

**UNIVERSIDADE ESTADUAL PAULISTA "JULIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CAMPUS DE JABOTICABAL**

**MODELOS PARA ESTIMAÇÃO DE COMPONENTES DE
(CO)VARIÂNCIA PARA PRODUÇÃO DE LEITE NO DIA DO
CONTROLE DE VACAS DA RAÇA HOLANDESA.**

**Annaiza Braga Bignardi
Médica Veterinária**

Jaboticabal - São Paulo - Brasil
2010

**UNIVERSIDADE ESTADUAL PAULISTA "JULIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CAMPUS DE JABOTICABAL**

**MODELOS PARA ESTIMAÇÃO DE COMPONENTES DE
(CO)VARIÂNCIA PARA PRODUÇÃO DE LEITE NO DIA DO
CONTROLE DE VACAS DA RAÇA HOLANDESA.**

Annaiza Braga Bignardi

Orientadora: **Profa. Dra. Lucia Galvão de Albuquerque**

Co-orientadores: **Dra. Lenira El Faro Zadra**

Dr. Roberto Augusto de Almeida Torres Júnior

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

Jaboticabal - São Paulo – Brasil
2010

Bignardi, Annaiza Braga
B592m Modelos para estimação de componentes de (co)variância para
produção de leite no dia do controle de vacas da raça Holandesa /
Annaiza Braga Bignardi. -- Jaboticabal, 2010
xiii, 71 f. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de
Ciências Agrárias e Veterinárias, 2010

Orientador: Lucia Galvão de Albuquerque

Banca examinadora: Fabyano Fonseca e Silva, Henrique Nunes
Oliveira, Humberto Tonhati, Maria Eugênia Zerlotti Mercadante
Bibliografia

1. Vaca Holandesa-funções de covariância. 2. Leite-produção. 3.
Splines. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e
Veterinárias.

CDU 636.2:613.287



UNIVERSIDADE ESTADUAL PAULISTA
CÂMPUS DE JABOTICABAL
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS



CERTIFICADO DE APROVAÇÃO

TÍTULO: MODELOS PARA ESTIMAÇÃO DE COMPONENTES DE (CO)VARIÂNCIA PARA PRODUÇÃO DE LEITE NO DIA DO CONTROLE DE VACAS DA RAÇA HOLANDESA.

AUTORA: ANNAIZA BRAGA BIGNARDI

ORIENTADORA: Dra. LUCIA GALVAO DE ALBUQUERQUE

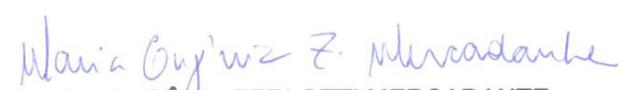
Co-Orientador(a): Dr(a). LENIRA EL FARO ZADRA

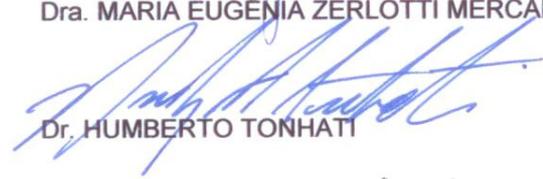
Co-Orientador(a): Dr. ROBERTO AUGUSTO DE ALMEIDA TORRES JUNIOR

Aprovado como parte das exigências para obtenção do Título de DOUTOR em GENÉTICA E MELHORAMENTO ANIMAL pela Comissão Examinadora:


Dra. LUCIA GALVAO DE ALBUQUERQUE

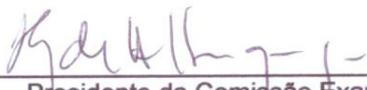

Dr. FABYANO FONSECA E SILVA


Dra. MARIA EUGÊNIA ZERLOTTI MERCADANTE


Dr. HUMBERTO TONHATI


Dr. HENRIQUE NUNES DE OLIVEIRA

Data da realização: 26 de fevereiro de 2010.



Presidente da Comissão Examinadora

Dra. LUCIA GALVAO DE ALBUQUERQUE

DADOS CURRICULARES DO AUTOR

ANNAIZA BRAGA BIGNARDI – solteira, nascida em 04 de abril de 1978, na cidade de Bebedouro-SP, filha de Antonio Carlos Bignardi e Maria de Lourdes Braga Bignardi. Iniciou em março de 1999 o curso de Medicina Veterinária na Universidade Estadual de Londrina – Londrina-PR e em 22 de novembro de 2003 obteve o título de Médica Veterinária. Em março de 2004 ingressou no Programa de pós-graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias – UNESP – Campus de Jaboticabal-SP, obtendo o grau de Mestre em 22 de fevereiro de 2006 sob orientação da Prof^a. Dr^a. Lucia Galvão de Albuquerque. Em 01 de março de 2006 iniciou o doutorado no mesmo programa de pós-graduação, obtendo o grau de Doutor em 26 de fevereiro de 2010 sob orientação da Prof^a. Dr^a. Lucia Galvão de Albuquerque. Em outubro de 2006 fez um curso de especialização em Genética e Melhoramento Animal no "Instituto Nacional de Investigación y Tecnologia Agraria y Alimentaria", INIA, em Madrid na Espanha. No ano de 2009 fez estágio de doutorado "sandwich" na University of Wisconsin, campus de Madison, WI, EUA, sob orientação do Prof. Dr. Guilherme Jordão de Magalhães Rosa.

*“As pessoas não entendem que o final de uma etapa é que torna possível o próximo passo”
(Khalil Gibran)*

A vida é feita por etapas, estamos sempre terminando e iniciando uma diferente, por isso devemos viver cada momento de nossa vida com o máximo de responsabilidade, pois as experiências de cada etapa vão sendo acumuladas para a próxima, e dependendo da forma como a conduzimos sofreremos as conseqüências.

Não existe dia igual a outro, por mais que achemos que estamos na rotina, ele chega sempre novo, cheio de novas oportunidades. Para isso existe a noite intercalando o dia, para filtrá-lo e trazê-lo de volta totalmente novo. O mal em geral, é que nos prendemos demais nas coisas que não deram certo e não percebemos que as portas continuam abertas para recomeçarmos e fazermos diferente. Costumamos desistir, logo na primeira vez que a má sorte nos dá as caras.

A vida é difícil mesmo e o mundo tem suas complicações, mas que graça teria se tudo fosse fácil, se tivéssemos tudo nas mãos?

Certamente reclamaríamos também, pois somos por natureza insatisfeitos.

Se chove queremos sol,

Se tem sol, queremos chuva,

Se faz calor queremos frio;

Se faz frio queremos calor;

E assim essa eterna insatisfação nos acompanha vida à fora.

O segredo da felicidade e da satisfação pessoal reside na capacidade de nos apaixonarmos por cada coisa que decidimos fazer.

Se tomamos a decisão de fazer algo, para que saia bem feito, o mínimo que se deve fazer é apaixonar-se já pela sua decisão.

Não decida fazer nada, pela qual não seja capaz de apaixonar-se.

Devemos nos esforçar o máximo para nos aperfeiçoar naquilo que sabemos fazer. Como disse alguém, que desconheço "Se você só sabe fritar um ovo, frite-o tão bem, para que você seja reconhecido mundialmente pela pessoa que melhor sabe fritar um ovo"

Gostar do que faz e fazê-lo bem feito, é a forma mais eficaz de conseguir realizar-se. E não ficar esperando que chegue o dia mágico em que você terá sorte e tudo se resolverá pra você.

O dia mágico, a hora mágica, o momento mágico é agora, só depende de você e da sua força de vontade.

(autor desconhecido)

Aos meus queridos pais, Antonio Carlos e Maria de Lourdes pela educação, amor, carinho que me proporcionaram, pelo incentivo de minha vida acadêmica e pelos ensinamentos que formaram os alicerces de minha história. Ao meu irmão Antonio Carlos, às minhas avós Anna e Izaura, à minha querida priminha Anna Carolina e ao amor da minha vida Mário, fontes de toda força e vontade que carrego comigo ao despertar de cada manhã.

Dedico e Ofereço

AGRADECIMENTOS

À Deus, por tornar tudo possível.

Aos meus pais, por terem acreditado em meus ideais, pelo exemplo de amor, a quem devo tudo que sou. Amo muito vocês.

Ao Programa de Pós-graduação em Genética e Melhoramento Animal da Faculdade de Ciências Agrárias e Veterinárias “Júlio de Mesquita Filho” – UNESP, pela oportunidade de realizar mais esta etapa da minha vida.

À Fundação de Amparo à Pesquisa do Estado de São Paulo pela concessão da bolsa de estudos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de Estágio de Doutorando - PDEE (“sandwich”) na University of Wisconsin em Madison – WI - EUA.

À Prof^ª. Dr^ª. Lucia Galvão de Albuquerque, pela orientação, pelos inúmeros ensinamentos, confiança e amizade, obrigada.

À Dr^ª. Lenira El Faro Zadra pela orientação, paciência, pela ajuda durante a realização deste trabalho e, principalmente, pela amizade.

Ao Dr. Roberto Augusto de Almeida Torres Júnior pela orientação e pelos ensinamentos.

Ao Prof. Dr. Guilherme Jordão de Magalhães Rosa pela orientação, pelos inúmeros ensinamentos e por ter me acolhido tão bem durante meu estágio “sandwich” na University of Wisconsin em Madison – WI - EUA.

Ao Prof. Dr. Fabyano Fonseca e Silva pelos inúmeros ensinamentos e principalmente pela amizade.

À Clínica do Leite, – Departamento de Zootecnia – USP/ESALQ, por intermédio do Prof. Dr. Paulo Fernando Machado, pela concessão dos dados que geraram este trabalho. Agropecuária Agrindus S.A e demais fazendas participantes pelo fornecimento dos seus bancos de dados.

Aos professores Fabyano Fonseca e Silva, Henrique Nunes Oliveira, Humberto Tonhati e Maria Eugênia Zerlotti Mercadante pelos ensinamentos, pela disponibilidade em participar da banca examinadora e pelas sugestões que contribuíram para aprimorar este trabalho.

Ao Prof. Danísio Prado Munari e Prof. João Ademir de Oliveira pelas valiosas sugestões prestadas no Exame de Qualificação.

Aos professores dos Departamentos de Zootecnia e Ciências Exatas pelos ensinamentos transmitidos.

Ao meu lindo, Mário Luiz Santana Júnior, pela compreensão, paciência, amizade, companheirismo e carinho nos momentos mais difíceis, desde o momento que o conheci, por tocar meu coração e ser a razão do meu sorriso. “Te amo muito”.

À Denise, obrigada por ser alguém muito especial em minha vida e espero que continue pela vida inteira, mesmo à distância, pois você é uma irmã que escolhi e que mora dentro do meu coração.

Às amigas de todas as horas Monyka e Arione “bixote”, obrigada pelo carinho, pela amizade eterna e mesmo longe na distância, próximas no coração... sempre!

Aos colegas e amigos da “Salinha” de Melhoramento Animal: Márcio Cinachi, Fernando, Fernanda, Marcos, Luciana, Fábio, Rodrigo, Ana Paula, Luís, Márcio Ribeiro, Severino, Diogo, Inaê, Rafael, Tomás pela amizade, ajuda e convívio.

Aos meus “orientados” de iniciação científica Daniel e Natália.

Aos colegas e amigos do departamento de Melhoramento Animal e Exatas: Roberta, Murote, Raul, Leonardo, Beto, Naudim, Geovany, Henry, Vanessa, Christina, Eveline, Gregório, Dimas, Matilde, Iara, Francisco, Davi, Guilherme, Diego, Daniela.

Às minhas queridas amigas e roomates de Madison (“602D”) Fernanda e Daniela e as agregadas (“933D”) Natália, Fumi, Luciana and “Myself” e Paulinha por serem pessoas tão maravilhosas para conviver durante todo este tempo.

Às pessoas incríveis que conheci em Madison, WI nos EUA, durante meu estágio de “sandwich”, em especial, Guilherme, Natália, Fabyano, Angélica, Bruno, Sandra, Márcio, Anibal, Walquíria, Fernando, Leandro, Alexandre, Theo, Thaís.

As minhas grandes amigas Laura, Michele, Raquel, Fernanda e Angélica pela amizade e por estarem sempre torcendo por mim.

Aos meus companheiros de viagem Lidiane e Carlos pela amizade e pelas caronas.

Aos meus filhinhos Brisa, Juma, Kurt e Lady pela lealdade, por sempre estarem ao meu lado e me fazer companhia.

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

SUMÁRIO

	Página
CAPÍTULO 1 – CONSIDERAÇÕES GERAIS.....	01
Introdução.....	01
Objetivos.....	03
Revisão de Literatura.....	03
Referências Bibliográficas.....	15
CAPÍTULO 2 – ANÁLISES MULTICARACTERÍSTICAS DA PRODUÇÃO DE LEITE NO DIA DO CONTROLE DE VACAS DA RAÇA HOLANDESA UTILIZANDO ANÁLISE DE FATORES E COMPONENTES PRINCIPAIS.....	23
Introdução.....	24
Material e Métodos.....	25
Resultados e Discussão.....	28
Conclusões.....	34
Referências.....	35
CAPÍTULO 3 – PARÂMETROS GENÉTICOS PARA PRODUÇÃO DE LEITE EM VACAS DA RAÇA HOLANDESA EMPREGANDO MODELOS DE REGRESSÃO ALEATÓRIA UTILIZANDO FUNÇÕES B-SPLINES E AMOSTRAGEM DE GIBBS.....	39
Introdução.....	40
Material e Métodos.....	42
Resultados e Discussão.....	45
Conclusões.....	53
Referências.....	54

MODELOS PARA ESTIMAÇÃO DE COMPONENTES DE (CO)VARIÂNCIA PARA PRODUÇÃO DE LEITE NO DIA DO CONTROLE DE VACAS DA RAÇA HOLANDESA

RESUMO – Parâmetros genéticos para a produção de leite no dia do controle (PLDC) de primeiras lactações de vacas da raça Holandesa foram estimados utilizando os modelos multicaracterísticas e modelos de regressão aleatória (MRA). Para os modelos multicaracterísticas foram analisados 15.896 controles mensais de produção de leite de 1.820 primeiras lactações de vacas da raça Holandesa. As análises foram realizadas por meio de sete modelos: multicaracterísticas padrão, três modelos de posto reduzido ajustando os primeiros 2,3 e 4 componentes principais genéticos e três modelos utilizando análise de fatores com 2,3 e 4 fatores. Para todos os modelos foram considerados os efeitos aleatórios genético aditivo e residual, e os efeitos sistemáticos do grupo de contemporâneo e da idade da vaca ao parto (efeito linear e quadrática) e do número de dias em lactação (efeito linear). A matriz de (co)variâncias residual, para todos os modelos, foi assumida ter posto completo. Os resultados indicam que somente dois componentes principais são requeridos para modelar a estrutura de (co)variâncias genéticas entre as produções de leite no dia do controle. Além disso, o modelo de posto reduzido diminui consideravelmente o número de parâmetros, sem reduzir a qualidade de ajuste. Para os MRA foram analisados 152.145 controles semanais de produção de leite de 7.317 primeiras lactações de vacas da raça Holandesa, provenientes de rebanhos da região Sudeste do Brasil. As produções de leite no dia do controle (PLDC) foram consideradas em 44 classes semanais de dia em lactação. Os grupos de contemporâneos foram definidos como rebanho-ano-semana do controle compondo 2.539 classes e, contendo, no mínimo, seis animais. O modelo utilizado incluiu os efeitos aleatórios genético aditivo direto, de ambiente permanente e o residual. Foram considerados como efeitos fixos, o grupo de contemporâneos, os efeitos linear e quadrático da covariável idade da vaca ao parto e a curva média de lactação da população, modelada por meio de polinômios ortogonais de Legendre de quarta ordem. Os efeitos aleatórios genético aditivo e de ambiente permanente foram modelados por meio de funções B-splines e inferência bayesiana. A estrutura residual

foi considerada heterogênea, contendo 6 classes de variâncias. Embora o modelo empregando B-spline cúbica com 6 nós foi o mais adequado de acordo com os critérios de comparação, um modelo mais parcimonioso empregando B-splines quadrática com 7 nós é suficiente para modelar as mudanças nas variâncias com o dia em leite para este conjunto de dados.

Palavras-chaves: Funções de covariância, Modelos reduzidos, Parâmetros genéticos, Splines, Test-day

MODELS FOR ESTIMATION OF COVARIANCE COMPONENTS OF FIRST LACTATIONS TEST DAY MILK YIELD OF HOLSTEIN CATTLE

SUMMARY – Genetic parameters for first lactation test-day milk yields of Holstein cattle were estimated using multivariate and random regression models (RRM). For multivariate model a total of 15,896 individual monthly test-day milk yields (10 test-days), from 1,820 complete first lactations of Holstein cattle. A standard multivariate analysis, reduced rank analyses fitting the first 2, 3 and 4 genetic principal components, and analyses that fitted a factor analytic structure considering 2, 3 and 4 factors, were carried out. All models included also fixed effects of the contemporary groups, age of cow (linear and quadratic effects) and days in milk (linear effect). The residual covariance matrix was assumed to have full rank throughout. The results indicate that only two principal components are required to model the genetic covariance structure among the test-days milk yield. Furthermore, reduced rank model allows decreasing the number of parameter without reducing the goodness of fit considerably. For RRM A total of 152,145 weekly test-day milk yield records from 7,317 first lactations of Holstein cows distributed across 93 herds in southeastern Brazil were analyzed. Test-day milk yields were classified into 44 weekly classes of days in milk (DIM). The contemporary groups were defined as herd-year-week of test-day. The model included direct additive genetic, permanent environmental and residual effects as random and fixed effects of contemporary group and age of cow at calving as covariable. Mean trends were modeled by a cubic regression on orthogonal polynomials of DIM, and additive genetic and permanent environmental effects were estimated using B-splines. Residual variances were modeled by step function with 6 variance classes. Although all the model selection criteria utilized indicated the model employing cubic B-splines to both random effects, with 6 knots, a more parsimonious model employing quadratic B-splines to both random effects, with 6 knots seems to suffice to model the changes in variances with days in milk for this data set.

Keywords: Covariance functions, Genetic parameters, Reduced rank model, Splines, Test-day

CAPÍTULO 1

CONSIDERAÇÕES GERAIS

1- INTRODUÇÃO

A produção de leite é um exemplo típico de medida repetida ou de dados longitudinais (medidas que são feitas seqüencialmente no mesmo animal, ao longo do tempo). As características medidas várias vezes durante a vida do animal têm sido estudadas de diversas formas. Abordagem mais simples seria analisá-las através de modelo de repetibilidade assumindo que todas as medidas representam a mesma característica. Outra alternativa seria o emprego de modelos multicaracterísticas utiliza a produção de leite no dia do controle (PLDC), denominados “test-day models” (TDM).

A maior desvantagem dos TDM é a dificuldade computacional de analisar um elevado volume de dados e a necessidade de estimar muito mais parâmetros, uma vez que utilizando modelo multicaracterísticas completo cada PLDC é considerada como uma característica distinta, assim o número de efeitos no modelo aumenta com o número de controles considerados. Atualmente, existem diferentes abordagens para diminuir a dimensão da matriz de (co)variâncias, como análise de componentes principais (modelos de posto reduzido) e a análise de fatores. A análise de componentes principais consiste na transformação do conjunto original de variáveis correlacionadas, num novo conjunto de variáveis as quais são combinações lineares das variáveis originais, não são correlacionadas entre si, e explicam a máxima quantidade de variação (KIRKPATRICK & MEYER, 2004). Na análise de fatores cada variável resposta é representada por função linear de pequena quantidade de fatores comuns, os quais descrevem a correlação entre as variáveis, e de fatores específicos (MEYER, 2009).

Os modelos de regressão aleatória (MRA) têm sido recomendados para modelar dados longitudinais, como as PLDC, e também vem sendo utilizados

comumente nas avaliações genéticas de bovinos leiteiros (INTERBULL, 2008). Os MRA permitem o ajuste de curvas de lactação aleatórias para cada indivíduo, expressas como desvios de uma curva média da população ou de grupos de indivíduos. Ao se ajustar um MRA assume-se implicitamente uma estrutura de covariâncias entre as observações, e esta é determinada pelas covariâncias entre os coeficientes de regressão, podendo ser caracterizada como função de covariâncias.

Em vários trabalhos, são apresentadas estimativas de parâmetros genéticos para a produção de leite no dia do controle, aplicando MRA com base nos polinômios de Legendre (COBUCI et al., 2005; ARAÚJO et al., 2006, EL FARO et al., 2008; BIGNARDI et al., 2009a). O maior problema desses modelos é a utilização de polinômios de alto grau e a desuniformidade de ajuste ao longo da curva. Segundo MEYER (2005) análises de regressão aleatória com polinômios de ordem cúbica, quártica ou mesmo de ordens mais elevadas têm levado a estimativas de componentes de variância e parâmetros genéticos erráticos e implausíveis. Isto tem sido especialmente prevalente em conjuntos de dados que contêm poucos registros nas últimas idades, em conjuntos de dados com poucos registros por animais e em análises em que os animais têm menos registros do que a ordem dos polinômios ajustados.

Alternativas para diminuir o grau dos polinômios estão sendo estudadas. Uma delas é a utilização de funções “splines” ou também chamados de polinômios segmentados, que são curvas constituídas por segmentos individuais de polinômios de grau reduzido e que se unem em pontos específicos, denominados nós (De BOOR, 1978). As funções “splines” caracterizam-se por atender quase integralmente as propriedades ótimas de uma função: possuir interpretação biológica simples, ser de fácil estimação (linear nos parâmetros) e redução da multicolinearidade (FULLER, 1969; SCHENKEL, 1999). Além disso, elas permitem, de forma geral, tanto um bom ajuste local como também um bom ajuste global.

2- OBJETIVOS

Objetivos gerais:

O objetivo geral proposto neste trabalho foi estudar a variação genética da produção de leite no dia do controle de vacas da raça Holandesa, utilizando modelos multicaracterísticas e modelos de regressão aleatória, visando obter modelos adequados e parcimoniosos para estimar parâmetros genéticos e prever valores genéticos dos animais.

Objetivos específicos:

- Estimar parâmetros genéticos para a produção de leite no dia do controle de vacas da raça Holandesa utilizando análise multicaracterística padrão, análise de fatores e análise de componentes principais.

- Estimar funções de covariâncias e parâmetros genéticos para a produção de leite em vacas da raça Holandesa aplicando funções B-splines e amostragem de Gibbs em modelos de regressão aleatória.

3- REVISÃO DE LITERATURA

3.1 Análises de componentes principais

A análise de componentes principais, de acordo com MORRISON (1976), consiste de um método de transformar variáveis correlacionadas em outro grupo de variáveis não correlacionadas, servindo ainda para a obtenção de combinações lineares das variáveis originais com variabilidade relativamente grande, além de ser uma ferramenta para redução da dimensionalidade dos dados. A técnica de

componentes principais foi originalmente descrita por PEARSON (1966) e, posteriormente, aplicada por HOTELLING (1935), em diversas áreas.

De acordo com CRUZ (1990), os componentes principais são independentes entre si e são estimados com o propósito de reter, em ordem de estimação, o máximo da informação, ou seja, o primeiro componente principal possui a máxima variância, tendo o segundo componente principal a segunda maior variância e, assim, sucessivamente, de forma que toda a informação está contida nos primeiros componentes.

KIRKPATRICK & MEYER (2004) mostraram que é possível estimar componentes principais genéticos através da simples reparametrização do modelo linear misto usual. Os autores relatam ainda que essa técnica é aplicável para qualquer análise multicaracterísticas, onde os efeitos genéticos são correlacionados. Dependendo da magnitude de correlação genética, um pequeno subconjunto de componentes principais geralmente é suficiente para capturar a maior parte da variação genética. MEYER (2005) relatou que as vantagens dessa técnica são: primeiramente estimando os m primeiros componentes principais produz um modelo mais parcimonioso, reduzindo o número de parâmetros de $q(q+1)/2$ (modelo multicaracterísticas padrão) para $m(2q-m+1)/2$. Em segundo lugar, as variâncias tendem a aumentar com o número de parâmetros estimados. Se os componentes principais têm omitido autovalores próximos de zero, viés nas estimativas em decorrência de tal omissão é insignificante e pode ser compensado pela redução das variâncias, resultando assim em menor quadrado médio do resíduo em comparação com análises de posto completo. Em terceiro lugar, as necessidades computacionais são bem menores do que as requeridas para um modelo multicaracterísticas de posto completo. Isso pode facilitar as análises de grandes conjuntos de dados, produzindo assim estimativas mais precisas dos parâmetros genéticos.

Dado em notação matricial o modelo multicaracterísticas padrão:

$$y = X\beta + Zu + e \quad (1)$$

em que y representa o vetor das observações, β é o vetor de soluções para os efeitos fixos; $u =$ é o vetor de soluções para os efeitos aleatórios genéticos aditivos; X e Z são as matrizes de incidência, para os efeitos fixos e aleatório de animal, respectivamente, e e é o vetor do efeito aleatório residual, A o numerador da matriz de parentesco e com $Var(u) = \Sigma \otimes A$ onde Σ é a matriz de (co)variâncias.

O Modelo de Componentes Principais pode ser obtido pela reparametrização do modelo em (1), a qual produz um modelo equivalente que contempla os componentes principais. Tal modelo, em notação matricial, é representado pela expressão:

$$y = X\beta + Z^* u^* + e \quad (2)$$

em que $Z^* = Z(E \otimes I)$, $u^* = (E \otimes I) u$, $Var(u^*) = (I \otimes A)$ e $\Sigma = E\Lambda E'$. Nestas definições, E correspondendo a matriz dos autovetores (v) e Λ a matriz diagonal dos autovalores λ_i . Assim, a matriz de (co)variâncias genética (Σ) pode ser decomposta em termos de E e Λ , de forma que $EE' = I$. Considerando m componentes principais, a matriz E é substituída por E_m , a qual compreende as m primeiras colunas (v_1, \dots, v_m) de E , portanto, E_m é então utilizada para definir Z^* e u^* (KIRKPATRICK and MEYER, 2005).

3.1 Análises de fatores

Inicialmente a análise de fatores esteve ligada ao problema da avaliação de escalas cognitivas e foi responsabilidade de uma série de pesquisadores da área de psicologia (SPEARMAN, 1904; THURSTONE, 1935 e BURT, 1941). No início, os métodos apresentavam uma característica mais empírica do que formal. LAWLEY (1942) publica um primeiro trabalho com maior rigor matemático (em termos de inferência estatística), o que fez com que se aumentasse a aceitação dessas técnicas.

A análise de fatores é definida por REYS (1997) como uma técnica estatística multivariada que, a partir da estrutura de dependência existente entre as variáveis de interesse (em geral representada pelas correlações ou (co)variâncias entre essas variáveis), permite a criação de um conjunto menor de variáveis (variáveis latentes, ou fatores) obtidas como função das variáveis originais. Além disso, é possível saber o quanto cada fator está associado a cada variável (fatores específicos) e o quanto o conjunto de fatores explica da variabilidade geral dos dados originais (fatores comuns).

Conforme FERNANDES & LIMA (1991) na composição desses fatores tem-se que:

- as variáveis mais correlacionadas se combinam dentro do mesmo fator, promovendo uma redução do número inicial de variáveis em um número reduzido de fatores;

- as variáveis que compõem um determinado fator são praticamente independentes das que constituem outro fator;

- a derivação dos fatores se processa visando maximizar a porcentagem de variância total relativa a cada fator consecutivo;

- os fatores não são correlacionados entre si.

Inicialmente, estudando o efeito da interação x ambiente em plantas SMITH et al. (2001) mostraram que é possível utilizar a análise de fatores em um modelo misto. No melhoramento animal, MEYER (2007) utilizou a análise de fatores, assim como análise de componentes principais, para estudar 14 características de carcaça em bovinos Angus, e concluiu que, no máximo, oito componentes principais são necessários para modelar de estrutura de covariância genética entre as características. MEYER (2009) demonstrou que é possível empregar a análise de fatores em um modelo misto para estudar o efeito da interação genótipo ambiente em melhoramento animal assim como fez SMITH et al. (2001) em melhoramento de plantas.

No Modelo de Análise de Fatores, o efeito genético pode ser explicado por m fatores comuns (c) e q fatores específicos (s). Assume-se que a matriz de (co)variâncias para os efeitos genéticos é dada por $\Sigma = \Gamma\Gamma' + \Psi$, em que: Γ é a matriz de cargas fatoriais, cuja dimensão é $m \times q$; e Ψ é a matriz das variâncias dos efeitos específicos, sendo esta uma matriz diagonal de ordem m . O modelo em questão pode ser representado em notação matricial como:

$$y = X\beta + Z^+c + Z^o s + e \quad (5)$$

em que: c e s representam, respectivamente, os fatores comuns e fatores específicos. Assim, tem-se: $Z^+ = Z^o(\Gamma \otimes I)$, $c = (I_N \otimes \Gamma' \Sigma^{-1})\mu$, $s = (I_N \otimes \Psi \Sigma^{-1})\mu$, $Var(c) = I$ e $Var(s) = \Psi \otimes I$ (THOMPSON et al., 2003 e MEYER, 2009).

3.2 Modelos de regressão aleatória e inferência bayesiana

Os modelos de regressão aleatória (MRA) fornecem ou possibilitam a predição de valores genéticos para a curva de lactação como um todo, para qualquer idade desejada e para funções da curva de lactação, em contraste com os modelos multi-característica, que fornecem predições pontuais. As regressões são ajustadas em função do período de lactação, utilizando-se polinômios ordinários ou outras funções lineares, e modelam trajetórias para a média populacional (regressões fixas) e para cada animal (regressão aleatória). Outra vantagem dos MRA é que a estrutura de covariâncias entre as medidas pode ser caracterizada por uma função de covariâncias, o que possibilita modelar de forma mais acurada a estrutura de covariâncias das características, dando maior flexibilidade na utilização de dados longitudinais (VAN der WERF & SCHAEFFER, 1997).

Introduzidas por KIRPATRICK & HECKMAN (1989), as funções de (co)variâncias são funções contínuas em relação ao tempo que representam as variâncias e as covariâncias de características mensuradas em diferentes pontos em uma trajetória (curva de lactação ou crescimento). MEYER (1998) mostraram que um MRA é um caso especial das funções de covariâncias e que os coeficientes das

funções de covariâncias podem ser estimados diretamente dos MRA por máxima verossimilhança restrita (REML).

Outro método para se estimar os coeficientes de funções de covariância é a inferência Bayesiana. Em inferência Bayesiana não existe qualquer distinção entre efeitos fixos ou aleatórios, sendo que os parâmetros a serem estimados são considerados variáveis aleatórias (GIANOLA & FERNANDO, 1986) que devem ser estimadas considerando as incertezas associadas a elas. O objetivo da inferência Bayesiana é encontrar a distribuição de probabilidade a posteriori do parâmetro ($f(\theta|y)$), em que: θ é o parâmetro e y é o vetor de observações. Quando essa distribuição é obtida, pode-se determinar o menor intervalo no qual a probabilidade de encontrar θ é maior que 95%, o qual é denominada região da distribuição a posteriori de alta probabilidade (highest posterior density(HPD)).

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

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

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

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

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

sendo, $f(\theta)$ a função densidade de probabilidade, a priori de θ , que reflete a relativa incerteza sobre os possíveis valores de θ , antes do vetor y ter sido observado, $f(y|\theta)$ é a função densidade conjunta de probabilidade de y condicional a θ , que representa a verossimilhança, ou seja, a quantidade de informação contida em y para o conhecimento de θ , $f(\theta|y)$ é a função de densidade de probabilidade a posteriori, que reúne em si, o estado de incerteza sobre θ dado o conhecimento prévio sobre os parâmetros e a informação contida no vetor de observações, e \propto é o operador que indica proporcionalidade. Assim a distribuição posteriori de θ contém todas as informações necessárias para se fazer inferência a respeito de θ . No

entanto, em situação multiparamétrica é necessário integrar a função $f(\theta|y)$, para obtenção de distribuições marginais a posteriori de cada parâmetro.

Métodos Bayesianos baseados na metodologia da Cadeia de Markov Monte Carlo (MCMC), dentre os quais se destacam a Amostragem de Gibbs (Gibbs Sampling), são aplicados para obter as amostras das distribuições marginais a posteriori e vêm sendo propostos como alternativa para solucionar questões relacionadas à estimação genética componentes de (co)variância. De acordo com CASELLA & GEORGE (1992), o amostrador de Gibbs é um procedimento numérico para obter amostra da distribuição conjunta ou marginal de todos os parâmetros do modelo, por repetidas amostragens de todas as distribuições condicionais completas formando uma cadeia de Markov. De acordo com VAN TASSEL et al. (1995) o uso do método amostrador de Gibbs na estimação de componentes de covariância, apresenta várias vantagens em relação aos métodos usuais que empregam a teoria tradicional (BLUP/REML): não requer soluções para as equações de modelo misto; permite a análise de conjuntos de dados maiores do que quando se usa REML com técnicas de matrizes esparsas; propicia estimativas diretas e acuradas dos componentes de covariância e intervalos de credibilidade para essas estimativas e apresenta facilidade computacional.

O amostrador de Gibbs tem sido intensamente empregado na estimação de parâmetros genéticos em estudos que utilizam MRA (JAMROZIK & SCHAEFFER, 1997; STRABEL & MISZTAL, 1999; REKAYA et al. 1999; JAMROZIK et al., 2001; LÓPEZ-ROMERO et al., 2003; MEYER, 2005)

3.3 Polinômios segmentados (funções “Splines”)

O nome “spline” originou-se de uma tira de madeira longa, fina e flexível usada tradicionalmente para desenhar curvas. “Splines” são geralmente definidas como curvas constituídas por segmentos individuais que são facilmente unidos; os segmentos são dados por polinômios e os pontos nos quais eles estão unidos são chamados de nós (DODGE et al., 2003).

FULLER (1969) reportou que funções “splines” caracterizam-se por atender quase integralmente as propriedades ótimas de uma função: ser contínua; possuir primeiras derivadas contínuas; ser de fácil estimação (linear nos parâmetros); e permitir com facilidade o cálculo dos estimadores. Segundo SCHENKEL (1999), as vantagens dos polinômios segmentados sobre os polinômios ordinários são:

- Quebra da multicolinearidade, aumentando a confiabilidade dos valores estimados, uma vez que minimiza o fator de inflação da variância e o viés dos estimadores.

- O comportamento da função que descreve a relação das variáveis em uma região torna-se bastante independente do comportamento em outras regiões com características distintas, o que torna a capacidade de aderência da função “spline” aos pontos superior a dos polinômios ordinários.

- Funções não lineares e polinômios ordinários de graus elevados podem ser aproximados por polinômios segmentados que são lineares nos parâmetros, de fácil estimação, com propriedades ótimas e de simples interpretação biológica.

Um particular tipo de função splines chamada B-spline é preferida a outros tipos de spline, devido suas ótimas propriedades numéricas (RUPERT et al., 2003, EILERS & MARX, 2005). De acordo com RICE & WU (2001) a utilização de B-splines é adequada para modelar efeitos aleatórios em análises de modelos mistos, e também eficientes na estimação de funções de covariância.

Como descritas por MEYER, (2005), as B-splines de grau $p=0$ têm valores iguais à unidade para todos os pontos (t) em um dado intervalo, e zero caso contrário. Para o k -ésimo intervalo dado por nós T_k e T_{k+1} com $T_k \leq T_{k+1}$:

$$B_{k,0}(t) = \begin{cases} 1 & \text{se } T_k \leq t < T_{k+1}, \\ 0 & \text{caso contrário} \end{cases}$$

As funções bases de grau $p > 0$ podem ser representadas por:

$$B_{k,p}(t) = \frac{t - T_k}{T_{k+p} - T_k} B_{k,p-1}(t) + \frac{T_{k+p+1} - t}{T_{k+p+1} - T_{k+1}} B_{k+1,p-1}(t)$$

WHITE et al. (1999) utilizaram uma abordagem não-paramétrica por meio do uso de funções splines cúbicas para o ajuste da curva de lactação. O modelo completo incluiu 10 nós e, além deste, os autores utilizaram também modelos reduzidos com 3, 4, 5, 6, e 8 nós localizados na média do dia em lactação de cada controle, comparando os ajustes destes modelos com o obtido usando a curva exponencial de Wilmink. A herdabilidade foi de 0,23 no primeiro controle, 0,20 do quinto ao nono e 0,27 no último controle. Os modelos de splines que continham mais que 8 nós obtiveram melhor ajuste do que a curva exponencial de Wilmink.

MEYER (2005) estimou parâmetros genéticos para pesos do nascimento aos 820 dias de idade em bovinos de corte da raça Angus, aplicando funções B-splines para descrever as trajetórias para os efeitos aleatórios de um modelo de regressão aleatória. O autor reportou que este método é menos susceptível a problemas freqüentemente observados em análises utilizando polinômios ortogonais de Legendre. Esta menor susceptibilidade se deve ao fato do modelo conter segmentos individuais de polinômios de menores graus e um melhor controle da influência global das observações individuais. Um bom ajuste dependerá de uma boa escolha dos graus do polinômio e os locais onde os nós serão inseridos.

Estudando sete funções, dentre elas a de Wood, Wilmink, Ali e Schaeffer, três polinômios de Legendre (ordens 3 a 5) e uma spline cúbica para modelar a curva de lactação utilizando controles diários de quatro rebanhos do norte de Portugal, SILVESTRE et al., (2006) concluíram que, embora a performance de todos os modelos tenha sido aceitável, a função spline e as funções com 5 parâmetros (Ali e Schaeffer e polinômio de Legendre) mostraram-se superiores.

DeGROOT et al. (2007) utilizando registros de controles no dia do controle para a produção de leite de 3 lactações em animais da raça Holandesa, concluíram que um MRA empregando uma função “spline” cúbica com 5 nós predeterminado nos dias 0, 50, 135, 220, e 305 foi eficiente em descrever a produção de leite no decorrer da lactação. Entretanto, BOHMANOVA et al. (2008) comparando MRA empregando polinômios de Legendre e “splines” lineares para modelar as (co)variâncias de dados de produção de leite, reportaram que de acordo com o critério DIC (Deviance

Information Criterion) o melhor modelo seria o que empregou “spline” linear com 6 nós. Todavia, os outros critérios utilizados mostraram pequenas diferenças entre todos os modelos.

4- PARÂMETROS GENÉTICOS

4.1 - Estimativas de Herdabilidade

As herdabilidade estimadas para produção no dia do controle em análises uni e multicaracterísticas, em vacas Holandesas, têm se apresentado próximas ou superiores às encontradas para produção acumulada, sendo o meio da lactação, mais herdável que o início e o final (OLORI, 1997; KETTUNEN et al., 1998; MACHADO et al., 1998; FERREIRA et al., 2003; MELO et al., 2005).

As estimativas de parâmetros genéticos obtidas de MRA, de modo geral, são dependentes das funções de regressão utilizadas e da estrutura de (co)variâncias para efeitos genético aditivo, de ambiente permanente e residual. Em estudos iniciais empregando MRA (JAMROZIK & SCHAEFFER, 1997), onde o resíduo era modelado considerando estrutura de variâncias homogêneas, foi observado que houve problemas na estimação dos demais componentes, como a superestimação das variâncias genéticas aditivas. Desta forma, em vários estudos (OLORI et al., 1999; BROTHERSTONE et al., 2000; EL FARO et al., 2008; BIGNARDI et al., 2009a) foi proposta a utilização de estruturas de variâncias residuais heterogêneas, modelando-as por agrupamento de classes que contenham variações semelhantes, com números reduzidos de classes de variâncias, ou através de funções de variâncias residuais, empregando polinômios ortogonais ou ordinários.

Quando se considerou o efeito de ambiente permanente como uma estrutura de variâncias homogênea, alguns autores (JAMROZIK & SCHAEFFER, 1997; KETTUNEN et al., 2000) constataram que não foi possível separar a variação

genética da não genética, proporcionando uma superestimação dos valores de herdabilidade, e quando o efeito foi modelado por uma estrutura de variâncias heterogênea, ocorreu melhor partição das variâncias e os valores de herdabilidade foram mais similares aos resultados obtidos mediante análises usando “test-day models” ordinário.

Na maioria dos trabalhos têm-se utilizado MRA com base nos polinômios ortogonais de Legendre para estudar a estrutura de (co)variâncias para efeitos aleatórios genético aditivo e de ambiente permanente (STRABEL & MISZTAL, 1999; OLORI et al., 1999; ARAÚJO et al., 2006; EL FARO et al., 2008; BIGNARDI et al., 2009a). Funções paramétricas como polinomial logarítmica, de ALI & SCHAEFFER (1987) e a exponencial, de WILMINK (1987), também têm sido utilizadas e proporcionado estimativas de parâmetros genéticos que diferem dos obtidos pelos polinômios ortogonais de Legendre (BROTHERSTONE et al., 2000 e ARAÚJO et al., 2006).

Vários autores têm relatado que as estimativas de herdabilidade obtidas usando MRA têm sido maiores no início e/ou no final da lactação, indicando que as maiores variações genéticas ocorrem nesses períodos (JAMROZIK & SCHAEFFER, 1997; REKAYA et al., 1999; STRABEL & MISZTAL, 1999; KETTUNEN et al., 2000; DeGROOT et al., 2007; EL FARO et al., 2008, BIGNARDI et al., 2009a). Entretanto, SILVESTRE et al. (2005) utilizando um MRA empregando “spline” cúbica na raça Holandesa em Portugal e COSTA et al. (2008), em estudos realizados usando MRA utilizando polinômios de Legendre na raça Holandesa no Brasil, reportaram que os valores de herdabilidade foram maiores na fase intermediária da lactação.

4.2 Estimativas de Correlações

As correlações fenotípicas assim como as correlações genéticas entre as produções de leite em cada controle têm-se mostrado elevadas e altas quanto mais próximos os controles, e decrescendo, gradativamente, com o aumento no intervalo

entre os mesmos (KETTUNEN et al., 2000; BROTHERSTONE et al., 2000; JAKOBSEN et al., 2002; ARAÚJO et al., 2006; BIGNARDI et al., 2009).

Alguns autores, dentre eles, BROTHERSTONE et al. (2000) e KETTUNEN et al. (2000) e no Brasil COBUCI et al. (2005) e BIGNARDI et al. (2009a,b) estimaram correlações genéticas negativas entre as PLDC utilizando MRA empregando diferentes funções em bovinos da raça Holandesa. BROTHERSTONE et al. (2000) reportaram que quando utilizaram a função paramétrica de WILMINK (1987) ocorreram estimativas de correlações genéticas negativas entre o primeiro controle e os controles a partir dos 42 dias de lactação, e atribuíram isso ao fato de que funções paramétricas não conseguem modelar a associação entre as produções no início e fim da lactação. KETTUNEN et al. (2000) utilizando polinômios ortogonais de Legendre ou função de ALI & SCHAEFFER (1987) para o efeito genético aditivo e mantendo o efeito de ambiente permanente constante ao longo da lactação, também observaram estimativas de correlações genéticas negativas para a produção entre os controles mais distantes e este fato poderia ser atribuído a uma superestimação da variância genética aditiva nos extremos da curva de lactação devido à constância do efeito de ambiente no decorrer da lactação.

COBUCI et al. (2005) não obtiveram estimativas de correlações genéticas negativas quando utilizaram uma função de WILMINK (1987) para ajustar os efeitos genéticos aditivo e de ambiente permanente. Contudo, quando utilizaram um MRA que considerava a variância de ambiente permanente constante, houve estimativas de correlações genéticas negativas. BIGNARDI et al. (2009a) utilizando MRA empregando polinômios de Legendre para os efeitos aleatórios, reportaram estimativas negativas entre controles do início da lactação e controles do final da lactação e atribuíram isso ao fato de as vacas nesse período inicial da lactação apresentarem déficit de energia e também ao estresse devido ao parto.

5. REFERÊNCIAS BIBLIOGRÁFICAS

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

ARAÚJO, C. V.; TORRES, R. A.; COSTA, C. N.; et. al. Uso de modelos de regressão aleatória para descrever a variação genética da produção de leite na raça Holandesa. **Revista Brasileira de Zootecnia**, vol.35, no.3, supl, p.975-981, 2006.

BIGNARDI, A.B.; EL FARO, L.; ALBUQUERQUE, L.G.; CARDOSO, V.L.; MACHADO, P.F. Random regression models to estimate test-day milk yield genetic parameters Holstein cows in southeastern Brazil. **Livestock Science**, v.123, p.1-7, 2009a.

BIGNARDI, A.B.; EL FARO, L.; ALBUQUERQUE, L.G.; CARDOSO, V.L.; MACHADO, P.F. Parametric correlation functions to model the structure of permanent environmental (co)variances in milk yield random regression models. **Journal Dairy Science**, v.92, p.4634–4640, 2009b.

BOHMANOVA, J.; MIGLIOR, F.; JAMROZIK, J.; MISZTAL, I.; SULLIVAN, P. G. Comparison of random regression models with Legendre polynomials and linear splines for production traits and somatic cell score of Canadian Holstein cows. **Journal Dairy Science**, v.91, p.3627–3638, 2008.

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

BURT, C. – **The vector of the mind: an introduction to factor analysis in psychology**. Mc.Millan, New York, 1941

CASELLA, G.; GEORGE. E. I., Explaining the Gibbs Sampler. **The American Statistician**, v.46 p.167-174, 1992.

COBUCI, J.A.; EUCLYDES, R.F.; LOPES, P.S.; COSTA, C.N.; TORRES, R.A.; PEREIRA, C. S. Estimation of genetic parameters for test-day milk in Holstein cows using a random regression model. **Genetics and Molecular Biology**, vol. 28, n^o1, p. 75-83, 2005.

COSTA, C.N.; MELO, C.N.R.; PACHER, I.U. et al. Genetic parameters for test day milk yield of first lactation Holstein cows estimated by random regression using Legendre polynomials. **Revista Brasileira de Zootecnia**, v.37, p.602-608, 2008.

CRUZ, C.D. **Aplicação de algumas técnicas multivariadas no melhoramento de plantas**. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, p.188, 1990.

De BOOR, C. **A practical guide to splines**. Springer-Verlag, New York, 1978.

DEGROOT, B. J.; KEOWN, J. F.; VAN VLECK, L. D.; KACHMAN S. D. Estimates of genetic parameters for Holstein cows for test-day yield traits with a random regression cubic spline model. **Genetics Molecular Research** v.6, p.434–444, 2007.

DODGE Y., **The Oxford Dictionary of Statistical Terms**, Oxford University Press, Oxford and New York, 2003.

EILERS, PHC; MARX, BD **Splines, knots, and penalties**. (2005) 10 jan. 2008. Capturado em 10 jan. 2008. On-line. Disponível na Internet: http://www.stat.lsu.edu/faculty/bmarx/splines_knots_penalties.pdf

EL FARO, L., ALBUQUERQUE, L.G., CARDOSO, V.L. Variance component estimates for test-day milk yield applying random regression models. **Genetics Molecular Biology** 33 (3), 665–673 (suppl). 2008.

FERNANDES, T.A.G.; LIMA, J.E. Uso de análise multivariada para a identificação de sistemas de produção. **Pesquisa Agropecuária Brasileira**, Brasília, v.26,n.10, p.1823-1836, 1991.

FERREIRA, W. J.; TEIXEIRA, N. M.; EUCLYDES, R. F. Avaliação genética de bovinos da raça holandesa usando a produção de leite no dia do controle. **Revista Brasileira de Zootecnia**, vol.32, n.2, p.295-303, 2003.

FULLER W.A., Grafted polynomials as approximating functions, **Australian Journal Agriculture Economic**. v.13 p.35–46, 1969.

GIANOLA, D., FERNANDO, R. Bayesian methods in animal breeding theory. **Journal Animal Science**, v.63, p.217-244, 1986.

HOTELLING, H. The most predictable criterion. **Journal of Educational Psychology**. v.26, p.139-142, 1935.

INTERBULL. 10 jan. 2008. Capturado em 10 jan. 2008. On-line. Disponível na Internet: <http://www-interbull.slu.se>.

JAMROZIK, J.; SCHAEFFER, L.R. Estimates of genetic parameters for a test day model with random regressions for production of first lactation Holsteins. **Journal Dairy Science**, v.80, p.762-770, 1997.

JAMROZIK, G. J.; GIANOLA, D.; 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.

KETTUNEN, A.,MÄNTYSAARI, E., PÖSÖ J. Estimation of genetic parameters for daily milk yield of primiparous Ayrshire cows by random regression test-day models. **Livestock Production Science**,v.66, p.251-261, 2000.

KIRKPATRICK, M.; HECKMAN, N. A quantitative genetic model for growth, shape and other infinite-dimensional characters. **Journal of Mathematical Biology**, v.27, p.429-450, 1989.

KIRKPATRICK, M.; MEYER, K. Direct estimation of genetic principal components: simplified analysis of complex phenotypes. **Genetics**. v.168, p.2295–2306, 2004.

KIRKPATRICK, M.; MEYER, K. Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. **Genetics Selection Evolution** v.21, p.1-30, 2005.

LAWLEY, D.N. Further estimation in factor analysis. **Proceedings of the Royal Society of Edinburgh**, Series A. v.61, pp.176-185, 1942.

LOPEZ-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.

MACHADO, S.G.; FREITAS, M.A.R.; GADINI, C.H Genetic parameters of test day milk yields of Holstein cows. **Genetics and Molecular Biology**, v.22, n°3, p.383-386,1999.

MELO, C.M.R; COSTA, C.N; PACKER, U.I.; MACHADO, P.F. Parâmetros genéticos para as produções de leite no dia do controle e da primeira lactação de vacas da raça Holandesa. **Revista Brasileira de Zootecnia** v.34, p.796-806, 2005

MEYER, K. Estimating covariance functions for longitudinal data using a random regression model. **Genetic Selection Evolution**, v. 30, p.221-240, 1998.

MEYER, K. Random regression analyses using B-splines to model growth of Australian Angus cattle. **Genetics Selection Evolution**, vol.37, p.473-500, 2005.

MEYER, K. Multivariate analyses of carcass traits for Angus cattle fitting reduced rank and factor analytic models. **Journal Animal Breeding Genetics**, v. 124, p.50-63, 2007.

MEYER, K. Factor-analytic models for genotype \times environment type problems and structured covariance matrices. **Genetics Selection Evolution** v.41, p.21-32, 2009.

MORRISON, D.F. **Multivariate statistical methods**. 2.ed New York: McGraw-Hill company, p.415, 1976.

OLORI, V. E.; HILL W. G.; MCGUIRK, B. J.; BROTHERSTONE S. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. **Livestock Production Science**, v.61, p.53–63, 1999.

PEARSON, E.S.; HARTLEY, H.O. **Biometrika Tables for Statisticians**, Vol. 1 ed. Cambridge University Press, New York, 1966.

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.

REYS, E. **Estatística multivariada aplicada**. Lisboa: Edições Silabo, p.343, 1997.

RICE, J.R. **The approximation of functions**. Vol.2, Massachusetts, Addison-Wesley 1969.

RICE, J.A., WU C.O., Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, v.57, p.253-259, 2001.

RUPPERT D., WAND M.P., CARROLL R.J., **Semiparametric Regression**, Cambridge University Press, New York, 2003.

SCHENKEL, F.S. **Utilização de polinômios segmentados na pesquisa zootécnica (considerações teóricas e práticas)**. Universidade Federal do Rio Grande do Sul – Faculdade de Agronomia – Departamento de Zootecnia, Seminário em Zootecnia, Porto Alegre-RS, 1999.

SILVESTRE, A. M., PETIM-BATISTA, F., COLAÇO, J. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. **Journal Dairy Science**, v.89, p.1813-1821, 2006.

SMITH, A.B.; CULLIS, B.R.; THOMPSON R., Analysing variety by environment data using multiplicative mixed models and adjustments for spatial field trends, **Biometrics**, v.57, p.1138–1147, 2001.

SPEARMAN, C. "General intelligence" objectively determined and measured. **American Journal of Psychology**, v.15, p.201-293, 1904.

STRABEL, T.; MISZTAL, I. Genetic parameters for first and second lactation milk yields of Polish Black and White cattle with random regression test-day models. **Journal Dairy Science**, v.82, p.2805-2810, 1999.

THOMPSON, R.; CULLIS, B.R.; SMITH, A.B; GILMOUR, A.R.; A sparse implementation of the Average Information algorithm for factor analytic and reduced rank variance models. **Australian New Zealand Journal Statistics**, v.45, p.445-459, 2003.

THURSTONE, L.L. **The vectors of mind**. Univ. Chicago Press, Chicago, 1935.

VAN DER WERF, J.; SCHAEFFER, L. **Random Regression in Animal Breeding**. Course Notes, Ontário: University of Guelph, p.70, 1997.

VAN TASSEL, C.P.; CASELLA, G.; POLLAK, E.J. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. **Journal of Dairy Science**, v.78, n.3, p.678-692, 1995.

WHITE I.M.S., THOMPSON R., BROTHERSTONE S., Genetic and environmental smoothing of lactation curves with cubic splines, **Journal Dairy Science** vol. 82 p.632–638, 1999.

WILMINK, J. B. M. Efficiency of selection for different cumulative milk, fat and protein yields in first lactation. **Livestock Production Science**, v.17, p.211-224, 1987.

CAPÍTULO 2 - ANÁLISES MULTICARACTERÍSTICAS DA PRODUÇÃO DE LEITE NO DIA DO CONTROLE DE VACAS DA RAÇA HOLANDESA UTILIZANDO ANÁLISE DE FATORES E COMPONENTES PRINCIPAIS

RESUMO – Foram analisados 15.896 controles mensais de produção de leite de 1.820 primeiras lactações de vacas da raça Holandesa. As análises foram realizadas por meio de sete modelos: multicaracterísticas padrão (MC), três modelos de posto reduzido ajustando os primeiros 2,3 e 4 componentes principais genéticos e três modelos utilizando análise de fatores com 2,3 e 4 fatores. Para todos os modelos foram considerados os efeitos aleatórios genético aditivo e residual, e os efeitos sistemáticos do grupo de contemporâneo e da idade da vaca ao parto (efeito linear e quadrática) e do número de dias em lactação (efeito linear). A matriz de (co)variâncias residual, para todos os modelos, foi assumida ter posto completo. Entre todos os modelos, o modelo ajustando dois primeiros componentes principais (MCP2) foi o melhor de acordo com os critérios de comparação. As estimativas de variâncias fenotípicas, genéticas e residuais para os modelos MC MCP2 foram similares. As estimativas das herdabilidades oscilaram entre 0,12 e 0,34 e entre 0,11 e 0,31, respectivamente, para os modelos MC, MCP2. As estimativas de correlação genética obtidas com o modelo MCP2 foram superiores as obtidas pelo modelo MC. Os resultados indicam que somente dois componentes principais são requeridos para modelar a estrutura de (co)variâncias genéticas entre as produções de leite no dia do controle. Além disso, o modelo de posto reduzido diminui consideravelmente o número de parâmetros, sem reduzir a qualidade de ajuste.

Palavras-chave: (Co)variância, Modelos reduzidos, Test-day

INTRODUÇÃO

Os modelos denominados “test-day models” (TDM), tem sido o método de escolha para análise de produção de leite no dia do controle (PLDC). Nos TDM é possível a inclusão de lactações incompletas, aumentando o número de filhas por touro, proporcionando, conseqüentemente, uma maior confiabilidade nos testes de touros (MEYER et al., 1989 e SWALVE, 1995). Com os TDM também é possível obter o encurtamento do intervalo de gerações (FERREIRA et al., 2003). Os TDM se tornam ainda mais importante em países que possuem rebanhos pequenos em que não há um esquema eficiente de coleta de dados de produção de leite, assim é possível utilizar todas as informações disponíveis na avaliação genética (BILAL AND KHAN, 2009).

A maior desvantagem dos TDM é a dificuldade computacional de analisar um elevado volume de dados e a necessidade de estimar muito mais parâmetros, uma vez que utilizando modelo multicaracterísticas completo cada PLDC é considerada como uma característica distinta, assim o número de efeitos no modelo aumenta linearmente com o número de controles considerados. Todavia, atualmente, existem diferentes abordagens para diminuir a dimensão dos dados, como a análise de fatores e análise de componentes principais (modelos de posto reduzido).

A análise de componentes principais tem como objetivo identificar os fatores que explicam a máxima quantidade de variação, e constitui na transformação de um conjunto de variáveis originais correlacionadas, num novo conjunto de variáveis as quais são combinações lineares das variáveis originais, mas não são correlacionadas entre si, eliminando assim as informações redundantes (KIRKPATRICK & MEYER, 2004). Por outro lado, a análise de fatores consiste em identificar as relações de covariância entre diversas variáveis em função de poucas, denominadas de fatores (MEYER, 2009). Na análise de fatores cada variável resposta é representada por uma função linear de uma pequena quantidade de fatores comuns, não observáveis, e de fatores específicos.

O objetivo do presente trabalho foi comparar as diferentes metodologias de análises multicaracterísticas, para a produção de leite no dia do controle de primeiras

lactações de vacas da raça Holandesa visando implementar a utilização destes modelos em programas de avaliação genética de bovinos de leite.

MATERIAL E MÉTODOS

Foram analisados 15.896 controles mensais de produção de leite de 1.820 primeiras lactações de vacas da raça Holandesa, pertencentes à Agropecuária Agrindus S.A., localizada no município de Descalvado, SP, com partos registrados no período de 1997 a 2005. As lactações foram truncadas aos 305 dias e os controles mensais de produção de leite, obtidos entre cinco e 305 dias após o parto, foram divididos em intervalos de, aproximadamente, 30 dias, totalizando dez controles (C1 a C10). Foi estabelecido que cada animal possuísse, pelo menos, sete controles individuais de produção na lactação, com o primeiro controle realizado até 45 dias após o parto. Os grupos de contemporâneos foram definidos como rebanho-ano-mês do controle, compondo 676 classes e, aplicou-se a restrição de que cada grupo deveria conter, no mínimo, oito animais. Um resumo da estrutura dos dados é apresentado na Tabela 1.

As análises foram realizadas por meio de sete modelos: multicaracterísticas padrão, três modelos de posto reduzido ajustando os primeiros 2,3 e 4 componentes principais genéticos e três modelos utilizando análise de fatores com 2,3 e 4 fatores. Para todos os modelos foram considerados os efeitos aleatórios genético aditivo e residual, e os efeitos sistemáticos do grupo de contemporâneo e da idade da vaca ao parto (efeito linear e quadrática) e do número de dias em lactação (efeito linear). A matriz de (co)variâncias residual, para todos os modelos, foi assumida ter posto completo.

Tabela 1 – Resumo da estrutura dos dados, número de animais (N), médias e desvios-padrão (DP) para as produções de leite no dia do controle (C1-C10) e médias e desvios-padrão de dias em lactação (DEL).

Controles	N	Produção de leite (kg)		DEL (dias)	
		Médias	DP	Médias	DP
C1	1695	24,13	5,49	16,19	3,14
C2	1789	29,81	6,49	44,88	1,13
C3	1755	31,23	6,21	74,93	1,07
C4	1724	31,84	6,19	104,77	1,46
C5	1661	32,06	6,39	134,87	1,86
C6	1582	31,37	6,46	164,90	1,69
C7	1506	30,39	6,96	194,86	1,70
C8	1455	29,56	6,85	224,88	1,93
C9	1415	27,94	7,30	254,94	1,58
C10	1314	25,41	7,35	284,07	3,41

O Modelo Multicaracterísticas (MC) utilizado pode ser representado em notação matricial como:

$$y = X\beta + Zu + e \quad (1)$$

em que y representa o vetor das observações para as 10 características (produções de leite no dia do controle), β é o vetor de soluções para os efeitos fixos; u é o vetor de soluções para os efeitos aleatórios genéticos aditivos; X e Z são as matrizes de incidência, para os efeitos fixos e aleatório de animal, respectivamente, e é o vetor do efeito aleatório residual, A o numerador da matriz de parentesco e com $Var(u) = \Sigma \otimes A$ onde Σ é a matriz de (co)variâncias.

As equações de modelos mistos para o MC podem ser representadas por:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + \Sigma^{-1} \otimes A^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix} \quad (2)$$

O Modelo de Componentes Principais (MCP) pode ser obtido pela reparametrização do modelo em (1), a qual produz um modelo equivalente que contempla os componentes principais. Tal modelo, em notação matricial, é representado pela expressão:

$$y = X\beta + Z^*u^* + e \quad (3)$$

em que $Z^* = Z(E \otimes I)$, $u^* = (E \otimes I)u$, $Var(u^*) = (I \otimes A)$ e $\Sigma = E\Lambda E'$. Nestas definições, E correspondendo a matriz dos autovetores (v) e Λ a matriz diagonal dos autovalores λ_i . Assim, a matriz de (co)variâncias genética (Σ) pode ser decomposta em termos de E e Λ , de forma que $EE' = I$. Considerando m componentes principais, a matriz E é substituída por E_m , a qual compreende as m primeiras colunas (v_1, \dots, v_m) de E , portanto, E_m é então utilizada para definir Z^* e u^* .

As equações de modelos mistos para o MCP são apresentadas em (3), sendo que o número de equações agora reduzido proporcionalmente a m , uma vez que Λ também é substituída por uma submatriz Λ_m , a qual consiste das m primeiros autovalores (KIRKPATRICK and MEYER, 2005). Tais equações são dadas por:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z^* \\ (Z^*)'R^{-1}X & (Z^*)'R^{-1}Z^* + \Lambda^{-1} \otimes A^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u}^* \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ (Z^*)'R^{-1}y \end{pmatrix} \quad (4)$$

No Modelo de Análise de Fatores (MAF), o efeito genético pode ser explicado por m fatores comuns (c) e q fatores específicos (s). Assume-se que a matriz de (co)variâncias para os efeitos genéticos é dada por $\Sigma = \Gamma\Gamma' + \Psi$, em que: Γ é a matriz de cargas fatoriais, cuja dimensão é $m \times q$; e Ψ é a matriz das variâncias dos efeitos específicos, sendo esta uma matriz diagonal de ordem m . O modelo em questão pode ser representado em notação matricial como:

$$y = X\beta + Z^+c + Z^os + e \quad (5)$$

em que: c e q representam, respectivamente, os fatores comuns e fatores específicos. Assim, tem-se: $Z^+ = Z^o(\Gamma \otimes I)$, $c = (I_N \otimes \Gamma' \Sigma^{-1})\mu$, $s = (I_N \otimes \Psi \Sigma^{-1})\mu$, $Var(c) = I$ e $Var(s) = \Psi \otimes I$ (THOMPSON et al., 2003 e MEYER, 2009). As equações dos modelos mistos para MAF podem ser representadas por:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}(Z^+) & X'R^{-1}X \\ (Z^+)'R^{-1}X & (Z^+)'R^{-1}(Z^+) + I_M \otimes A^{-1} & (Z^+)'R^{-1}Z^o \\ Z^o'R^{-1}X & Z^o'R^{-1}(Z^+) & Z^o'R^{-1}Z^o + \Psi^{-1} \otimes A^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{c} \\ \hat{s} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ (Z^+)'R^{-1}y \\ Z^o'R^{-1}Y \end{pmatrix} \quad (6)$$

Os componentes de (co)variâncias e os parâmetros genéticos foram estimados pelo método da máxima verossimilhança restrita (REML), utilizando o pacote Wombat (MEYER, 2006).

A comparação entre os modelos foi efetuada pelos critérios de informação de Akaike (AIC) e Bayesiano de Schwarz (BIC) (WOLFINGER, 1993). Os critérios de AIC e BIC permitem uma comparação entre os modelos não aninhados e penalizam aqueles com maior número de parâmetros, sendo que o BIC atribui uma penalidade mais rigorosa.

Os critérios de informação podem ser representados como:

$$\text{AIC} = -2\log L + 2p$$

$$\text{BIC} = -2\log L + p\log(N-r),$$

onde p é o número de parâmetros do modelo, N o número total de observações, r é o posto da matriz de incidência dos efeitos fixos no modelo e $\log L$ o logaritmo da máxima função de verossimilhança restrita. Menores valores para AIC e BIC indicam melhor ajuste.

Correlações entre classificações para os valores genéticos dos touros (correlação de posto) para as produções de leite no dia do controle e para a produção até 305 dias de lactação foram obtidas pela opção Spearman (SAS, 2002). O modelo para ajuste (P305) incluiu os efeitos fixos de grupo de contemporâneos (ano-estação-de-parto), época do parto e idade da vaca ao parto (efeito linear e quadrático), e o efeitos aleatório de animal e resíduo.

RESULTADOS E DISCUSSÃO

Os resultados das análises com os valores do logaritmo da função de verossimilhança ($\log L$), o critério de informação de Akaike (AIC) e o Bayesiano de Schwarz (BIC) estão apresentados na Tabela 2.

Tabela 2: Modelos, número de parâmetros (p) do modelo, função de verossimilhança ($2\log L$) e critérios de informação de Akaike (AIC) e Bayesiano de Schwarz (BIC).

Modelos*	p	Log L	AIC	BIC
MCP2	74	-31890	63928	64492
MCP3	82	-31886	63936	64562
MCP4	89	-31884	63946	64625
MC	110	-31883	63987	64826
MAF2	84	-31889	63946	64586
MAF3	92	-31886	63955	64657
MAF4	99	-31884	63966	64721

*MCPn: modelo ajustando n primeiros componentes principais, MC: modelo multicaracterísticas, MAFn: modelo análise de fatores com n fatores. Valores em negrito indicam o melhor modelo com base em AIC e BIC.

O modelo contendo o maior número de parâmetros (MC) promoveu o pior ajuste, de acordo com os dois critérios de comparação. Para os modelos que contém o número de fatores igual ao número de componentes principais, o valor de log L foi muito semelhante. Dentre os modelos avaliados, foi possível observar que os valores de ambos os critérios (AIC e BIC) diminuíram quando os números de parâmetros para os modelos MCP e MAF diminuíram. Entretanto nos MCP, os valores de AIC e BIC, assim como o número de parâmetros, foram sempre menores do que MAF.

Entre todos os modelos, o MCP2 foi o melhor de acordo com os critérios AIC e BIC e, além disso, esse modelo apresentou convergência mais rápida. Comparando-o ao MC, que por sua vez apresentou pior ajuste, há uma expressiva redução no número de parâmetros (36). Além disso, o modelo MCP2, apresentou convergência muito mais rápida que os modelos MC e MAF. MEYER (2009) também reportou que os MCP apresentaram convergência mais rápida que os MC e MAF.

As estimativas dos componentes de variância fenotípica, genética e residual para os modelos MC e MCP2, MAF2 estão apresentadas na Figura 1. As estimativas de variâncias fenotípicas para todos os modelos foram similares, sendo mais elevadas no final da lactação.

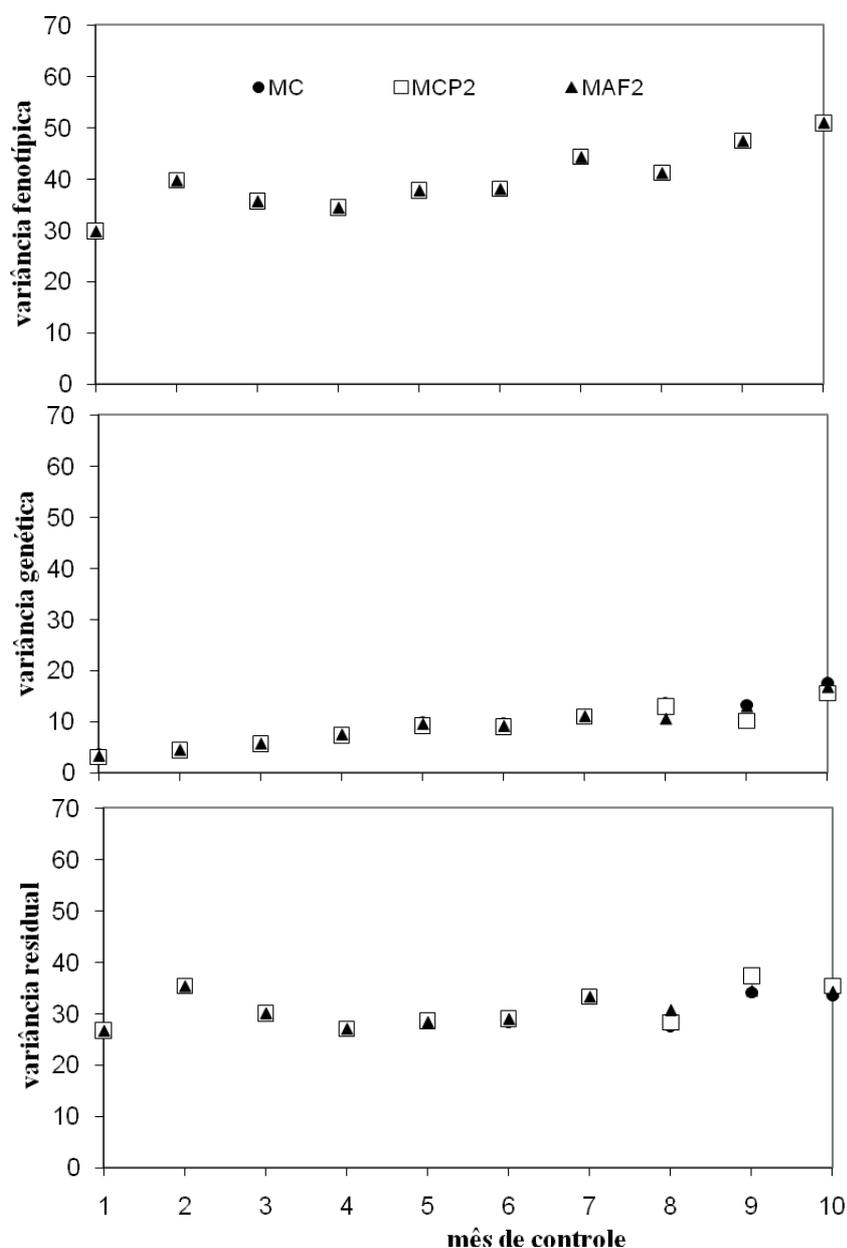


Figura 1. Estimativas dos componentes de variância fenotípica, genética, residual, para as produções mensais de leite, para o modelo multicaracterísticas (MC), para o modelo ajustando dois primeiros componentes principais (MCP2) e para o modelo ajustando dois fatores (MAF2).

Os maiores valores para as variâncias genéticas obtidas com os três modelos ocorreram no final da lactação. De modo geral, no final da lactação, tem-se menor número de registros de produção de leite, podendo ocorrer dificuldades no ajuste do

modelo, que resultam em estimativas de variâncias genéticas mais elevadas neste período. Como se pode observar, as variâncias genéticas obtidas com os três modelos foram praticamente idênticas até o 7º controle e mantiveram-se sem grandes oscilações do 8º até o último controle. As variâncias genéticas aditivas foram sempre menores que a residual e também constantes por todo o período de lactação. Tais tendências têm sido observadas na literatura por outros autores utilizando modelos de regressão aleatória em bovinos da raça Holandesa, dentre eles, JAMROZIK et al. (1997) e OLORI et al. (1999).

Em todos os modelos MAF as variâncias genéticas específicas obtidas foram responsáveis por pequena parte da variação genética. Para o modelo MAF2, as variâncias específicas contribuíram com somente com 0,01% e 0,03% da variação genética no 8º e 10º controle, respectivamente, e sendo praticamente zero nos demais. Segundo MEYER, (2009), se fatores específicos são assumidos como ausentes, o MAF pode ser reduzido para um MCP. Como no presente estudo os fatores específicos para os modelos MAF foram praticamente nulos, não seria vantajosa a utilização desses modelos.

As variâncias residuais apresentaram tendência semelhante à variância fenotípica, com pequenas oscilações entre as estimativas obtidas com os três modelos a partir do 8º mês de lactação.

As estimativas de herdabilidades obtidas com os modelos (MC e MCP2) dos controles mensais de produção aumentaram gradativamente até 8º controle da lactação, em seguida, diminuiu no 9º controle, e aumentaram novamente ao final da lactação (Figura 2). Assim como para as variâncias genéticas, as herdabilidades dos dois modelos foram praticamente iguais até o 7º controle e pequenas diferenças ocorreram após 8º controle.

As estimativas das herdabilidades oscilaram entre 0,12 e 0,34 e entre 0,11 e 0,31, respectivamente, para os modelos MC, MCP2. Valores semelhantes a esse foram relatados por OLORI et al. (1999) e KETTUNEN et al. (2000) e no Brasil por COBUCCI et al. (2005), ARAÚJO et al. (2006), MELO et al. (2007) e BIGNARDI et al. (2009a,b) utilizando MRA empregando diferentes funções em bovinos da raça Holandesa. MELO et al. (2005) e BIGNARDI et al. (2008) utilizando modelos uni-

característica para PLDC em bovinos da raça Holandesa no Brasil, também reportaram valores similares de herdabilidades que o presente estudo.

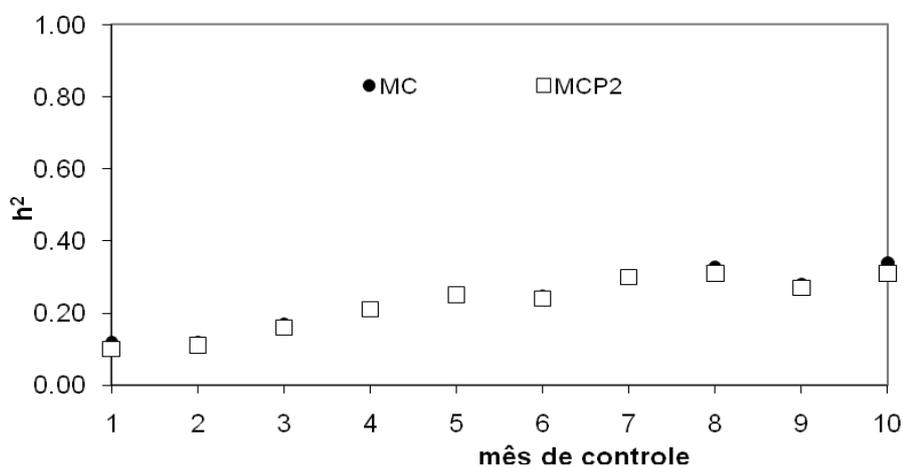


Figura 2. Estimativas de herdabilidades, para as produções mensais de leite, para o modelo multicaracterísticas (MC), para o modelo ajustando dois primeiros componentes principais (MCP2).

As estimativas de correlação genética obtidas com o modelo MCP2 foram superiores as obtidas pelo modelo MC, principalmente, nos controles adjacentes do final da lactação (Tabela 3). Nos dois modelos (MC e MCP2) as PLDC adjacentes no período intermediário da lactação apresentaram estimativas de correlações genéticas próximas a unidade. Estes resultados estão de acordo com aqueles relatados por REKAY A et al. (1999), OLORI et al. (1999) e no Brasil por COBUCI et al. (2005) e BIGNARDI et al., (2009a,b) em bovinos da raça Holandesa utilizando modelos de regressão aleatória. As correlações fenotípicas (não apresentadas) tanto para MC quanto para (MCP2) apresentaram a mesma tendência das correlações genéticas, com valores superiores entre os controles adjacentes. As correlações residuais foram inferiores às demais.

Pela decomposição canônica da matriz de covariância genética do MC, os seguintes autovalores são produzidos 83,95; 11,00; 2,97; 1,02; 0,43; 0,12; 0,00; 0,00; 0,00; 0,00, assim, o primeiro autovalor contabiliza grande parte da variância genética

(84,38%). Para o modelo MCP2 uma parte maior da variância genética é explicada pelo primeiro autovalor (89,09%).

Tabela 3 - Estimativas de correlações genéticas para as produções de leite no dia do controle (C1 a C10), para o modelo multicaracterísticas (acima da diagonal) e para o modelo ajustando dois primeiros componentes principais (abaixo da diagonal).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	-	0,90	0,80	0,74	0,63	0,48	0,55	0,40	0,28	0,12
C2	0,91	-	0,95	0,94	0,87	0,73	0,82	0,74	0,64	0,46
C3	0,88	1,00	-	0,97	0,96	0,82	0,89	0,76	0,67	0,57
C4	0,77	0,97	0,98	-	0,96	0,92	0,97	0,87	0,81	0,71
C5	0,77	0,97	0,98	1,00	-	0,90	0,95	0,85	0,78	0,71
C6	0,58	0,87	0,90	0,97	0,97	-	0,97	0,87	0,85	0,90
C7	0,61	0,88	0,91	0,97	0,98	1,00	-	0,93	0,91	0,84
C8	0,39	0,73	0,78	0,89	0,89	0,98	0,97	-	0,64	0,85
C9	0,27	0,64	0,69	0,82	0,82	0,94	0,93	0,99	-	0,88
C10	0,17	0,56	0,62	0,76	0,77	0,91	0,89	0,98	1,00	-

São apresentadas na Tabela 4, as correlações de ordem entre os valores genéticos preditos para a produção acumulada até 305 dias (P305), pelo método tradicional e as produções de leite no dia do controle, estimadas pelo MC e pelo MCP2, para todos os touros do rebanho (281). Comparando ambos os modelos (MC e MCP2) ao modelo padrão (P305), as correlações de ordem foram de moderada a alta, isso mostra que ocorre semelhança nas classificações dos touros entre esses modelos, podendo haver algumas diferenças no primeiro (C1) e no último controle (C10) com a P305. Como pode ser observado, não houve muita diferença na magnitude das correlações de ordem entre os valores genéticos usando os modelos MC e MCP2 na modelagem das produções de leite no dia do controle, pois todas as correlações de ordem foram próximas. Assim, a utilização do MCP2 levaria à seleção de praticamente os mesmos touros que no MC.

Assim como no presente estudo, MEYER (2007) utilizando modelo multicaracterísticas, análise de fatores e análise de componentes principais, para estudar 14 características de carcaça em bovinos Angus, observou que os componentes principais também promoveram um melhor ajuste na estrutura de covariâncias genéticas em comparação às demais abordagens.

Um aspecto importante a ser considerado em avaliações genéticas de um grande número de animais e características como programas de avaliação nacionais de leite, é o requerimento computacional. Em bovinos de leite as PLDC têm substituído a produção acumulada até 305 dias (P305). Além das PLDC, outras características como a percentagem de proteína, a percentagem gordura, os sólidos totais e contagem de células somáticas também vêm sendo incorporadas aos programas. O emprego de modelos de posto reduzido ao invés de modelos multicaracterísticas padrão, em programas de avaliação genética que utilizam a análise simultânea de várias características, seria muito vantajoso, uma vez que, esses modelos necessitam de uma menor demanda computacional, resultando em grande economia de tempo e, conseqüentemente, facilitando as análises de rotina.

Tabela 4 - Correlações de ordem entre os valores genéticos preditos para a produção acumulada até 305 dias (P305) e as produções de leite no dia do controle (C1-C10), pelo modelo multicaracterísticas padrão (MC) e pelo modelo ajustando dois componentes principais (MCP2) para todos os touros do rebanho, e entre MC e MCP2

Controles	MC – P305	MCP2 – P305	MC-MCP2
C1	76,63	74,19	96,82
C2	91,48	89,29	97,58
C3	93,24	92,94	99,68
C4	96,35	96,13	99,78
C5	94,50	93,67	99,12
C6	91,51	89,63	97,95
C7	96,85	91,46	99,95
C8	97,70	96,84	99,12
C9	94,82	93,25	98,35
C10	88,27	85,73	97,13

CONCLUSÃO

Os resultados indicam que somente dois componentes principais são requeridos para modelar a estrutura de (co)variâncias genéticas entre as produções de leite no dia do controle. Além disso, o modelo de posto reduzido diminui consideravelmente o número de parâmetros, sem reduzir a qualidade de ajuste.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO, C. V.; TORRES, R. A.; COSTA, C. N.; et. al. Uso de modelos de regressão aleatória para descrever a variação genética da produção de leite na raça Holandesa. **Revista Brasileira de Zootecnia**, vol.35, n.3, supl, p.975-981, 2006.

BIGNARDI, A.B.; EL FARO, L.; ALBUQUERQUE, L.G.; CARDOSO, V.L.; MACHADO, P.F. Modelos de dimensão finita para a estimação de parâmetros genéticos para a produção de leite de primeiras lactações de vacas da raça Holandesa. **Ciência Rural** v.38, p.1705–1710, 2008.

BIGNARDI, A.B.; EL FARO, L.; ALBUQUERQUE, L.G.; CARDOSO, V.L.; MACHADO, P.F. Random regression models to estimate test-day milk yield genetic parameters Holstein cows in southeastern Brazil. **Livestock Science**, v.123, p.1-7, 2009a.

BIGNARDI, A.B.; EL FARO, L.; ALBUQUERQUE, L.G.; CARDOSO, V.L.; MACHADO, P.F. Parametric correlation functions to model the structure of permanent environmental (co)variances in milk yield random regression models. **Journal Dairy Science**, v.92, p.4634–4640, 2009b.

BILAL, G.; KHAN, M. S.; Use of test-day milk yield for genetic evaluation in dairy cattle: a review. **Pakistan Veterinary Journal**, v.29 p.35-41, 2009.

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

FERREIRA, W. J.; TEIXEIRA, N. M.; EUCLYDES, R. F. et al. Avaliação genética de bovinos da raça holandesa usando a produção de leite no dia do controle. **Revista Brasileira de Zootecnia**, vol.32, n.2, p.295-303, 2003.

JAMROZIK, G. J.; KISTEMAKER, J.C.M.; DEKKERS; SCHAEFFER, L. R. Comparison of possible covariates for use in a random regression model for analyses of test day yields. **Journal Dairy Science**, v.80, p.2550-2556, 1997.

KETTUNEN, A.,MÄNTYSAARI, E., PÖSÖ J. Estimation of genetic parameters for daily milk yield of primiparous Ayrshire cows by random regression test-day models. **Livestock Production Science**,v.66, p.251-261, 2000.

KIRKPATRICK, M.; MEYER, K. Direct estimation of genetic principal components: simplified analysis of complex phenotypes. **Genetics**. v.168, p.2295–2306, 2004.

KIRKPATRICK, M.; MEYER, K. Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. **Genetics Selection Evolution** v.21, p.1-30, 2005.

MELO, C.M.R; COSTA, C.N; PACKER, U.I.; MACHADO, P.F. Parâmetros genéticos para as produções de leite no dia do controle e da primeira lactação de vacas da raça Holandesa. **Revista Brasileira de Zootecnia** v.34, p.796-806, 2005.

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**, v.1, p.325–334, 2007.

MEYER, K.; GRASER, H.U.; HAMMOND, K. Estimates of genetic parameters for first lactation test day production of Australian Black and White cows. **Livestock Production Science**, v.21, n.3, p.177-199, 1989.

MEYER K. "WOMBAT" – digging deep for quantitative genetic analyses using restricted maximum likelihood. CD ROM Eighth World Congress Genetics Applied Livestock Production, **Proceedings....** Communication No. 27-14, 2006.

MEYER, K. Multivariate analyses of carcass traits for Angus cattle fitting reduced rank and factor analytic models. **Journal Animal Breeding Genetics**, v. 124, p.50-63, 2007.

MEYER, K. Factor-analytic models for genotype \times environment type problems and structured covariance matrices. **Genetics Selection Evolution** v.41, p.21-32, 2009.

OLORI, V. E.; HILL W. G.; MCGUIRK, B. J.; BROTHERSTONE S. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. **Livestock Production Science**, v.61, p.53–63, 1999.

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.

SAS INSTITUTE, **SAS/STAT Software**: changes and enhancement through release 8.2. Cary, 2002.

SWALVE, H.H. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. **Journal of Dairy Science**, v.78, n.4, p.929-938, 1995.

THOMPSON, R.; CULLIS, B.R.; SMITH, A.B; GILMOUR, A.R.; A sparse implementation of the Average Information algorithm for factor analytic and reduced rank variance models. **Australian New Zealand Journal Statistics**, v.45, p.445-459, 2003.

WOLFINGER, R. Covariance structure selection in general mixed models. **Communications in Statistics**, v.22, n.4, p.1079-1106, 1993.

CAPÍTULO 3 - PARÂMETROS GENÉTICOS PARA PRODUÇÃO DE LEITE EM VACAS DA RAÇA HOLANDESA EMPREGANDO MODELOS DE REGRESSÃO ALEATÓRIA UTILIZANDO FUNÇÕES B-SPLINES E AMOSTRAGEM DE GIBBS

RESUMO – Foram analisados 152.145 controles semanais de produção de leite de 7.317 primeiras lactações de vacas da raça Holandesa, provenientes de rebanhos da região Sudeste do Brasil. As produções de leite no dia do controle (PLDC) foram consideradas em 44 classes semanais de dia em lactação. Os grupos de contemporâneos foram definidos como rebanho-ano-semana do controle compondo 2.539 classes e, contendo, no mínimo, seis animais. O modelo utilizado incluiu, os efeitos aleatórios genético aditivo direto, de ambiente permanente e o residual. Foram considerados como efeitos fixos, o grupo de contemporâneos, os efeitos linear e quadrático da covariável idade da vaca ao parto e a curva média de lactação da população, modelada por meio de polinômios ortogonais de Legendre de quarta ordem. Os efeitos aleatórios genético aditivo e de ambiente permanente foram modelados por meio de funções B-splines e inferência bayesiana. A estrutura residual foi considerada heterogênea, contendo 6 classes de variâncias. O modelo empregando B-spline cúbica com o número de coeficientes de regressão aleatória igual a oito tanto para o efeito aleatório genético aditivo como de ambiente permanente foi o mais adequado de acordo com os critérios de comparação. Entretanto o modelo empregando B-spline quadrática poderia ser adotado, uma vez que, foram constatadas poucas mudanças nos parâmetros genéticos estimados em relação ao modelo B-spline cúbica e, além disso, as estimativas a posteriori apresentaram menores intervalos de credibilidade nos pontos extremos da curva de lactação.

Palavras-chave: Funções de covariância, Parâmetros genéticos, Splines

INTRODUÇÃO

Em vários trabalhos têm sido apresentadas estimativas de parâmetros genéticos para a produção de leite no dia do controle aplicando modelos de regressão aleatória (MRA) com base nos polinômios de Legendre (EL FARO et al., 2003; COBUCCI et al., 2005; ARAÚJO et al., 2006; EL FARO et al., 2008; BIGNARDI et al., 2009a). Um grande problema desses modelos é a necessidade de utilização de polinômios de alto grau e a desuniformidade de ajuste ao longo da curva. Alternativas para diminuir o grau dos polinômios estão sendo estudadas e uma delas é a utilização de funções “spline” ou também chamados de polinômios segmentados (MEYER, 2005a; SILVESTRE et al., 2006; BOHMANOVA et al., 2008).

“Splines” são curvas constituídas por segmentos individuais de polinômios de grau reduzido e que se unem em pontos específicos, denominados nós (De BOOR, 1978). As funções “splines” caracterizam-se por atender quase integralmente as propriedades ótimas de uma função: possuir interpretação biológica simples, ser de fácil estimação (linear nos parâmetros) e redução da multicolinearidade (FULLER, 1969). Entretanto, o ajuste dependerá da escolha dos graus do polinômio e os locais onde os nós serão inseridos. Um tipo particular de função “splines” chamada B-spline é preferida a outros tipos de “spline”, devido suas ótimas propriedades numéricas (RUPERT et al., 2003; EILERS & MARX, 2005) e a sua utilização é adequada para modelar efeitos aleatórios em análises de modelos mistos, e também eficientes na estimação de funções de covariância (RICE & WU, 2001).

Segundo MISZTAL (2006), os MRA que empregam funções splines tendem a ser bastante afetados pela distribuição dos dados, do grau das funções e, mais ainda, pela escolha do número e da posição dos nós. RUPERT et al. (2003) recomendam que em modelos empregando splines penalizadas, a melhor escolha da posição dos nós seria em intervalos equidistantes. Conforme relatado por MEYER (2005b), o emprego de funções B-splines, que é um tipo de spline penalizada, consegue atenuar a influência da posição dos nós sobre as estimativas. MEYER (2005a), modelando a curva de crescimento em bovinos de corte com MRA

empregando funções B-splines, relatou que não houve diferenças nas estimativas dos parâmetros genéticos entre modelos com nós inseridos em posições eqüidistantes ou não eqüidistantes.

Uma das opções para estimação dos coeficientes de funções de covariância são os métodos Bayesianos, com base em Cadeia de Markov Monte Carlo (MCMC). A abordagem Bayesiana é capaz de produzir com exatidão as distribuições conjuntas e marginais posteriores dos parâmetros que compõem a função (BOX & TIAO, 1973). Na estatística Bayesiana não existe distinção entre efeitos fixos e aleatórios, todos os efeitos são consideradas como aleatórios. Efeitos considerados fixos no melhoramento animal (rebanho, ano, etc) são tidos, neste método, como efeitos aleatórios sobre as quais se tem pouco ou nenhum conhecimento a priori, ou seja, a priori o pesquisador é indiferente ao provável valor destes efeitos (GIANOLA & FERNANDO, 1986).

A aplicação dos métodos MCMC, dentro os quais se destaca a amostragem de Gibbs, está sendo muito utilizada no melhoramento animal (JAMROZIK et al., 2001; LÓPEZ-ROMERO et al., 2003; MEYER, 2005a). De acordo com CASELLA & GEORGE (1992), o amostrador de Gibbs é um procedimento numérico para obter amostra da distribuição conjunta ou marginal de todos os parâmetros do modelo, por repetidas amostragens de todas as distribuições condicionais completas formando uma cadeia de Markov. Segundo VAN TASSEL et al. (1995), o uso do método amostrador de Gibbs na estimação de componentes de covariância, apresenta várias vantagens em relação aos métodos usuais que empregam máxima verossilhança restrita (REML): não requer soluções para as equações de modelo misto; propicia estimativas diretas e acuradas dos componentes de (co)variância para essas estimativas.

O objetivo desse estudo foi modelar a variação da produção de leite no dia do controle de primeiras lactações de vacas da raça Holandesa por meio de modelos de regressão aleatória utilizando funções B-splines e inferência bayesiana, visando obter modelos adequados e parcimoniosos para estimar parâmetros genéticos.

MATERIAL E MÉTODOS

Foram analisados 152.145 controles semanais de produção de leite de 7.317 primeiras lactações de vacas da raça Holandesa, filhas de 612 touros, provenientes de 93 rebanhos dos estados de São Paulo e Minas Gerais, no Brasil. Os dados foram provenientes do Programa de Análise de Rebanhos Leiteiros e da Qualidade do Leite da ESALQ - USP, no período de 1996 a 2003. As produções de leite foram consideradas em classes semanais de dias em lactação, variando de 1 até 44 classes. Foi estabelecido que cada animal possuísse, pelo menos, cinco controles individuais de produção na lactação, com o primeiro controle realizado até 45 dias após o parto. Os grupos de contemporâneos foram definidos como rebanho-ano-semana do controle, compondo 2.539 grupos e, aplicou-se a restrição de que cada um deveria conter, no mínimo, seis animais. O arquivo de pedigree continha 17.688 animais na matriz de parentesco.

As análises foram realizadas por meio de modelos uni-características de regressão aleatória. Para todos os modelos foram incluídos os efeitos aleatórios genético aditivo e de ambiente permanente de animal e o resíduo. Os efeitos aleatórios genético aditivo e de ambiente permanente foram modelados por meio de funções B-splines. Foram considerados como efeitos fixos, o grupo de contemporâneos, os efeitos linear e quadrático da covariável idade da vaca ao parto e a curva média de lactação da população, esta modelada por meio de um polinômio ortogonal de quarta ordem. O resíduo foi considerado heterogêneo, contendo 6 classes de variâncias, sendo que as semanas de lactação foram agrupadas da seguinte forma: 1, 2, 3, 4-6, 7-12, 13-44 semanas.

As B-splines de grau $p=0$ têm valores iguais à unidade para todos os pontos (t) em um dado intervalo, e zero caso contrário. Para o k -ésimo intervalo dado por nós T_k e T_{k+1} com $T_k \leq T_{k+1}$:

$$B_{k,0}(t) = \begin{cases} 1 & \text{se } T_k \leq t < T_{k+1}, \\ 0 & \text{caso contrário} \end{cases}$$

As funções bases de grau $p > 0$ podem ser representadas por:

$$B_{k,p}(t) = \frac{t - T_k}{T_{k+p} - T_k} B_{k,p-1}(t) + \frac{T_{k+p+1} - t}{T_{k+p+1} - T_{k+1}} B_{k+1,p-1}(t)$$

Foram considerados polinômios linear (L), quadrático (Q) e cúbico (C) para cada segmento individual, com funções bases, respectivamente, de graus $p=1, 2$ e 3 . Polinômios de mesmo grau foram considerados no modelo para os efeitos aleatórios genético e de ambiente permanente. Os m nós foram escolhidos para dividir as semanas de lactação em $m-1$ intervalos equidistantes e os nós externos foram localizados na 1ª semana e 44ª semana de lactação para todos os modelos. Até sete nós foram utilizados por efeito, e o mesmo número de nós foram ajustados tanto para o genético aditivo quanto para o ambiente permanente. Os números de coeficientes de regressão aleatória para modelar a trajetória das funções bases linear, quadrática e cúbica, são dados, respectivamente, por $m, m+1$ e $m+2$.

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.

Para a análise utilizando inferência bayesiana pressupõem-se que:

- $y | b, a, c, \sigma_{e1}^2, \dots, \sigma_{e6}^2 \sim NMV(Xb + Za + Wc, R)$, com $R = \text{diag} \{ \sigma_{el}^2 \}$ onde $l = 1, \dots, 6$, é o número de classes do resíduo.

As pressuposições acerca das distribuições *a priori* para os parâmetros desconhecidos são consideradas como a seguir, como descrito por REKAYA et al. (1999):

- $b \propto$ 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;
 - $K_a | V_a, S_a^2 \sim W^{-1}(v_a, v_a S_a^2)$;
 - $K_c | V_c, S_c^2 \sim W^{-1}(v_c, v_c S_c^2)$, em que v_a, S_a^2 e v_c, S_c^2 representam o grau de liberdade e os valores *a priori* para as covariâncias dos coeficientes de regressão genético aditivo e de ambiente permanente, respectivamente.
 - $\sigma_{el}^2 | v_{el}, s_{el}^2 \sim X^{-2}(v_{el}, v_{el} s_{el}^2)$,

em que NMV , W^{-1} e X^{-2} representam, respectivamente, às distribuições normal multivariada, Wishart invertida e qui-quadrado invertida escalonada. Maiores detalhes sobre as distribuições condicionais completas a posteriori foram descritos por JAMROZIK & SCHAEFFER (1997) e REKAYA et al. (1999).

Os componentes de (co)variâncias e os parâmetros genéticos dos modelos foram obtidas 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. Uma única cadeia composta por 400.000 amostras foi gerada para cada análise.

O período de *burn-in* para obtenção das densidades marginais a posteriori foi determinado pelo critério de RAFTERY & LEWIS (1992) por meio do pacote computacional "BOA" (*Bayesian Output Analysis*) do programa R (R *Development Core Team*, 2008). Esse critério estima quantas iterações são necessárias para que o amostrador de Gibbs apresente convergência à distribuição estacionária, propondo especificações de quantidades necessárias para isto. O teste de HEIDELBERGER & WELCH (1983) e de GEWEKE (1992), disponíveis no pacote BOA, confirmaram a

convergência e também que a estacionariedade das cadeias de Gibbs foi alcançada. A autocorrelação entre as amostras foi baixa para todas as cadeias. Para todos os parâmetros o descarte das 10.000 primeiras amostras foi suficiente e a tomada de pontos amostrais em uma a cada 50 iterações.

Os diferentes modelos foram comparados pelo quadrado médio do resíduo (QMR) e por meio das estimativas de variância residual (JAMROZIK & SCHAEFFER, 2002). Utilizaram-se as médias ponderadas (pelo número de semanas em cada classe) das estimativas das variâncias residuais (VRP) de acordo com a expressão:

$$VRP = \frac{1}{TS} \sum_1^f \hat{\sigma}_{el}^2 NS_l,$$

em que TS = número total de semanas em lactação (44 semanas), $\hat{\sigma}_{el}^2$ = variância residual estimada para $l^{ésima}$ classe; f = número de classes de resíduos para cada modelo (6 classes); NS_l = número de semanas existente na $l^{ésima}$ classe.

A citação dos MRA segue o padrão: “BSX k”, em que X=L, Q e C são os graus do segmento polinomial linear, quadrático e cúbico, respectivamente, k especifica o número de coeficientes de regressão aleatória tanto para o efeito genético aditivo como para o efeito de ambiente permanente. Portanto, um modelo tendo a seguinte representação BSL4: constará de uma função B-spline linear, o número de coeficientes de regressão aleatória igual 4 tanto para o efeito aleatório genético aditivo como o de ambiente permanente, com 2 segmentos e 4 nós, uma vez que existem dois nós em pontos extremos, na primeira e 44^a semana.

RESULTADOS E DISCUSSÃO

A produção de leite média no dia do controle foi de 27,45 kg, com desvio padrão de 7,61 kg e coeficiente de variação de 27,73%. Houve aumento da produção da fase inicial (18,19 kg) até a 21^a semana de lactação (29,35 kg), decrescendo a partir daí até o final (23,63 kg), de acordo com a Figura 1. Houve maior variação da produção de leite no início (38,59%) e no final da lactação (32,52%). Como esperado, o número de animais foi menor no período final da lactação.

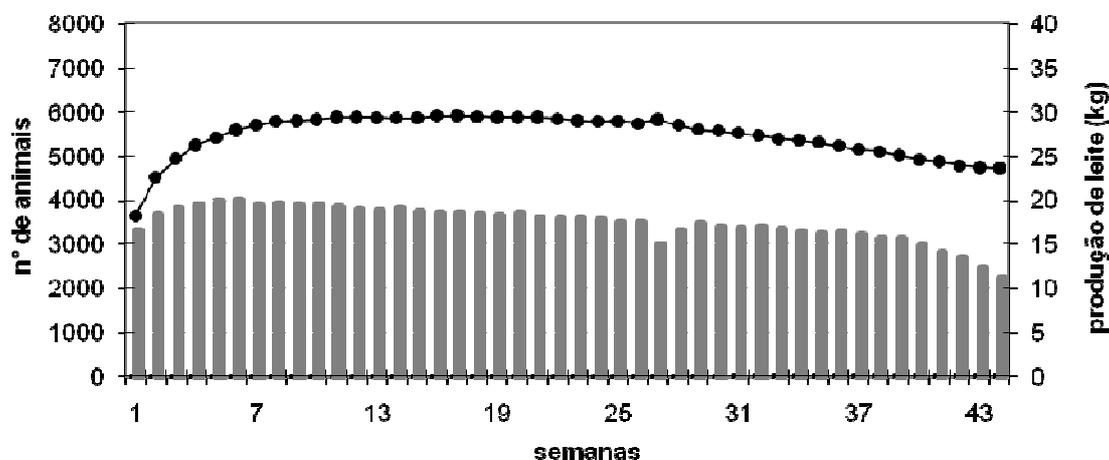


Figura 1. Número de observações (barra) e produção semanal de leite (linha).

Os quadrados médios do resíduo (QMR) e as estimativas de variância residual ponderada (VRP) para todos os modelos estão sumarizados na Tabela 1. Os modelos que possuem menor número de parâmetros foram os que apresentaram pior ajuste de acordo com QMR e VRP. Independente da ordem da B-spline (linear, quadrática ou cúbica) empregada, quando se aumentou nos modelos o número de nós tanto o QMR como o VRP apresentaram valores inferiores. Comparando os modelos com mesmo número de parâmetros, aqueles de ordem cúbica com menor número de segmentos mostraram se sempre superiores de acordo com QMR e VRP.

Entre todos os modelos estudados, o BSC8 foi o melhor de acordo com QMR e VRP. Uma vez que a diferença dos valores de QMR e VRP foi muito pequena entre o melhor modelo (BSC8) e o segundo melhor modelo (BSQ8), que continha o mesmo número de parâmetros e um maior número de segmentos, optou-se em apresentar e comparar as estimativas dos parâmetros genéticos dos mesmos.

Tabela 1. Modelos, número de parâmetros (n) do modelo, posição dos nós (pn) nas semanas, números de segmentos (ns), quadrado médio do resíduo (QMR) e estimativas das variâncias residuais (VRP).

Modelos *	n	pn	ns	QMR	VRP
BSL4	26	1,16,30,44	3	13,11	14,16
BSL5	36	1,12,23,34,44	4	12,36	13,56
BSL6	48	1,10,19,26,35,44	5	11,65	12,95
BSQ4	26	1,23,44	2	13,13	14,21
BSQ5	36	1,16,30,44	3	12,06	13,30
BSQ6	48	1,12,23,34,44	4	11,27	12,55
BSQ7	62	1,10,19,26,35,44	5	10,51	11,92
BSQ8	78	1,9,16,23,30,37,44	6	10,19	11,63
BSC5	36	1,23,44	2	11,89	13,14
BSC6	48	1,16,30,44	3	11,12	12,52
BSC7	62	1,12,23,34,44	4	10,51	11,89
BSC8	78	1,10,19,26,35,44	5	10,11	11,45

*BSXk: em que X=L, Q e C são os graus do segmento polinomial, k especifica o número de coeficientes de regressão aleatória tanto para o efeito genético aditivo como para o efeito de ambiente permanente.

As variâncias fenotípicas, genéticas e ambiente permanente dos dois melhores MRA (BSQ8 e BSC8) indicados pelos critérios QMR e VRP são apresentadas na Figura 2. As variâncias fenotípicas nos dois modelos foram altas nas duas primeiras semanas de lactação, diminuíram até o meio da lactação e voltaram a aumentar a partir da 20^a semana até o final da lactação.

As variâncias genéticas para ambos os modelos também foram maiores nas duas primeiras semanas de lactação, diminuíram até o meio da lactação e voltaram a aumentar a partir da 25^a semana diminuindo no final da lactação. Houve uma queda acentuada nas estimativas variâncias genéticas obtidas por esses dois modelos, até os 30 dias de lactação. Fato semelhante também foi observado por REKAYA et al. (1999) e KETTUNEN et al. (2000) e no Brasil por COBUCI et al. (2005), utilizando modelos de regressão aleatória empregando diferentes funções, em bovinos da raça Holandesa. Durante o pico da lactação (16^a e 17^a semana) observou-se menor variabilidade genética.

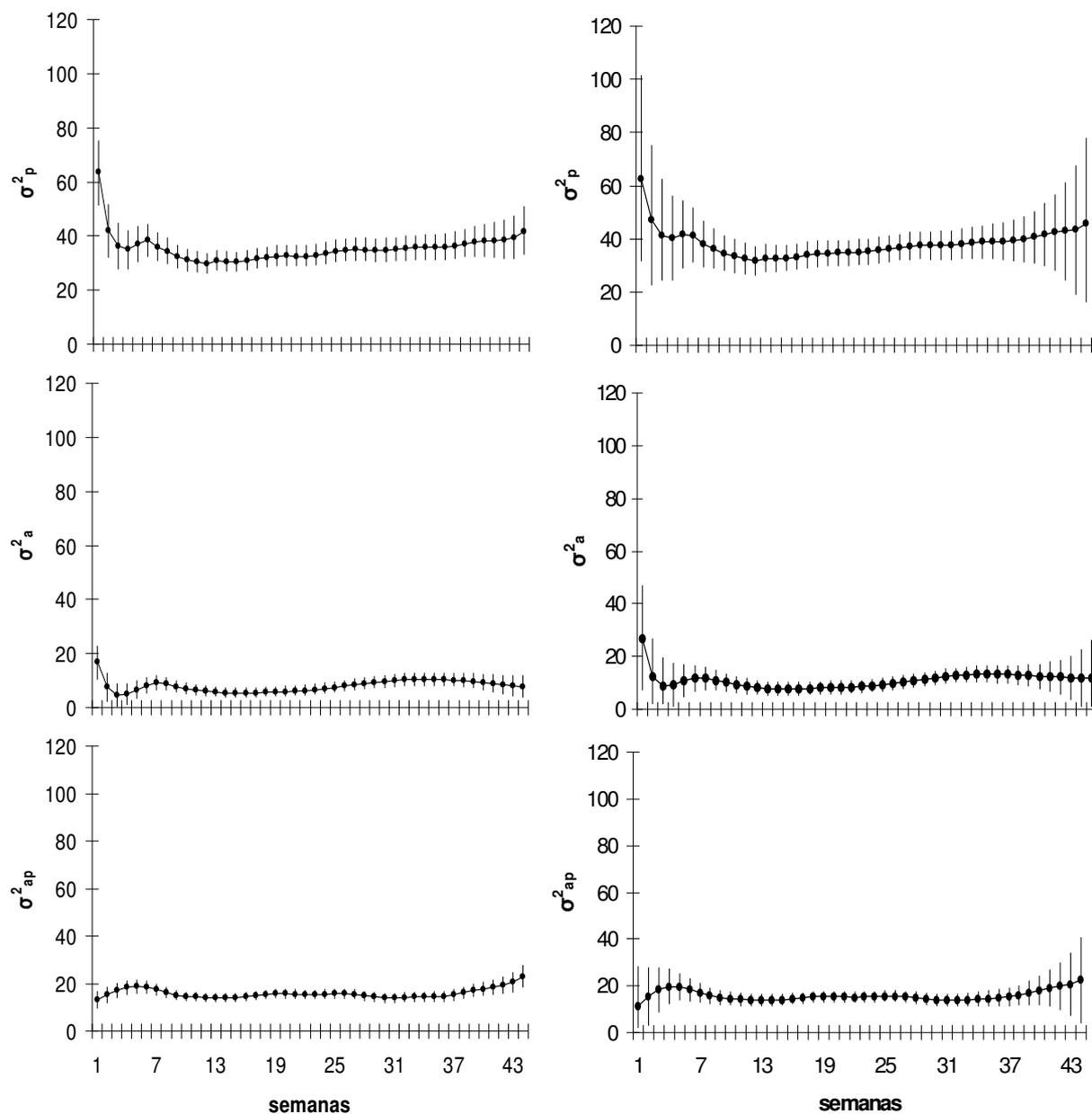


Figura 2 – Médias a posteriori (\bullet) das variâncias fenotípicas (σ^2_p), genéticas (σ^2_a) e ambiente permanente (σ^2_{ap}) e 95% de região de maior densidade a posteriori (barras) do modelo BSQ8 (à esquerda) e modelo BSC8 (à direita).

As variâncias genéticas aditivas foram sempre menores que as demais por todo o período de lactação. Tais tendências foram observadas na literatura em bovinos da raça Holandesa por outros autores, dentre eles, JAMROZIK et al. (1997) e OLORI et al. (1999).

As variâncias de ambiente permanente obtidas com os dois modelos foram praticamente constantes até a 37ª semana de lactação aumentando levemente à partir dessa semana até o final da lactação. Entretanto COBUCCI et al. (2005) empregando função de Wilmink e COSTA et al. (2008) utilizando polinômios de Legendre em bovinos da raça Holandesa no Brasil, reportaram maiores estimativas no final da lactação. Comparando com os resultados obtidos por BIGNARDI et al. (2009b), para o mesmo conjunto de dados, mas utilizando método da máxima verossimilhança restrita e MRA empregando polinômios de Legendre para efeito genético e uma função de correlação paramétrica estacionária associada a uma função de variância de ordem sete para efeito de ambiente permanente, pode-se concluir que os modelos BSQ8 e BSC8 proporcionaram estimativas de variâncias de ambiente permanente de magnitude inferiores praticamente em todo o período da lactação.

As estimativas residuais obtidas com o modelo BSQ8 (24,35; 19,14; 14,30; 11,34; 9,50 e 11,10 kg²), foram inferiores as obtidas com o modelo BSC8 (33,27; 18,87; 14,63; 11,73; 9,49 e 11,02 kg²), principalmente no início da lactação. Para os dois modelos as maiores estimativas de variâncias residuais foram no início da lactação. Resultados semelhantes foram reportados por OLORI et al. (1999) e por BIGNARDI et al. (2009a) empregando MRA ajustado com polinômios de Legendre e por BOHMANOVA et al. (2008) empregando função spline linear, em bovinos da raça Holandesa. Segundo LOPEZ-ROMERO et al. (2003), a heterogeneidade de variâncias residuais está relacionada com o estágio da lactação, e devido a um conjunto de fatores não especificados no modelo, tais como: características do período seco, estágio de prenhez, condição corporal ao parto, entre outros, fazem com que a variância residual seja maior nos extremos da lactação.

As estimativas de herdabilidade obtidas com os dois modelos apresentaram a mesma tendência durante a maior parte da lactação, diferindo somente no primeiro controle onde o modelo BSC8 apresentou estimativa superior (Figura 3). Após um decréscimo no início, houve aumento que resultou em maiores valores na fase intermediária da lactação, seguido de novo decréscimo até o final da lactação. Essa

diminuição no final da lactação nas estimativas de herdabilidade está associada ao aumento da variância de ambiente permanente e residual, promovendo um aumento na variância total e, conseqüente, uma diminuição da contribuição proporcional da variância genética aditiva.

Comparando com os resultados obtidos por BIGNARDI et al. (2009a) com MRA, empregando polinômios de Legendre para o mesmo conjunto de dados, as estimativas de herdabilidade oscilaram mais do que o modelo empregando função B-spline. Isso corrobora com o observado por MEYER (2005a), que estudando o crescimento em gado de corte relatou que as funções B-splines foram menos susceptíveis a problemas de estimativas erráticas nos extremos da curva, freqüentemente observados com polinômios de Legendre. Esta menor susceptibilidade se deve ao fato dos modelos utilizando B-splines conter segmentos individuais de polinômios de menor grau e melhor controle da influência global das observações individuais.

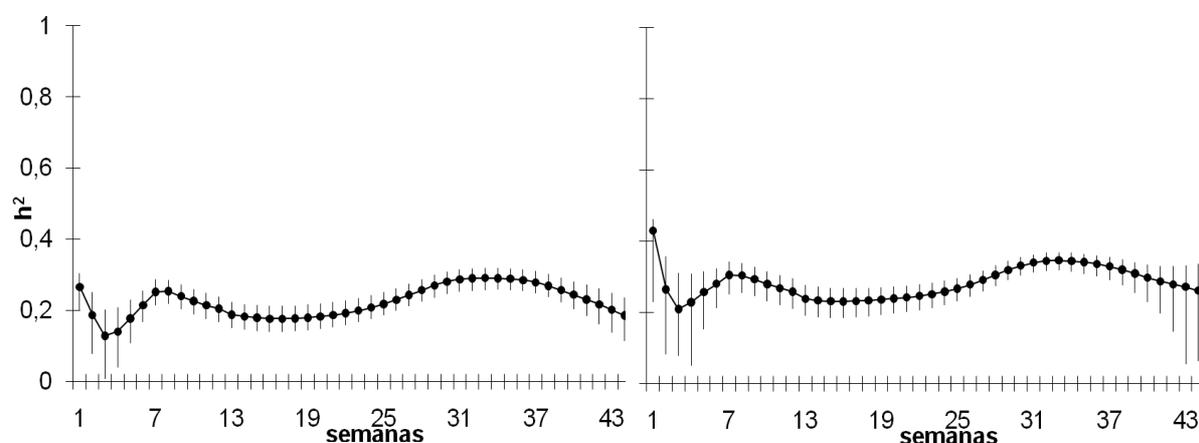


Figura 3 – Médias a posteriori das herdabilidades (•), 95% de região de maior densidade a posteriori (barras) do modelo BSQ8 (à esquerda) e modelo BSC8 (à direita).

As médias a posteriori das herdabilidades estimadas variaram de 0,21 (4ª semana) a 0,35 (34ª semana) com o modelo BSQ8, e de 0,21 (4ª semana) a 0,43 (1ª semana) com o modelo BSC8. Resultados semelhantes em bovinos da raça Holandesa foram encontrados por WHITE et al. (1999) empregando uma função

“spline” cúbica e por COBUCCI et al. (2005), ARAÚJO et al. (2006), COSTA et al. (2008) e BIGNARDI et al. (2009a,b), utilizando MRA empregando diferentes funções, em dados de produção de leite de no dia do controle no Brasil. Entretanto, SILVESTRE et al. (2005) utilizando MRA empregando uma função “spline” cúbica reportaram estimativas de herdabilidade de menor magnitude.

Tanto para as estimativas de todos os componentes de variância, quanto para as de herdabilidade, o modelo BSQ8, no período inicial e final da lactação, resultou em menor intervalo de credibilidade que o BSC8, o que sugere maior confiança nas estimativas fornecida pelo modelo BSQ8 nestes períodos. Isso indica que o modelo BSQ8 é mais eficaz no ajuste da produção de leite nos pontos extremos da lactação, conhecidamente, como os períodos mais problemáticos de serem modelados.

As estimativas de correlações fenotípicas e genéticas obtidas pelo modelo BSC8, em determinados dias de lactação, estão apresentadas na Tabela 3. As correlações fenotípicas foram mais altas entre as produções de leite nos dias de controle mais próximos, especialmente na fase intermediária da lactação.

As produções de leite adjacentes no período intermediário da lactação apresentaram estimativas de correlações genéticas próximas a unidade. Estes resultados estão de acordo com aqueles relatados em bovinos da raça Holandesa por REKAYA et al. (1999), OLORI et al. (1999) e no Brasil por COBUCCI et al. (2005) e BIGNARDI et al. (2009a). Foram observadas também correlações genéticas negativas entre as semanas iniciais e finais da lactação (não apresentadas), para os dois modelos, entretanto para o modelo BSQ8 observou-se menor frequência. Estas correlações negativas podem ser atribuídas à dificuldade de modelagem das produções de leite do início e do final da lactação. No início, além da vaca ainda estar sofrendo o estresse do parto também apresenta um déficit de energia. Correlações genéticas negativas também foram obtidas com MRA empregando diferentes funções na raça Holandesa por JAMROZIK & SCHAEFFER (1997), OLORI et al. (1999), BROTHERSTONE et al. (2000), KETTUNEN et al. (2000) e no Brasil por COBUCCI et al. (2005), COSTA et al. (2008) e BIGNARDI et al. (2009a,b).

Tabela 3 - Estimativas de correlações genéticas (acima da diagonal) e fenotípicas (abaixo da diagonal) entre as produções de leite no dia do controle em determinados dias de lactação (DEL) obtidas com o modelo BSC8.

DEL	30	60	90	120	150	180	210	240	270	305
30		0,85	0,66	0,43	0,33	0,34	0,36	0,35	0,27	0,26
60	0,61		0,93	0,77	0,70	0,70	0,70	0,69	0,60	0,52
90	0,48	0,65		0,93	0,86	0,80	0,77	0,76	0,69	0,63
120	0,42	0,56	0,62		0,95	0,85	0,79	0,76	0,67	0,57
150	0,37	0,49	0,56	0,63		0,95	0,90	0,88	0,78	0,63
180	0,33	0,46	0,50	0,55	0,65		0,98	0,96	0,88	0,71
210	0,30	0,44	0,47	0,51	0,60	0,66		0,99	0,91	0,74
240	0,28	0,40	0,44	0,48	0,53	0,58	0,67		0,94	0,79
270	0,24	0,37	0,39	0,40	0,45	0,50	0,57	0,63		0,91
305	0,22	0,31	0,34	0,36	0,38	0,41	0,47	0,54	0,65	

Quanto às correlações de ambiente permanente (Tabela 4), as estimativas entre as produções de leite foram menores conforme aumentou o intervalo dos controles. De modo geral, as estimativas das correlações de ambiente permanente foram de menor magnitude que as correlações genéticas. As estimativas de correlações fenotípicas, genéticas e de ambiente permanente obtidas pelo BSQ8 foram muito similares ao modelo BSC8.

Comparando-se as estimativas dos parâmetros obtidos com os modelos BSC8 e BSQ8, com mesmo número de parâmetros (78), mas ajustados com diferentes ordens (cúbica e quadrática, respectivamente) e número de segmentos (5 e 6, respectivamente), foi possível observar que praticamente não apresentaram diferenças, embora os valores de QMR e VRP indiquem superioridade do modelo BSC8.

Tabelas 4 - Estimativas de correlações de ambiente permanente entre as produções de leite no dia do controle em determinados dias de lactação (DEL) do obtidas com o modelo BSC8.

DEL	30	60	90	120	150	180	210	240	270	305
30		0,85	0,71	0,69	0,63	0,53	0,46	0,42	0,38	0,32
60			0,64	0,82	0,71	0,62	0,55	0,47	0,44	0,37
90				0,93	0,81	0,70	0,63	0,56	0,48	0,42
120					0,92	0,78	0,77	0,67	0,57	0,46
150						0,94	0,84	0,67	0,56	0,48
180							0,91	0,71	0,58	0,47
210								0,68	0,57	0,47
240									0,83	0,71
270										0,84
305										

CONCLUSÃO

A estrutura de covariâncias da produção de leite no dia do controle ao longo da lactação pode ser modelada por meio de modelos de regressão aleatória utilizando-se função B-splines e inferência bayesiana.

O modelo empregando B-spline cúbica com o número de coeficientes de regressão aleatória igual a oito tanto para o efeito aleatório genético aditivo como de ambiente permanente foi o mais adequado de acordo com os critérios de comparação. Entretanto o modelo empregando B-spline quadrática poderia ser adotado, uma vez que, foram constatadas poucas mudanças nos parâmetros genéticos estimados em relação ao modelo B-spline cúbica e, além disso, as estimativas a posteriori apresentaram menores intervalos de credibilidade nos pontos extremos da curva de lactação.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO, C. V.; TORRES, R. A.; COSTA, C. N.; et. al. Uso de modelos de regressão aleatória para descrever a variação genética da produção de leite na raça Holandesa. **Revista Brasileira de Zootecnia**, vol.35, n.3, supl, p.975-981, 2006.

BIGNARDI, A.B.; EL FARO, L.; ALBUQUERQUE, L.G.; CARDOSO, V.L.; MACHADO, P.F. Random regression models to estimate test-day milk yield genetic parameters Holstein cows in southeastern Brazil. **Livestock Science**, v.123, p.1-7, 2009a.

BIGNARDI, A.B.; EL FARO, L.; ALBUQUERQUE, L.G.; CARDOSO, V.L.; MACHADO, P.F. Parametric correlation functions to model the structure of permanent environmental (co)variances in milk yield random regression models. **Journal Dairy Science**, v.92, p.4634–4640, 2009b.

BOHMANOVA, J.; MIGLIOR, F.; JAMROZIK, J.; MISZTAL, I.; SULLIVAN, P. G. Comparison of random regression models with Legendre polynomials and linear splines for production traits and somatic cell score of Canadian Holstein cows. **Journal Dairy Science**, v.91, p.3627–3638, 2008.

BOX, G.E.; TIAO, G.C. **Bayesian Inference in Statistical Analysis**. New York: J.Wiley, 360p, 1973.

CASELLA, G.; GEORGE. E. I., Explaining the Gibbs Sampler. **The American Statistician**, v.46 p.167-174, 1992.

COBUCI, J.A.; EUCLYDES, R.F.; LOPES, P.S.; COSTA, C.N.; TORRES, R.A.; PEREIRA, C. S. Estimation of genetic parameters for test-day milk 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.N.R.; PACHER, I.U. et al. Genetic parameters for test day milk yield of first lactation Holstein cows estimated by random regression using Legendre polynomials. **Revista Brasileira de Zootecnia**, v.37, p.602-608, 2008.

De BOOR, C. **A practical guide to splines**. Springer-Verlag, New York, 1978.

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.

EL FARO, L.; ALBUQUERQUE, L.G.; CARDOSO, V.L. Variance component estimates for test-day milk yield applying random regression models. **Genetics Molecular Biology**, v.33, p.665–673, 2008.

EILERS, PHC; MARX, BD **Splines, knots, and penalties**. 10 jan. 2008. Capturado em 10 jan. 2008. On-line. Disponível na Internet: http://www.stat.lsu.edu/faculty/bmarx/splines_knots_penalties.pdf, 2005

FULLER W.A., Grafted polynomials as approximating functions. **Australian Journal Agriculture Economic**, v.13, p.35–46, 1969.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. (eds. Bernardo, J.M.; Berger, J.O.; Dawid, A.P.; Smith, A.F.M.), New York: Oxford University Press, **Bayesian Statistics**, v.4, p.625-631, 1992.

GIANOLA, D.; FERNANDO, R. Bayesian methods in animal breeding theory. **Journal Animal Science**, v.63, p.217-244, 1986.

HEIDELBERGER, P; WELCH, P Simulation Run Length Control in the Presence of an Initial Transient." **Operations Research**, v.31, p.1109-1144, 1983.

JAMROZIK, G. J.; KISTEMAKER, J.C.M.; DEKKERS; SCHAEFFER, L. R. Comparison of possible covariates for use in a random regression model for analyses of test day yields. **Journal Dairy Science**, v.80, p.2550-2556, 1997.

JAMROZIK, G. J.; GIANOLA, D.; 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.

JAMROZIK, J., L. R. SCHAEFER. Bayesian comparison of random regression models for test-day yields in dairy cattle. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7, 2002, Montpellier. **Proceedings...**Montpellier, 2002b. I CD-ROM.2002.

KETTUNEN, A.,MÄNTYSAARI, E., PÖSÖ J. Estimation of genetic parameters for daily milk yield of primiparous Ayrshire cows by random regression test-day models. **Livestock Production Science**,v.66, p.251-261, 2000.

LOPEZ-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.

MEYER K., "RRGIBBS" – A program for simple random regression analyses via Gibbs sampling. CD-ROM 7 World Congress. Genetics applied to livestock production. **Proceedings...** Communication No. 28–27, 2002.

MEYER, K. Random regression analyses using B-splines to model growth of Australian Angus cattle. **Genetics Selection Evolution** vol.37, p.473-500, 2005a.

MEYER, K. Advance in methodology for random regression analyses. **Australian Journal of experimental agriculture**, v. 45, p.847-858, 2005b.

MEYER K. "WOMBAT" – digging deep for quantitative genetic analyses using restricted maximum likelihood. CD ROM Eighth World Congress Genetics Applied Livestock Production, **Proceedings....** Communication No. 27-14, 2006.

MISZTAL, I Properties of random regression models using linear splines. **Journal Animal Breeding Genetics**, v.123, p. 74-80, 2006.

OLORI, V. E.; HILL W. G.; MCGUIRK, B. J.; BROTHERSTONE S. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. **Livestock Production Science**, v.61, p.53–63, 1999.

R Development Core Team, **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL<http://www.R-project.org>, 2008. Acesso: 20 jan. 2009.

RAFTERY, A. E.; LEWIS, S. **How many iterations in the Gibbs sampler?** Bayesian Statistics 4 (eds.: J. M. Bernardo et al.), Oxford: University Press, p.763-773. 1992.

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.

RICE, J.A.; WU C.O. Nonparametric mixed effects models for unequally sampled noisy curves. **Biometrics**, v.57, p.253-259, 2001.

RUPPERT D., WAND M.P., CARROLL R.J., **Semiparametric Regression**, Cambridge University Press, New York, 2003.

SILVESTRE, A. M.; PETIM-BATISTA, F.; COLAÇO, J. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. **Journal Dairy Science**, v.89, p.1813-1821, 2006.

VAN TASSEL, C.P.; CASELLA, G.; POLLAK, E.J. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. **Journal of Dairy Science**, v.78, n.3, p.678-692, 1995.

WHITE I.M.S.; THOMPSON R.; BROTHERSTONE S., Genetic and environmental smoothing of lactation curves with cubic splines. **Journal Dairy Science** v. 82, p.632–638, 1999.