-1616, 2002.
- JAMROZIK, J.; SCHAEFFER, L.R. Estimates of genetic parameters for a test day model with random regression for yield traits of first lactation Holstein. **Journal of Dairy Science**, v.80, n.4, p.762-770, 1997.
- JAMROZIK, J.; KISTEMAKER, G.J.; DEKKERS, J.C.M. et al. Comparison of possible covariates for use in a random regression model for analyses of test day yields. **Journal of Dairy Science**, v.80, p.2550-2556, 1997.

- JAMROZIK, J.; SCHAEFFER, L.R. Bayesian comparison of random regression models for test-day yields in dairy cattle. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier. **Proceedings...** Montpellier: 2002 (CD\_ROM).
- JENSEN, J. Genetic evaluation of dairy cattle using test-day models. **Journal Dairy Science**, v.84, p.2803-2812, 2001.
- LEGARRA, A.; MISZTAL, I.; BERTRAND, J.K. Constructing covariance functions for random regression models for growth in Gelbvieh beef cattle. **Journal of Animal Science**, v.82, p.1564-1571, 2004.
- LÓPEZ-ROMERO, P.; CARABAÑO, M.J. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein Friesian cattle. **Livestock Production Science**, v.82, p.81-96, 2003.
- LÓPEZ-ROMERO, P.; REKAYA, R.; CARABAÑO, M.J. Assessment of homogeneity vs. heterogeneity of residual variance in random regression test-day models in a Bayesian analysis. **Journal Dairy Science**, v.86, p.3374-3385, 2003.
- LIU, Y.X.; ZHANG, J.; SCHAEFFER, L.R.; YANG, R.Q.; ZHANG, W.L. Short Communication: Optimal random regression models for milk production in dairy cattle. **Journal Dairy Science**, v. 89, p.2233-2235, 2006.
- MELO, C.M.R.; PACKER, I.U, COSTA, C.N.; MACHADO, P.F. Genetic parameters for test day milk yields of first lactation Holstein cows by random regression models. **Animal**, p.325-334, 2007.
- MEYER, K. DXMRR – A program to estimate covariance functions for longitudinal data by REML. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 6., 1998a, Armidale. **Proceedings...**Armidale: University of New England, 1998a. CD-ROM.
- MEYER, K. Estimating covariance functions for longitudinal data using a random regression model. **Genetic Selection Evolution**, v.30, p.221-240, 1998b.

- MEYER, K. Estimates of genetic covariance function for growth of Angus cattle. **Journal Animal Breeding Genetics**, v.122, p. 73-85, 2005.
- NUNEZ-ANTÓN, V.N.; ZIMMERMAN, D.L. Modelling nonstationary longitudinal data. **Biometrics** , v.56, p.699-705, 2000.
- OLORI, V.E.; HILL, W.G.; McGUIRK, S. et al. Estimating variance components for test day milk records by restricted maximum likelihood with random regression animal model. **Livestock Production Science**, v.61, p.53-63, 1999.
- SCHAWARZ, G. Estimating the dimension of the model. **Ann. Stat.** n.6, p. 127-132, 1998.
- TONHATI, H.; MENDONZA-SANCHES,G.; SESANA, B.C.; ALBUQUERQUE, L.G. Programa de mejoramento genético de búfalos lechero em el Brasil. In: SIMPÓSIO DE BÚFALOS DAS AMÉRICAS, 3., 2006, Medellin. **Proceeding...**Medelin: Associação Brasileira de Criadores de Búfalos, 2006. p. 115-122.
- WILMINK, J.B.M. Efficiency of selection for different cumulative milk, fat and protein yields in first lactation. **Livestock Production Science**, v.17, n.3, p.211-224, 1987.

## Capítulo 2

### Estimação de Parâmetros Genéticos para Produção de Leite de Búfalos da Raça Murrah Usando Inferência Bayesiana

**Resumo** – Modelos de regressão aleatória foram utilizados para estimar parâmetros genéticos da produção de leite no dia do controle em búfalos da raça Murrah, por meio da metodologia Bayesiana. O conjunto de dados incluiu 17.935 registros da produção de leite no dia do controle, provenientes da primeira lactação de 1.433 búfalas. Foram testados 12 modelos que usaram diferentes combinações dos polinômios ortogonais de ordens três, quatro, cinco, seis e sete para os efeitos genético aditivo e de ambiente permanente. Todos os modelos incluíram os efeitos fixos de grupo contemporâneo, número de ordenhas e a covariável idade da vaca ao parto. Além disso, a variância residual foi considerada heterogênea com seis classes. Os modelos foram comparados pelas estimativas do quadrado médio do resíduo e pelas estimativas dos componentes de variâncias, herdabilidades, correlações, autovalores e autofunções. Os resultados obtidos indicam que as mudanças nas ordens de ajuste dos efeitos aleatórios genético aditivo e de ambiente permanente influenciam a predição de parâmetros genéticos. Adicionalmente, os modelos mais parametrizados apresentam dificuldade de convergência para a distribuição estacionária de equilíbrio. O modelo que utilizou ordem de ajuste para o efeito genético aditivo e de ambiente permanente igual quatro e seis, respectivamente, pode ser utilizado para descrever as mudanças que ocorrem nas variâncias ao longo da curva de lactação de búfalos.

Palavras-chave: produção de leite, parâmetros genéticos, inferência Bayesiana

## Chapter 2

### **Estimation of Genetic Parameters for Milk Yield of Buffaloes Murrah Breed Using Bayesian Inference**

**Abstract** – Random regression models were used to estimate genetic parameters for test-day milk yield of Murrah breed, through Bayesian methods. Data comprised 17,935 test-day milk records of 1,433 buffaloes. Twelve models were tested using different combinations of three, fourth, fifth, sixth and seventh orders of the orthogonal polynomial for additive genetic and permanent environmental effects. All models included fixed effects of contemporary group, milking number and age of buffaloes as covariate. Besides, the residual variances were considered heterogeneous with six classes. Models were compared by residual mean square errors and by variance components, heritabilities, correlations, eigenvalues and eigenfunctions estimates. Results indicated that changes in the adjustment orders for genetic-addictive and permanent environmental random effects influenced the prediction of genetic parameters. Additionally, models with higher number of parameters present difficulty to converge for equilibrium stationary distribution. The model that used the fourth and sixth order polynomial for the additive genetic and permanent environmental effects, respectively, is appropriate to describe the changes that happen in the variances along buffaloes lactation curve.

Key Words: genetic parameters, inference Bayesian, milk yield

## Introdução

A população mundial estimada de búfalos é de 177 milhões (13% do rebanho bovino mundial), sendo a Índia possuidora do maior número de cabeças (57%) e da maior produção de leite da espécie (65%) no mundo (FAO, 2008). No entanto, é na Itália que se encontra o mais moderno programa de melhoramento genético de búfalos, onde a seleção iniciou em 1980 e, aproximadamente, oito ciclos de teste de progênie foram executados.

No Brasil não houve um processo intensivo de melhoramento genético de búfalos, como observado em bovinos, isso se deve ao alto custo para a realização do controle leiteiro oficial e limitações de infra-estrutura para a realização dos registros de produção e inseminação artificial (Peeva, 2002). Em 2005 o Ministério da Agricultura e a Associação Brasileira de Criadores de Búfalos implementaram o “Programa de Melhoramento Genético de Búfalos Leiteiros no Brasil”, com o propósito principal de avaliar e selecionar genótipos superiores para a produção de leite e constituintes, mediante estruturação e execução de um programa utilizando teste de progênie (Tonhati et al., 2006). Todavia, o programa enfrenta as dificuldades de se trabalhar com um banco de dados ainda pequeno, visto que a melhoria da acurácia do valor genético predito depende da manipulação de grande quantidade de informação de qualidade; remoção de fatores de variação ambiental; incorporação de indivíduos aparentados e do método utilizado para a predição dos valores genéticos.

Sobre este aspecto, o teorema de Bayes fornece uma solução para o problema de tamanhos finitos de amostras. A forma usual do teorema de Bayes é dada por:  $f(\theta|y) \propto f(y|\theta)f(\theta)$ , em que  $f(\theta|y)$  é a função densidade de probabilidade *a posteriori*, que reúne em si, o estado de incerteza sobre  $\theta$ , dado o conhecimento prévio sobre os parâmetros e a informação contida no vetor de observações;  $f(y|\theta)$  é a função densidade de probabilidade de  $y$  condicional a  $\theta$ , que representa a verossimilhança, ou seja, a quantidade de informação contida em  $y$  para o conhecimento de  $\theta$ ; e  $f(\theta)$  é a função densidade de probabilidade *a priori* de  $\theta$ , que reflete a relativa incerteza sobre os possíveis valores de  $\theta$ , antes do vetor  $y$  ter sido observado.

Para cada conjunto de dados, grande ou pequeno, existe uma distribuição a *posteriori* exata da qual se pode realizar inferências. Quando se analisa um conjunto de dados com grande volume, a informação a *priori* tende a ser subjugada pela função de verossimilhança no estabelecimento da *posteriori*. Neste caso, as estimativas a respeito dos parâmetros em  $\theta$ , são próximas àquelas obtidas por métodos frequentistas, o que pode não acontecer quando o tamanho amostral é finito, pois a estimação de Máxima Verossimilhança, somente apresenta propriedades bem definidas quando as amostras são grandes (Gianola & Fernando, 1986).

No entanto, para se fazer qualquer inferência a respeito de  $\theta$  é necessário integrar  $f(\theta|y)$ , procedimento este difícil de ser executado, o que inviabilizou por muito tempo a aplicação da inferência Bayesiana. A metodologia da Cadeia de Markov Monte Carlo ou MCMC (*Markov Chain Monte Carlo*), que usa a amostragem Gibbs, permite fazer inferência a respeito de  $f(\theta|y)$  sem a necessidade de integração. A amostragem de Gibbs é uma técnica de integração numérica por simulação, muito usual em situações em que a integração analítica completa é impossível.

Neste trabalho, objetivou-se estimar, por meio de modelos de regressão aleatória, funções de covariância para os efeitos genético aditivo e de ambiente permanente, descritas por polinômios ortogonais de Legendre, e parâmetros genéticos para as produções semanais de leite de búfalas da raça Murrah, utilizando inferência Bayesiana.

### **Material e Métodos**

Foram analisados 17.935 controles mensais de produção de leite de 1.433 primeiras lactações de búfalas da raça Murrah, provenientes de 11 rebanhos localizados no estado de São Paulo, com partos registrados no período de 1985 a 2006. Os rebanhos foram supervisionados pelos técnicos do programa de controle leiteiro de búfalos do Departamento de Zootecnia da Faculdade de Ciências Agrárias e Veterinária (UNESP, Jaboticabal).

Para o ajuste dos modelos de regressão aleatória, as produções de leite no dia do controle da primeira lactação de búfalas foram consideradas em classes semanais, sendo utilizados os registros obtidos entre a 2ª e a 43ª

semana após o parto de búfalas com idade entre 24 a 48 meses com, no mínimo, quatro controles de produção durante a lactação, sendo o primeiro registro realizado até 75 dias após o parto. Adicionalmente, foram excluídos animais com três desvios-padrão acima ou abaixo da média de produção de leite da semana de lactação. A média do número de animais dentro dos 708 grupos contemporâneos, formados pela fazenda, ano e mês do controle, foi igual a 27, sendo o número mínimo igual a cinco e o máximo igual a 66 animais. Para verificar a influência da idade da vaca ao parto, número de ordenha e da semana de lactação sobre a característica analisada foram realizadas análises preliminares, pelo método de quadrados mínimos.

As estimativas de componentes de (co)variâncias foram obtidas pela análise Bayesiana, utilizando-se o programa computacional RRGIBBS (Meyer, 2002). Este programa gera cadeias de Markov para os parâmetros de um modelo de regressão aleatória, por meio da amostragem de Gibbs. Quatro cadeias de Markov com 150.000 amostras foram geradas para cada modelo, assumindo *flat priors* para a matriz de (co)variâncias dos efeitos aleatórios genético aditivos e de ambiente permanente e para os componentes de variâncias residuais. De cada uma das quatro cadeias foram descartadas as primeiras 50.000 amostras como período de *burn-in* e a média *a posteriori* da primeira cadeia foi usada como *flat priors* para gerar a segunda cadeia e assim por diante. Para os modelos mais parametrizados a primeira cadeia de Markov foi gerada com 400.00 amostras, descartando-se as 150.000 primeiras amostras como período de *burn-in*, pois estes modelos apresentaram dificuldade de convergência para a distribuição estacionária de equilíbrio. O período de *burn-in* para obtenção das densidades marginais *a posteriori* foi determinado por análise gráfica e pelo programa GIBANAL (Van Kaam, 1998).

Em todos os modelos de regressão aleatória analisados foram considerados como efeitos sistemáticos o grupo contemporâneo (708 níveis), número de ordenhas (dois níveis), os efeitos linear e quadrático da covariável idade da búfala ao parto e a trajetória média da população modelada por polinômios ortogonais de Legendre, de ordem cúbica. A forma matricial do modelo animal, unicaracterístico, de regressão aleatória é dada por:

$$y = Xb + Za + Wc + e,$$

em que:  $y$  é o vetor das  $N$  observações, medidas em  $N_d$  animais;  $b$  é o vetor de efeitos sistemáticos e dos coeficientes da regressão fixa;  $a$  é o vetor de coeficientes de regressão aleatória genético aditivo;  $c$  é o vetor dos coeficientes de regressão aleatória de ambiente permanente;  $e$  é o vetor de efeito aleatório residual; e  $X$ ,  $Z$  e  $W$  são as matrizes de incidência dos efeitos fixos e aleatórios genético aditivo e de ambiente permanente.

Pressupõem-se que:

- $y | b, a, c, \sigma_{el}^2, \dots, \sigma_{ef}^2 \sim NMV(Xb + Za + Wc, R)$ , com  $R = \text{diag} \{ \sigma_{el}^2 \}$ , sendo essa a função de verossimilhança.
- $b \propto \text{constante}$ .
- $a | K_a \sim NMV(0, G)$ , com  $G = A \otimes K_a$ , em que  $A$  é a matriz de numeradores dos coeficientes de parentesco entre os animais e  $K_a$  é a matriz contendo as (co)variâncias entre os coeficientes de regressão aleatória genético aditivo;
- $c | K_c \sim NMV(0, C)$ , com  $C = I \otimes K_c$ , em que  $I$  é uma matriz identidade e  $K_c$  é a matriz contendo as (co)variâncias entre os coeficientes de regressão aleatória de ambiente permanente e  $NMV$  refere-se à distribuição normal multivariada;
- $K_a | V_a, S_a^2 \sim W^{-1}(v_a, v_a S_a^2)$
- $K_c | V_c, S_c^2 \sim W^{-1}(v_c, v_c S_c^2)$ , em que  $v_a, S_a^2$  e  $v_c, S_c^2$  representam o grau de confiança e os valores *a priori* para as covariâncias dos coeficientes de regressão genético aditivo e de ambiente permanente, respectivamente, e  $W^{-1}$  é a distribuição de Wishart invertida;
- $\sigma_{el}^2 | v_{el}, s_{el}^2 \sim X^{-2}(v_{el}, v_{el} s_{el}^2)$  com  $l = 1, 2, \dots, f$ , em que  $f$  é o número de classes do resíduo e  $X^{-2}$  é a distribuição Qui-quadrado invertida escalonada; e

Diferentes ordens de ajuste dos polinômios ortogonais de Legendre para a regressão dos efeitos aleatórios foram avaliadas. A notação utilizada foi

PL(ijl), em que i e j indica a ordem de ajuste para o efeito genético aditivo e de ambiente permanente, respectivamente; e l o número de classes de variância residual, como apresentado na Tabela 1. A variância residual foi considerada heterogênea, com seis classes, constituídas pelas semanas 2; 3 e 4; 5; 6 a 14; 15 a 34; e 35 a 43, seguindo a recomendação feita por Sesana (Comunicação Pessoal), num estudo sobre a estrutura da variância residual, usando polinômios ortogonais de Legendre, em que as estimativas dos componentes de co(variâncias) foram obtidas por meio da máxima verossimilhança restrita.

Tabela 1 – Ordens de ajuste dos polinômios de Legendre para as regressões aleatórias genético aditivo ( $k_a$ ) e de ambiente permanente ( $k_c$ ), número de classes de variância residual (e) e total de parâmetros a serem estimados para cada modelo (p)

<b>Modelo</b>	$k_a$	$k_c$	<b>e</b>	<b>p</b>
PL336	3	3	6	18
PL346	3	4	6	22
PL356	3	5	6	27
PL366	3	6	6	33
PL446	4	4	6	26
PL456	4	5	6	31
PL466	4	6	6	37
PL546	5	4	6	31
PL556	5	5	6	36
PL566	5	6	6	42
PL656	6	5	6	42
PL676	6	7	6	55

PL(ijl), em que i e j indicam a ordem de ajuste para o efeito genético aditivo e de ambiente permanente, respectivamente; e l é o número de classes de variância residual.

A região da distribuição a *posteriori* de alta densidade foi determinada, com 95% de credibilidade, para cada parâmetro do modelo. O Quadrado médio do resíduo e as estimativas de parâmetros genéticos foram utilizados para decidir qual o modelo foi mais adequado em modelar as mudanças nas variâncias genéticas, de ambiente permanente e residuais com o decorrer dos dias em lactação das búfalas. Também foi calculado o grau de interdependência entre os parâmetros do modelo, ou seja, a complexidade, dado pelo segundo termo da medida da informação teórica da complexidade (Bozdogan, 2000), como segue:

$$C_{Modelo} = C_{K_a} + C_{K_c}, \quad \text{sendo:} \quad C_{K_a} \text{ ou } C_{K_c} = \frac{p(K)}{2} \ln \left[ \frac{\text{traço}(K)}{p(K)} \right] - \frac{1}{2} \ln|K|;$$

$p(K)$  é o posto da matriz  $K$ . A maior simplicidade é alcançada quando  $K_a$  e  $K_c$  tendem a uma matriz identidade, sugerindo que os parâmetros são ortogonais e podem ser estimados com igual precisão.

### Análises Univariadas

Para que as análises unicaracterísticas pudessem ser comparadas com as de regressão aleatória, os dados foram extraídos do mesmo conjunto de dados. Assim, foram realizadas nove análises unicaracterísticas por meio da máxima verossimilhança restrita, considerando as idades de 7 a 60; 30 a 90; 60 a 120; 90 a 150; 120 a 180; 150 a 210; 180 a 240; 210 a 270; e 240 a 300 dias de lactação.

### **Resultados e Discussão**

Na Figura 1 são apresentados o número de observações, a média e o desvio-padrão para produção de leite referente a cada semana do controle. O período de pico de produção de leite nas búfalas ocorreu por volta da semana 11, em que os animais produziram, em média, 7,30kg de leite por controle. Após este período a produção diminuiu com o aumento dos dias em lactação.

O desvio padrão da média da produção de leite no dia do controle foi praticamente constante, com tendência a decrescer a partir da 30ª semana. No final da lactação houve uma redução no número de observações e somente as búfalas com maior persistência e com produções similares de leite permaneceram lactantes, o que resultou em diminuição do desvio-padrão. Além das semanas finais, baixo número de observações também foi verificado no início da lactação.

Os quadrados médios do resíduo (QMR) e as medidas de complexidade para todos os modelos são apresentados na Tabela 2. O grau de complexidade tendeu a ser maior quanto maior o número de parâmetros a serem estimados. Esse comportamento não foi sempre encontrado quando se compararam o PL356, PL366 e PL466 com alguns modelos mais parametrizados.

O PL676, com maior número de parâmetros a serem estimados, resultou em menor QMR. Todavia, o número de parâmetros não foi o que mais influenciou o decréscimo na variância residual. Nota-se que para o PL366, PL466, PL566, PL656 e PL676, as estimativas de QMR foram baixas e semelhantes, sendo que para estes modelos o número de parâmetros estimados variou de 33 a 55.

Ao ajustar um polinômio quadrático para o efeito genético ( $k_a = 3$ ), independente do grau de ajuste do efeito de ambiente permanente, as estimativas de variância genética foram constantes ao longo da lactação (dados não apresentados). Resultado este, biologicamente improvável, de acordo com alguns trabalhos encontrados na literatura (Jamrozik & Schaeffer, 1997; López-Romero et al., 2003; Roos et al., 2004; Assis et al., 2006). Adicionalmente, esses modelos foram os que apresentaram maiores estimativas de quadrado médio do resíduo, exceto o PL366. Aumentando a ordem de  $k_a$  para quatro, cinco e seis estimativas mais prováveis de variâncias genéticas foram geradas.

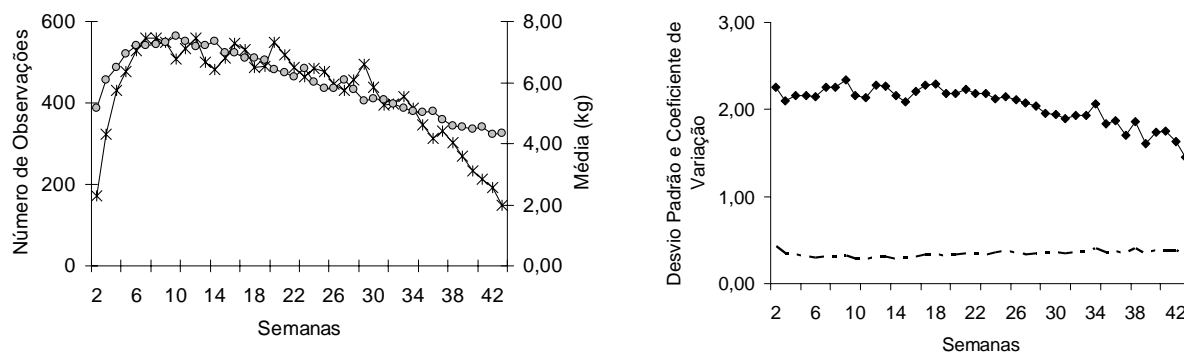


Figura 1 - Número de registros (\*), média (°), desvio-padrão(◆) e coeficiente de variação dividido por 100 (- - -) da produção de leite referente à semana do controle de primeiras lactações de búfalas da raça Murrah.

O PL656 quando comparado com o modelo com o mesmo número de parâmetros (PL566) produziu maiores estimativas de variâncias genéticas no início e no final da lactação. Além disso, as estimativas de herdabilidades produzidas pelo PL656 flutuaram consideravelmente ao longo da lactação (dados não apresentados), sugerindo que utilizar ordem de ajuste do ambiente permanente menor que a do efeito genético aditivo pode não ser adequado

para modelar as mudanças nas variâncias que ocorrem nos distintos controles, o que tem sido constatado para a produção de leite no dia do controle de bovinos (Van der Werf et al., 1998, López-Romero & Carabaño, 2003) e caprinos (Assis et al., 2006).

Tabela 2 – Modelos Ajustados com os seus respectivos número de parâmetros estimados (p), quadrado médio do resíduo (QMR) e medida da complexidade

<b>Modelo</b>	<b>p</b>	<b>QMR</b>	<b>Complexidade</b>
PL336	18	0,91	3,34
PL346	22	0,84	4,21
PL356	27	0,78	6,07
PL366	33	0,74	7,77
PL446	26	0,83	6,61
PL456	31	0,78	8,14
PL466	37	0,73	9,61
PL546	31	0,78	7,95
PL556	36	0,77	11,52
PL566	42	0,73	12,79
PL656	42	0,74	12,69
PL676	55	0,71	16,99

PL(ijl), em que i e j indicam a ordem de ajuste para o efeito genético aditivo e de ambiente permanente, respectivamente; e l é o número de classes de variância residual.

### *Componentes de Variância e Herdabilidade*

As características da distribuição a *posteriori* das variâncias e covariâncias entre os coeficientes de regressão para os efeitos aleatórios de todos os modelos foram bastante semelhantes. O desvio da média a *posteriori* em relação à respectiva mediana (percentile 50%) fornece informação sobre a simetria da distribuição marginal a *posteriori*. De modo geral, as estimativas de média e mediana para os coeficientes de aleatória genético aditivo e de ambiente permanente e para as classes de variâncias residuais foram bem próximas.

A correlação serial informa o grau de associação entre as amostras que compõe a Cadeia de Markov Monte Carlo. Neste estudo, as correlações seriais foram altas, sendo os maiores correlações observadas para os coeficientes de regressão aleatória genético aditivo, o que resultou num menor tamanho efetivo de amostra para este efeito quando comparado ao de ambiente permanente. Estes resultados foram similares ao relatado por López-Romero et al (2003).

Outro aspecto interessante de ressaltar, é que os modelos com maior número de parâmetros apresentaram dificuldade de convergência para a distribuição estacionária de equilíbrio, necessitando de um maior número de amostras e de maior período de *burn-in* na primeira vez que a Cadeia de Markov Monte Carlo foi gerada.

Pelos motivos anteriormente discutidos, apenas as médias a *posteriori* das variâncias fenotípicas ( $\sigma_p^2$ ), genéticas ( $\sigma_a^2$ ), de ambiente permanente ( $\sigma_c^2$ ) e residuais ( $\sigma_e^2$ ) estimadas para a produção de leite no dia do controle pelos modelos PL466, PL566 e PL676 são apresentadas na Figura 2. Os três modelos diferiram pouco na partição da variância fenotípica em variância genética, de ambiente permanente e residual. As maiores diferenças foram verificadas no início e no final da lactação, sendo as estimativas produzidas pelo PL566 e PL676 mais semelhantes.

As variâncias fenotípicas geradas pelos três modelos variaram de 2,43 a 5,27kg<sup>2</sup>, sendo maior na 2ª semana, diminuindo bruscamente em seguida e mantendo-se praticamente constante a partir da semana seis. Esse comportamento foi semelhante ao obtido para as variâncias de ambiente permanente, que foram superiores às estimativas de variâncias genéticas, exceto nas extremidades da curva de lactação. As variâncias genéticas estimadas pelos modelos de regressão aleatória apresentaram a mesma tendência que as estimadas pelas análises unicaracterísticas.

A menor estimativa de variância residual na segunda semana resultou do modelo mais parametrizado, ou seja, do PL676 (1,29kg<sup>2</sup>), como consequência do aumento da ordem de ajuste do efeito de ambiente permanente. López-Romero et al (2003) verificaram que, com o aumento de  $k_c$ , a variância residual foi significativamente reduzida, principalmente no início da lactação. Essa redução não foi observada quando a ordem  $k_a$  foi aumentada. Os autores ainda afirmaram que se a ordem de ajuste do efeito de ambiente permanente não é grande o suficiente para adaptar as mudanças das variâncias ao longo da lactação, a variância de ambiente temporário pode ser erroneamente superestimada. Todavia, o PL676 produziu a menor estimativa de variância residual apenas na segunda semana, nas semanas subseqüentes,

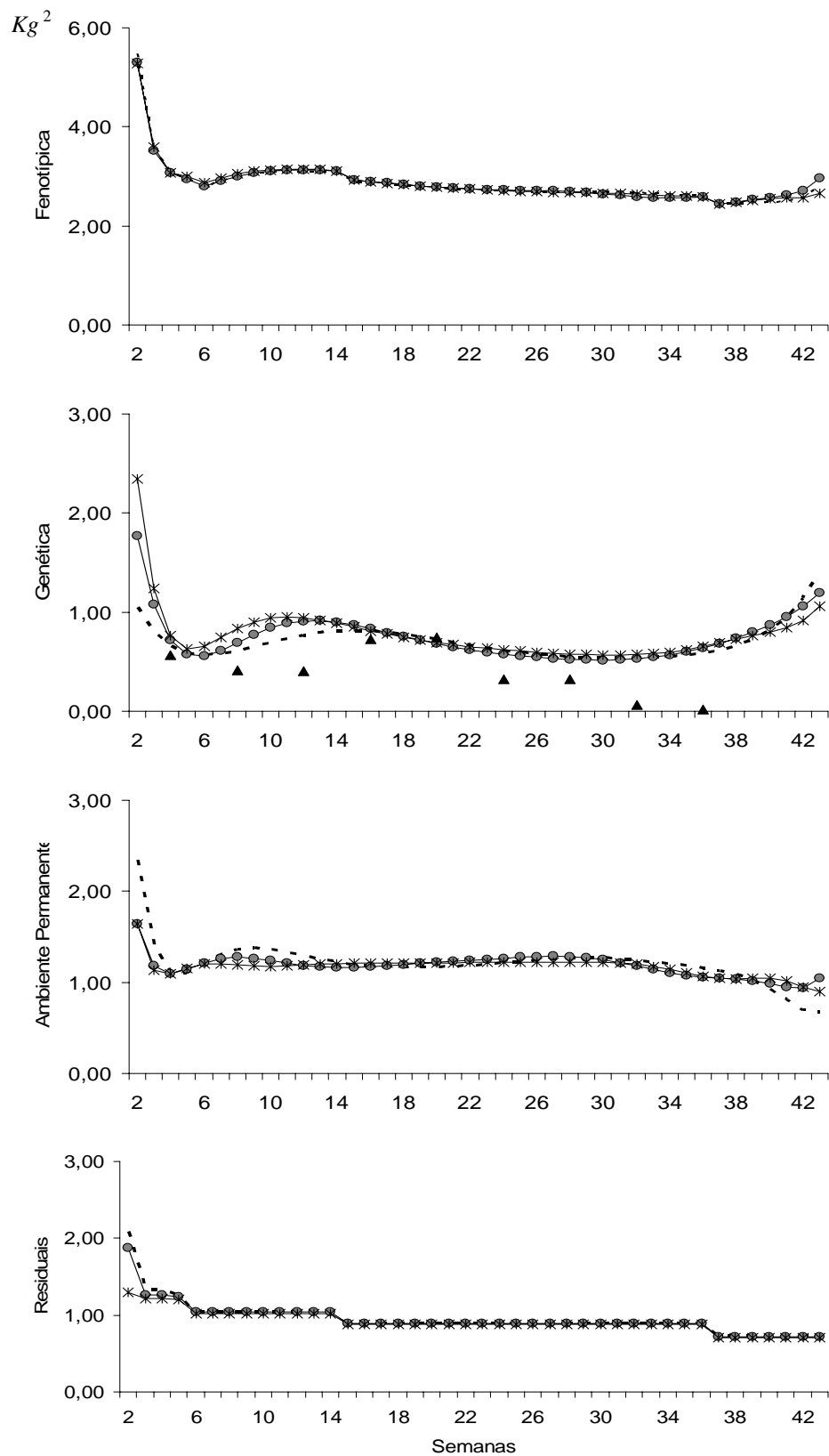


Figura 2 – Médias *a Posteriori* das variâncias fenotípicas, residuais, genéticas ( $\sigma_a^2$ ) e de ambiente permanente estimadas usando os modelos PL466 (---), PL566 (◐) e PL676 (\*). Estimativas provenientes das análises unicaracterísticas da produção de leite no dia do controle (▲).

os três modelos produziram estimativas similares, indicando que  $k_c$  igual a seis pode ser suficiente para modelar as mudanças nas variâncias de ambiente permanente ao longo da lactação.

De modo geral, as estimativas para as demais classes de variância residual foram ao redor de 1,30; 1,23; 1,03; 0,88 e 0,71kg<sup>2</sup>, portanto, a influência de fatores ambientais, não explicados pelo modelo, na variância total diminuiu com o decorrer da lactação. Brotherstone et al. (2000) ressaltaram que as estimativas de variâncias residuais têm sido maiores no início da lactação, provavelmente devido a fatores relacionados com a recuperação após o parto e adaptação à ordenha. Comportamento semelhante da variância residual foi relatado por Assis et al (2006). Jamrozik & Schaeffer (1997) observaram maiores variâncias residuais também no final da lactação.

No final da lactação, as estimativas de variâncias genéticas aumentaram, enquanto as de variâncias residuais diminuíram. Como anteriormente mencionado, no final da lactação, somente as búfalas com maior persistência e com nível de produção de leite similar tendem a permanecer lactantes. Provavelmente, estes animais sejam influenciados de maneira semelhante pelos fatores ambientais (alimentação, clima, manejo, etc.), sendo a maior parte das diferenças observadas, com relação a produção de leite, atribuída aos fatores de origem genética e não aos ambientais.

Na Figura 3 são apresentadas às médias *a posteriori* das herdabilidades ( $h^2$ ) e das frações da variância fenotípica referente à variância de ambiente permanente ( $c^2$ ) estimadas pelo PL446, PL566 e PL676. As diferenças nas herdabilidades refletiram o padrão das estimativas dos componentes de variância anteriormente observadas, sendo principalmente influenciadas pelas diferenças nas estimativas de variância genética entre os modelos.

Como nas variâncias, as diferenças nas estimativas de herdabilidades, entre os modelos, foram verificadas no início (0,34 e 0,45) e no final da lactação (0,40 e 0,50), onde as herdabilidades foram bastante elevadas, exceto as obtidas pelo PL466 no início da lactação (0,19). Após a segunda semana, os valores de herdabilidades decresceram até a 6ª semana (0,20), período caracterizado pelo início da pico de produção de leite (Figura 1). Da 6ª até a 12ª semana as herdabilidades aumentaram e depois voltaram a diminuir até a

31ª semana (0,21). Durante a fase de maior produção de leite (6ª a 14ª semana), verificou-se uma pequena diferença nas herdabilidades produzidas pelos três modelos, em que o PL566 produziu estimativas maiores do que as fornecidas pelo PL466 e menores do que as obtidas pelo PL676.

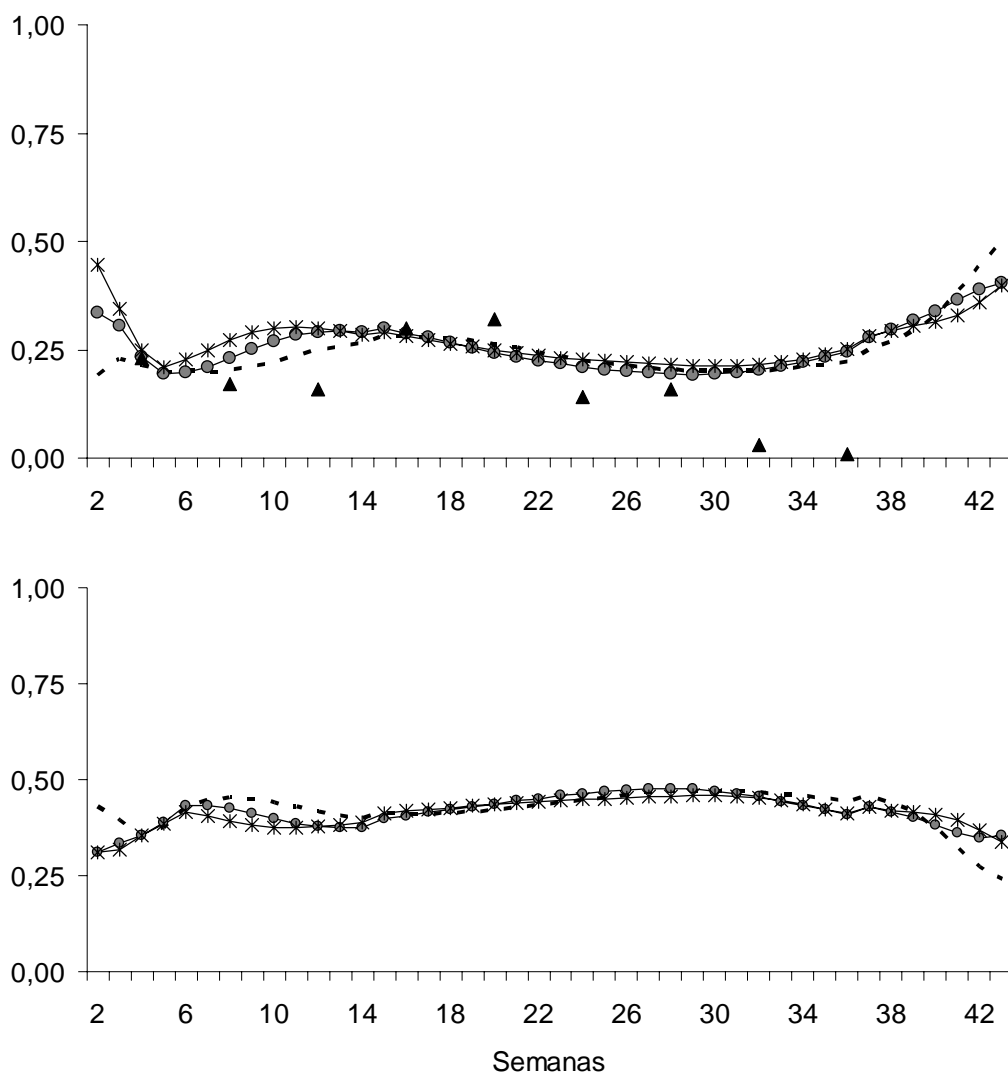


Figura 3 – Médias *a posteriori* das herdabilidades (acima) e frações da variância fenotípica referente à variância de ambiente permanente (abaixo) obtidas usando os modelos PL466 (--), PL566 (◻) e PL676 (\*). Estimativas provenientes das análises unicaracterísticas da produção de leite no dia do controle (▲)

Ao se ignorar as herdabilidades para a produção de leite nas semanas iniciais e finais da lactação, pois os três modelos pareceram superestimá-las, as estimativas variaram de 0,19 a 0,30. As maiores estimativas foram obtidas entre a 9ª e a 19ª semanas, para os três modelos de regressão aleatória. O

mesmo comportamento das estimativas de herdabilidades foi verificado para as análises unicaracterísticas até a 28ª semana da lactação.

Existem poucas referências na literatura sobre estimação de parâmetros genéticos para a produção de leite em búfalos, as quais se restringem a produção total na lactação, com estimativas de herdabilidade variando de 0,14 a 0,40 (Rosati & Van Vleck, 1998; Tonhati et al., 2000; Peeva, 2002) e a produção no dia do controle, usando análises uni ou bicaracterísticas, com herdabilidades oscilando de 0,01 a 0,24 (Hurtado-Lugo et al., 2006; Aspilcueta-Borquis et al., 2007).

Ainda na Figura 3, maior influência de ambiente permanente foi observada na 6ª (0,43) e da 22ª a 31ª semanas de lactação (0,43 a 0,48). Estas semanas também foram caracterizadas pela menor influência genética, indicando forte correlação amostral entre  $h^2$  e  $c^2$ . Meyer (2005) verificou menor variância de ambiente permanente para os modelos com maior ordem de  $k_c$ . Esta tendência não ficou evidente neste estudo, quando se aumentou  $k_c$  de seis (PL466 e PL566) para sete (PL676), sugerindo, novamente, que utilizar  $k_c$  igual a seis pode ser suficiente para modelar as mudanças nas variâncias de ambiente permanente no decorrer dos dias em lactação, sem causar grandes alterações nas estimativas de variâncias genéticas e residuais.

Tanto para as estimativas de componentes de variâncias, quanto para as herdabilidades e frações da variância fenotípica referente à variância de ambiente permanente, maiores diferenças entre os modelos de regressão aleatória foram observadas no início e final da lactação. Possivelmente, este comportamento seja explicado pelo baixo número de registros de produção de leite disponíveis, para análise, nestes períodos (Figura 1), corroborando com o relatado por Meyer (2005) que observou maiores diferenças entre os modelos nas idades com menos registros. Durante o pico de lactação (11ª semana), período de maior concentração de registros, os modelos diferiram, de forma menos expressiva que a observada nas extremidades da curva da lactação, nas estimativas de  $\sigma_c^2$  e  $\sigma_a^2$  (Figura 2), o que se refletiu nas estimativas de

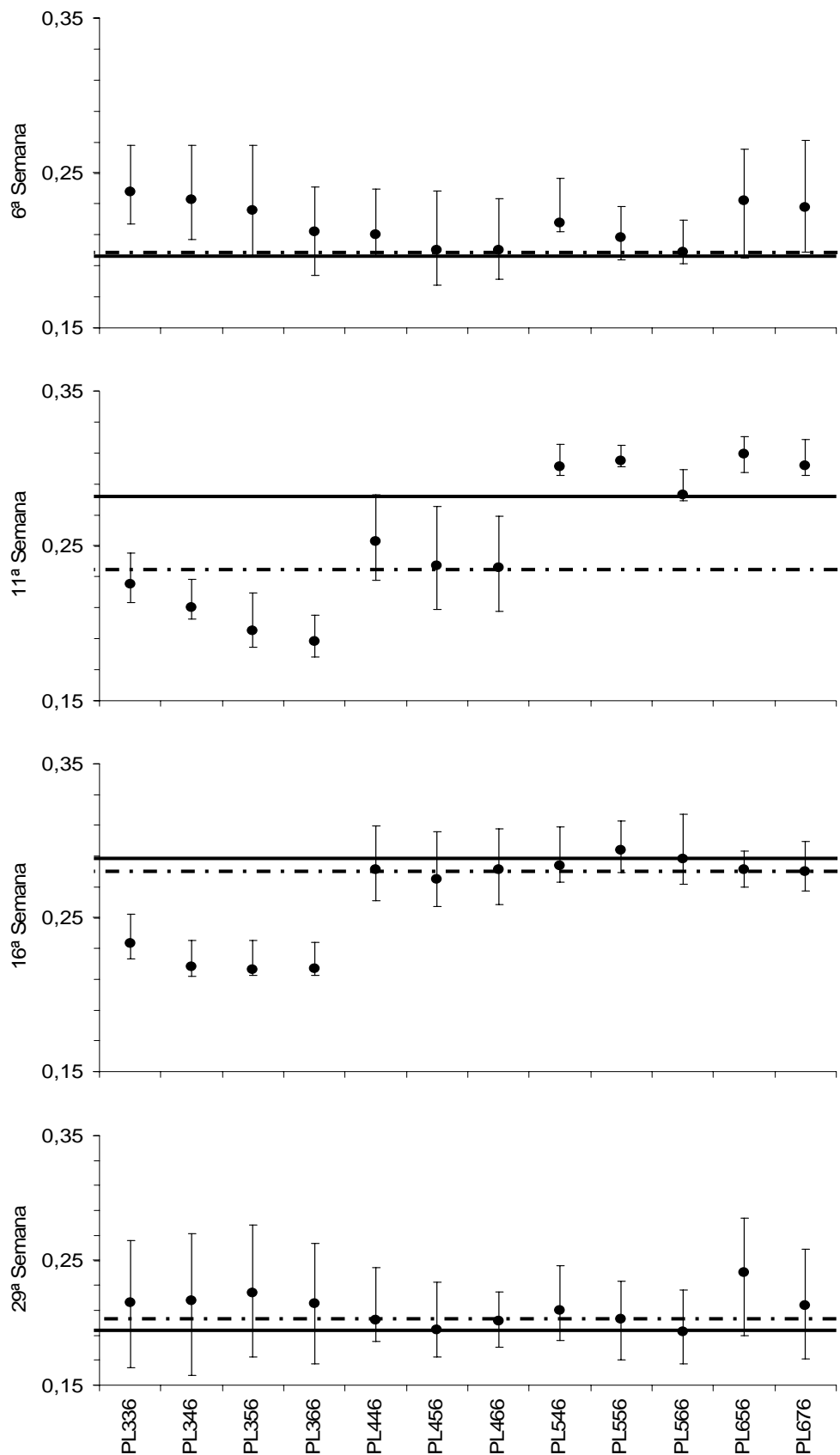


Figura 4 – Médias a posteriori das herdabilidades (•), 95% de região de maior densidade a posteriori (barras) de modelos com diferentes ordens de ajuste dos polinômios para 6ª, 11ª, 16ª e 29ª semanas de lactação. As linhas contínuas e pontilhadas fornecem a média a posteriori das herdabilidades estimadas pelos PL566 e PL466, respectivamente.

herdabilidades. Essa diferença pode ser melhor visualizada na Figura 4, que mostra as estimativas de herdabilidades e suas regiões aproximadas de confiança para as semanas 6, 11, 16 e 29.

De modo geral, nas semanas 6, 16 e 29, os modelos PL466 e PL566 geraram médias a *posteriori* das herdabilidades próximas e dentro da região de confiança dos demais modelos, exceto daqueles que utilizaram  $k_a$  igual a três nas semanas seis e 16. Na 11ª semana o PL566 resultou em menor região de confiança do que o PL466, o que sugere maior credibilidade na estimativa fornecida pelo PL566 na 11ª semana do controle.

### *Correlações*

A ordem de ajuste para os efeitos aleatórios também afetaram as estimativas de correlações genéticas, de ambiente permanente e fenotípicas como ilustrado na Figura 5. Os três modelos de regressão aleatória produziram correlações genéticas negativas entre as semanas iniciais e finais da lactação. Porém, o PL566 e PL676 geraram correlações genéticas negativas entre a 2ª e as semanas do meio da lactação (a partir da 9ª semana). A grande quantidade de correlações genéticas negativas observadas na segunda semana, para o PL566 e PL676, pode ser atribuído à maior inabilidade de polinômios de altas ordens em modelar as extremidades da curva de lactação e, também, por que nesta fase a búfala está sofrendo com o estresse do parto e com o déficit energético, como relatado por Bignardi (2006) em bovinos. Esses resultados seguem a mesma tendência descrita na literatura por outros pesquisadores (Rekaya et al., 1999; Brotherstone et al., 2000; e Lopez-Romero & Carabaño, 2003; Bignardi, 2006).

De modo geral, exceto nas primeiras semanas, as estimativas de correlações genéticas, de ambiente permanente e fenotípica decresceram quanto maior o período entre as semanas do controle. Este comportamento pode ser melhor visualizado na Figura 6, que mostra as correlações genéticas e de ambiente permanente entre a produção de leite na semana 11 (pico de produção) e as demais semanas de controle. As correlações genéticas produzidas pelo PL466 se mantiveram altas até a 25ª semana, decrescendo,

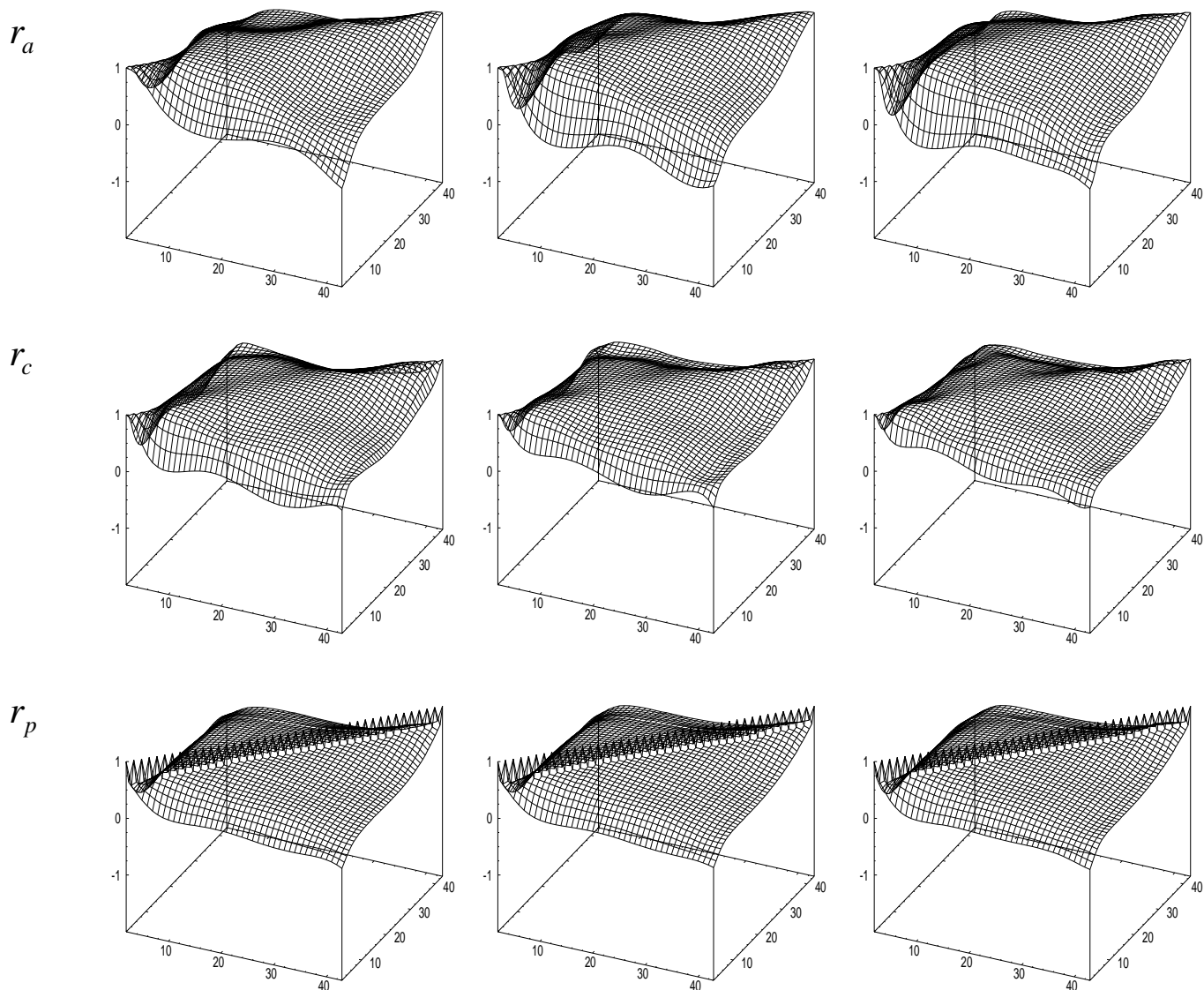


Figura 5 – Médias *a posteriori* das correlações genéticas, de ambiente permanente e fenotípicas para obtidas usando os modelos PL466 (esquerda), PL566 (meio) e PL676 (direita).

de forma mais rápida, nas semanas subseqüentes. Já as estimativas de correlações genéticas obtidas pelo PL566 e PL676 começaram a decrescer lentamente a partir da 15<sup>a</sup> semana. A correlação genética entre a 11<sup>a</sup> e a 37<sup>a</sup> semana foi igual a 0,57 para o PL566, 0,46 para o PL676 e 0,41 para o PL466, sugerindo que a seleção para maiores produções de leite no início da lactação pode produzir efeitos positivos na produção de leite nas semanas finais (Olori et al., 1999 e Assis et al., 2006). Com relação às correlações de ambiente permanente, diferenças foram verificadas nas primeiras quatro semanas com

menores estimativas para o PL466. Nas demais semanas as correlações de ambiente permanente obtidas pelos três modelos foram parecidas.

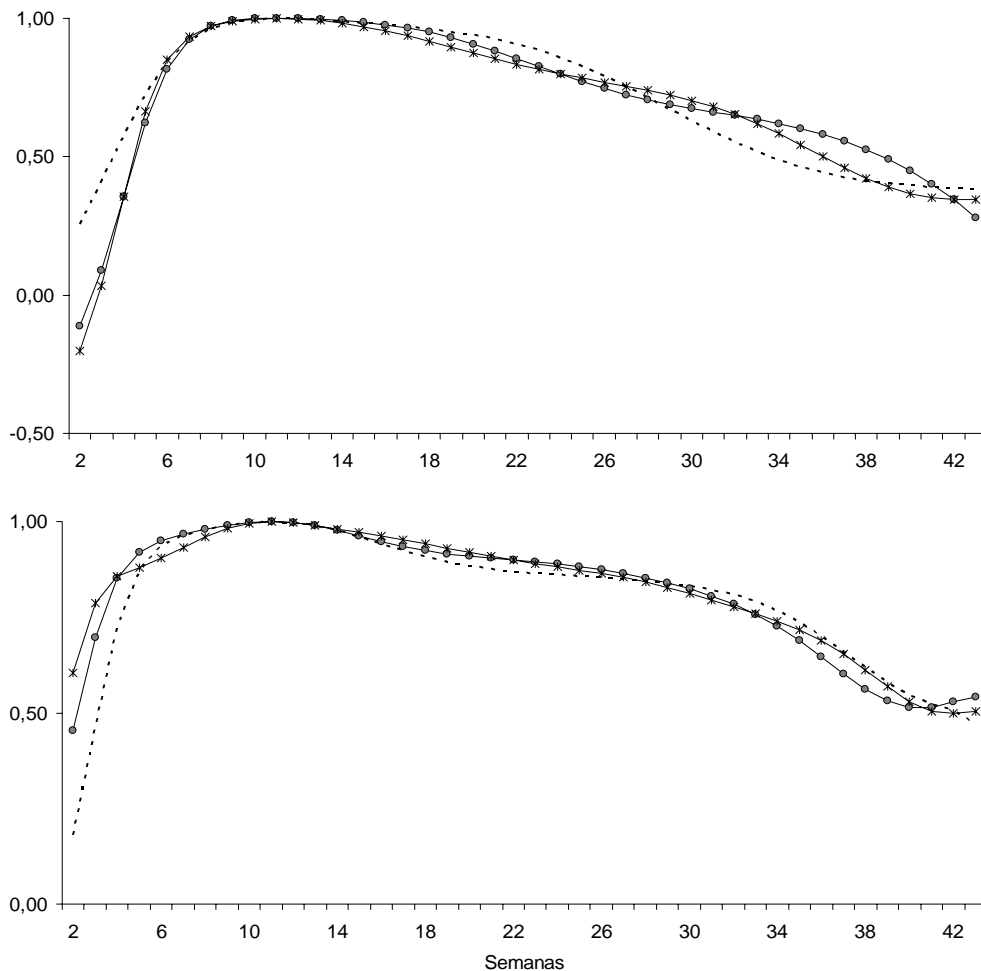


Figura 6 – Médias *a posteriori* das correlações genéticas (acima) e de ambiente permanente (abaixo) para a produção de leite na 11<sup>a</sup> semana com as produções das semanas 2 a 43, obtidas usando os modelos PL466 (--) PL566 (°) e PL676 (\*).

#### *Autovalores e Autofunções*

As estimativas dos dois primeiros autovalores das funções de covariância genética e de ambiente permanente e suas regiões de confiança são sumarizadas na Figura 7. As estimativas obtidas pelos três modelos foram praticamente as mesmas, em que os dois primeiros autovalores da matriz de coeficientes da função de covariância genética aditiva responderam por 88, 87 e 85% da variação genética total para o PL466, PL566 e PL676, respectivamente e, os dois primeiros autovalores da matriz de coeficientes da função de covariância de ambiente permanente foram responsáveis por,

aproximadamente, 89% da variação de ambiente permanente para os três modelos.

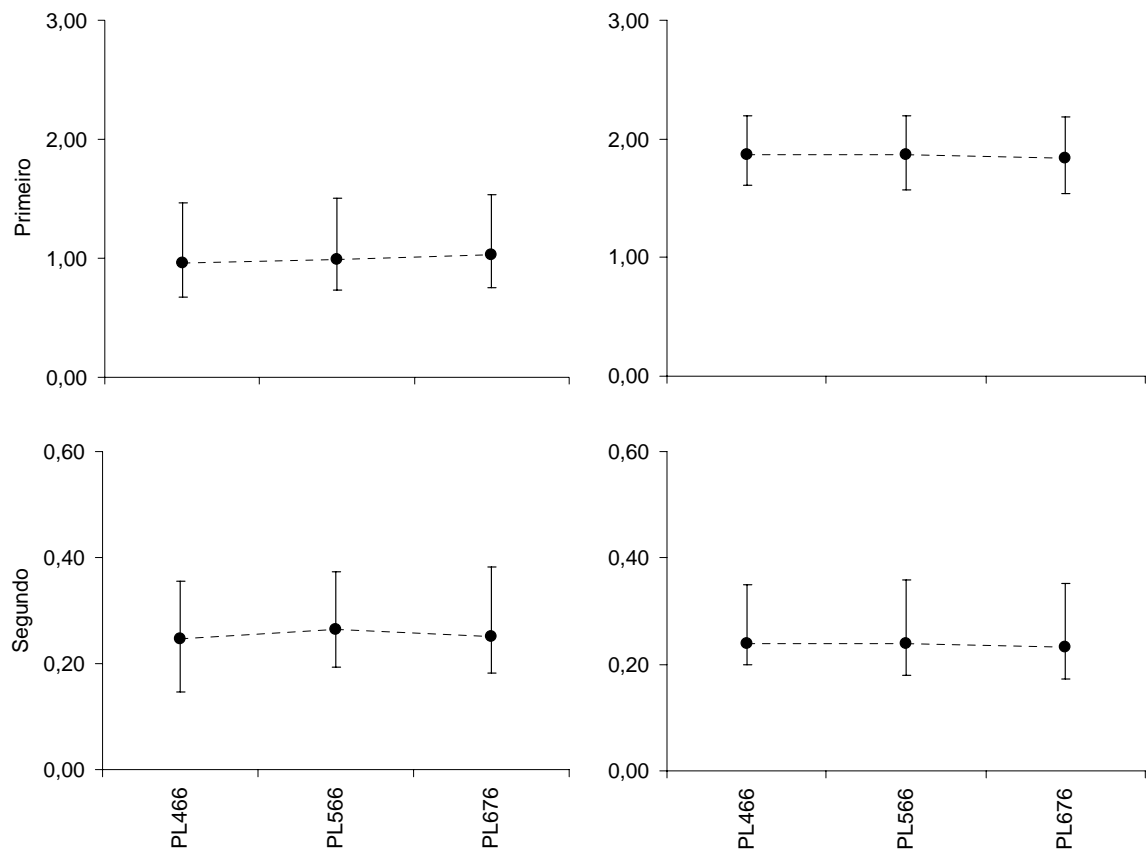


Figura 7 – Estimativas dos dois primeiros autovalores (•), 95% de região de maior densidade a *posteriori* (barras), para função de covariância genética (esquerda) e de ambiente permanente (direita) dos modelos PL466, PL566 e PL676.

As duas principais autofunções para o efeito genético aditivo, referentes aos modelos PL466, PL566 e PL676 são apresentadas na Figura 8. As autofunções são funções contínuas, cujos coeficientes são formados pelos elementos dos autovetores das matrizes de coeficientes das funções de covariância. Para cada autofunção existe um autovalor, que representa a proporção de variação total que ela é responsável. O sinal que a autofunção assume indica a direção da seleção no decorrer da lactação (Van der Werf & Schaeffer, 1997) e se há antagonismo entre produções em diferentes períodos.

A primeira autofunção obtida pelo PL466 foi positiva e tendeu a oscilar ao longo da trajetória da lactação. As primeiras autofunções produzidas pelos PL566 e PL676 foram negativas apenas na segunda semana, sendo este o provável motivo das correlações genéticas negativas observadas entre a

segunda semana e o meio da lactação. De modo geral, a primeira autofunção aumentou nas semanas iniciais da lactação e só decresceu após a ocorrência do pico de produção de leite, se mantendo praticamente constante nas semanas subseqüentes. Os valores relativamente constantes após período de pico de produção indicam que a produção de leite é controlada principalmente por genes de efeitos similares (Olori et al., 1999). A primeira autofunção positiva sugere que a produção de leite em todas as semanas pode ser mudada na mesma direção. Além disso, com aproximadamente 70% da variação genética atribuível a primeira autofunção, esta mudança é facilmente alcançada.

Os segundos autovalores das funções de covariância genética, fornecidos pelos modelos PL466, PL566 e PL676, foi responsável por 17, 18 e 16% da soma dos autovalores, respectivamente. A autofunção correspondente foi negativa até o meio da lactação (26ª semana) e positiva nas demais semanas, para os três modelos. Essa mudança no sinal sugere a associação com um fator, que ocasiona efeitos contrastantes da produção de leite no início e no final da lactação. A seleção em tal fator (ex: persistência) pode ser usada para mudar a forma da curva de lactação (Olori et al., 1999). Os demais autovalores juntos foram responsáveis por menos de 15% da soma total, sugerindo que o potencial de mudança no padrão da produção de leite ao longo da lactação selecionando-se a terceira, quarta, quinta e sexta autofunções é pequeno, uma vez que pouco contribuíram para a variação genética total.

Os resultados ilustram claramente os efeitos da escolha de modelos de regressão aleatória nas estimativas de parâmetros genéticos. Como apontado por Misztal et al (2000) consideráveis diferenças existem entre estimativas de parâmetros genéticos obtidas por diferentes modelos de regressão aleatória. Estas diferenças podem ser atribuídas ao reduzido número de registros de produção de leite e/ou aos modelos e funções usadas para descrever a regressão aleatória. Meyer (2005) menciona que a escolha do melhor modelo, inevitavelmente, é a ponderação entre as características do modelo que se desejam e a quantidade de informações disponíveis nos dados.

Neste sentido, o PL466, PL566 e PL676 produziram estimativas de parâmetros genéticos similares, com as maiores diferenças observadas no início e no final da lactação e uma pequena diferença nas semanas relativas a

ocorrência do pico de produção de leite. O  $k_c$  igual a sete não proporcionou mudanças nas estimativas de variâncias quando comparado com os modelos que utilizaram  $k_c$  igual a seis, não justificando, dessa forma a adoção deste modelo para descrever as mudanças nas variâncias ao longo da lactação. Adicionalmente, o PL676 e o PL566 resultaram em maior grau de interdependência entre os coeficientes de regressão aleatória genético aditivo e de ambiente permanente, resultando em maior estimativa da complexidade do modelo, como ilustrado na Tabela 2, sendo, possivelmente, esta a razão para a dificuldade de convergência para distribuição estacionária de equilíbrio observada ao processar as análises com estes modelos.

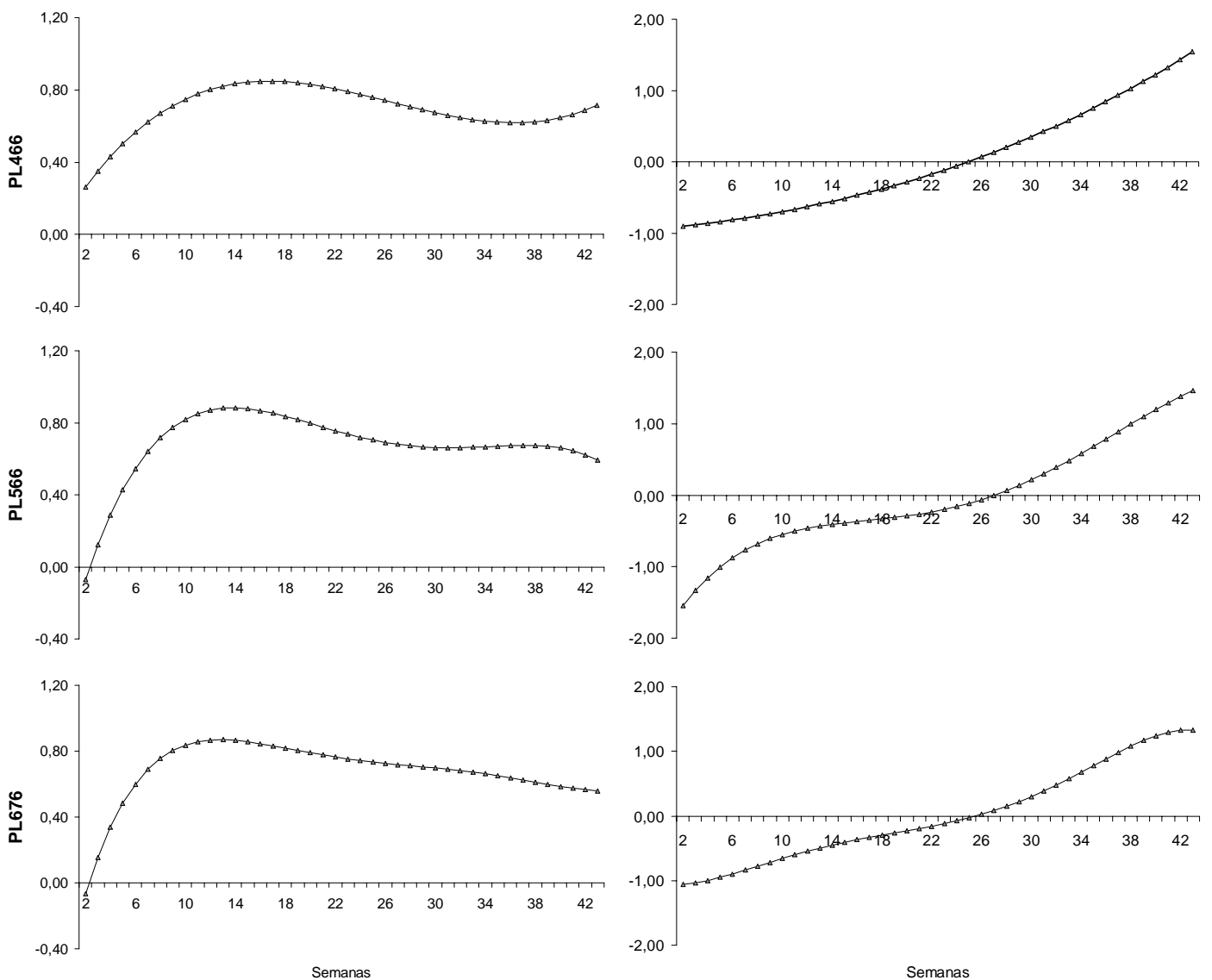


Figura 8 – Estimativas da primeira (esquerda) e segunda (direita) autofunções genética obtidas usando os modelos PL466, PL566 e PL676.

## Conclusões

Os modelos de regressão aleatória detectaram variabilidade genética e existência de correlações genéticas positivas entre as produções de leite no dia do controle, sugerindo que a seleção de búfalas geneticamente superiores para a produção de leite num determinado dia do controle pode resultar em resposta a seleção positiva em qualquer ponto da curva.

O PL466 pode ser utilizado para descrever as mudanças que ocorrem nas variâncias ao longo da curva de lactação.

Os modelos mais parametrizados apresentam dificuldade de convergência para a distribuição estacionária de equilíbrio.

## Referências Bibliográficas

- ASPILCUETA-BORQUIS, R.; TANAKA, A.; ALBUQUERQUE, L.G. et al. Genetic parameters for milk, fat and protein yield analyzed by test day models for murrah buffaloes. In: WORLD BUFFALO CONGRESS, 8., 2007, Caserta:Itália. **Proceeding...**Caserta:Itália, 2007.
- ASSIS, G.M.L.; ALBUQUERQUE, L.G.; SARMENTO, J.L.R.; et al. Estimação de parâmetros genéticos em caprinos leiteiros por meio de análise de regressão aleatória utilizando-se a amostragem de Gibbs. **Revista Brasileira de Zootecnia**, v. 35, n.3, p. 706-714, 2006.
- BIGNARDI, A.B. **Estimação de parâmetros genéticos para produção de leite de primeiras lactações de vacas da raça Holandesa usando modelos de dimensão infinita**. 2006. 86f. Dissertação (Mestrado em Zootecnia) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, 2006.
- BOZDOGAN, H. Akaike's information criterion and recent developments in information complexity. **Journal of Mathematical Psychology**, v.44, p.62-91, 2000.
- BROTHERSTONE, S.; WHITE, I.M.S.; MEYER, K. Genetic modeling of daily milk yield using orthogonal polynomials and parametric curves. **Journal Animal Science**, v.70, p.407-415, 2000.

- FAO - Food and Agriculture Organization. <http://faostat.fao.org>. Acessado em janeiro de 2008.
- GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v.63, p.217-244, 1986.
- HURTADO-LUGO, N.; CERÓN-MUÑOZ, M.; GUTIÉRREZ-VALENCIA, A. Estimación de parámetros genéticos para la producción de leche en el día del control en búfalos de la Costa Atlántica Colombiana. **Livestock Research for Rural Development**, v.18, n.3, 2006.
- JAMROZIK, J.; SCHAEFFER, L.R. Estimates of genetic parameters for a test day model with random regression for yield traits of first lactation Holstein. **Journal of Dairy Science**, v.80, n.4, p.762-770, 1997.
- LÓPEZ-ROMERO, P.; REKAYA, R.; CARABAÑO, M.J. Assessment of homogeneity vs heterogeneity of residual variance in random regression test-day models in a Bayesian analysis. **Journal Dairy Science**, v.83, p.3374-3385, 2003.
- LÓPEZ-ROMERO, P.; CARABAÑO, M.J. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein–Friesian cattle. **Livestock Production Science**, v.82, p. 81–96, 2003.
- MEYER, K. “RRGIBBS” – A program for simple random regression analyses via Gibbs Sampling. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier. **Proceedings...**Montpellier:2002. CD-ROM.
- MEYER, K. Estimates of genetic covariance function for growth of Angus cattle. **Journal Animal Breeding Genetics**, v.122, p. 73-85, 2005.
- MISZTAL, I.; STRABEL, T.; JAMROZIK, J. et al. Strategies for estimating the parameters needed for different test-day models, **Journal of Dairy Science**, v.83, p.1125-1134, 2000.

- OLORI, V.E.; HILL, W.G.; McGUIRK, S. et al. Estimating variance components for test day milk records by restricted maximum likelihood with random regression animal model. **Livestock Production Science**, v.61, p.53-63, 1999.
- PEEVA, T. Genetic improvement of buffaloes in Bulgaria. In: BUFALO SYMPOSIUM OF AMERICAS, 1, 2002, Belém: Pará. **Proceedings...** Belém: Pará, 2002.
- REKAYA, R.; CARABANO, M.J.; TORO, M.A. Random regression in animal models for test-day production in dairy cattle. Proceedings of the fifth world congress on genetics applied to livestock production. **Proceedings...** Guelph, v.18, p.443-446. 1999.
- ROSATI, A.; VAN VLECK, L.D. Estimation of genetic parameters for milk, fat, protein and mozzarella cheese production in the Italian river buffalo population. In: WORLD CONGRESS ON GENETIC APPLIED TO LIVESTOCK PRODUCTION, 6, 1998, Armidale: Australia. **Proceedings....** Armidale: Australia, 1998. 24459.pdf (CD-ROM)
- ROOS, A. P. W. ; HARBERS, A. G. F.; JONG, G. Random herd curves in a test-day model for milk, fat, and protein production of dairy cattle in the Netherlands. **Journal Dairy Science**, v.87, p.2693–2701, 2004.
- TONHATI, H.; et al. Parâmetros genéticos para a produção de leite, gordura e proteína em bubalinos. **Revista Brasileira de Zootecnia.**, v.29, n.6, p.2051-2056, 2000.
- TONHATI, H.; MENDONZA-SANCHES, G.; SESANA, R.C.; ALBUQUERQUE, L.G. Programa de melhoramento genético de búfalos lechero em el Brasil. In: SIMPÓSIO DE BÚFALOS DAS AMÉRICAS, 3., 2006, Medelin. **Proceeding...** Medelin: Associação Brasileira de Criadores de Búfalos, 2006. p. 115-122.
- Van der Werf, J.; SCHAEFFER, L. **Random Regression in Animal Breeding.** Course Notes, Ontario: University of Guelph, p.70, 1997.

Van der Werf, J.H.J., Goddard, M.E., Meyer, K.. The use of covariance function and random regressions for genetic evaluation of milk production based on test day records. **Journal Dairy Science**, v.8, p.3300–3308, 1998.

Van KAAM, J.B.C.H.M. “GIBANAL” – **Analyzing program for Markov Chain Monte Carlo Sequences (Version 2.10)**. Manual, Department of Animal Sciences, Wageningen Agricultural University, 1998.

## CONCLUSÕES GERAIS

Os modelos de regressão aleatória foram adequados em descrever a variação genética da produção de leite de búfalas, com exceção das semanas iniciais e finais da lactação, sendo recomendados para modelar a produção de leite no dia do controle de búfalas, em programas de avaliação genética.

Para o rebanho de búfalos em estudo, detectou-se variabilidade genética para a produção de leite no dia do controle, com estimativas de herdabilidades de magnitude moderada ao longo de toda a trajetória da lactação, o que viabiliza a predição de ganhos genéticos e o uso da seleção para indicar os animais geneticamente superiores.

Os modelos que empregaram as funções paramétricas e polinômios ortogonais de Legendre diferiram no agrupamento de classes de variâncias residuais, sugerindo que diferentes modelos requerem diferentes estruturas de variância residual.

O modelo que considerou a função de Ali e Schaeffer, com resíduo homogêneo (AS\_HO) e o que usou o polinômio de Legendre usando ordem de ajuste do efeito genético aditivo e de ambiente permanente igual a quatro e seis, respectivamente, com seis classes de variância residual (PL466) são indicados para descrever as mudanças que ocorrem nas variâncias ao longo da curva de lactação. Esses modelos apresentaram estimativas bastante semelhantes, exceto nas extremidades da curva de lactação.

Ao se trabalhar com inferência Bayesiana, os modelos mais parametrizados apresentam dificuldade de convergência para a distribuição estacionária de equilíbrio.