

Utilização de Modelos de Regressão Aleatória para Produção de Leite no Dia do Controle, com Diferentes Estruturas de Variâncias Residuais¹

Lenira El Faro^{2,3}, Lucia Galvão de Albuquerque^{4, 5}

RESUMO - Foram utilizados quatorze modelos de regressão aleatória, para ajustar 86.598 dados de produção de leite no dia do controle de 2.155 primeiras lactações de vacas Caracu, truncadas aos 305 dias. Os modelos incluíram os efeitos fixos de grupo contemporâneo e a covariável idade da vaca ao parto. Uma regressão ortogonal de ordem cúbica foi usada para modelar a trajetória média da população. Os efeitos genéticos aditivos e de ambiente permanente foram modelados por meio de regressões aleatórias, usando polinômios ortogonais de Legendre, de ordens cúbicas. Diferentes estruturas de variâncias residuais foram testadas e consideradas por meio de classes contendo 1, 10, 15 e 43 variâncias residuais e de funções de variâncias (FV) usando polinômios ordinários e ortogonais, cujas ordens variaram de quadrática até sêxtupla. Os modelos foram comparados usando o teste da razão de verossimilhança, o Critério de Informação de Akaike e o Critério de Informação Bayesiano de Schwarz. Os testes indicaram que, quanto maior a ordem da função de variâncias, melhor o ajuste. Dos polinômios ordinários, a função de sexta ordem foi superior. Os modelos com classes de variâncias residuais foram aparentemente superiores àqueles com funções de variância. O modelo com homogeneidade de variâncias foi inadequado. O modelo com 15 classes heterogêneas foi o que melhor ajustou às variâncias residuais, entretanto, os parâmetros genéticos estimados foram muito próximos para os modelos com 10, 15 ou 43 classes de variâncias ou com FV de sexta ordem.

Palavras-chave: funções de covariância, parâmetros genéticos, produção de leite

Random Regression Test-Day Models for Milk Yield Records, with Different Structure of Residual Variances

ABSTRACT - Fourteen random regression models were used to adjust 86,595 test-day milk records of 2,155 first lactation of native Caracu cows. The models include fixed effects of contemporary group and age of cow as covariable. A cubic regression on Legendre orthogonal polynomial of days in milk was used to model the mean trend and the additive genetic and permanent environmental regressions. Different structures of residual variances were tried and considered through homogeneous variances or heterogeneous variances, modeled as a step function with 10, 15 and 43 classes or variance functions, using ordinary and orthogonal polynomials of different orders (quadratic to sixty). Models were compared by Likelihood ratio test, Akaike's Information Criterion and Bayesian Information Criterion. These tests indicated that functions with higher order improved the change in log-likelihood. The models with step functions were superior to models with residual variance functions. Homogeneous residual variances were not adequate. The model using a step function with 15 heterogeneous variances presented the best fit. However, the genetic parameters estimated by the models with 10, 15 or 43 classes or with a sixty order variance function were similar.

Key Words: covariance functions, genetic parameters, milk yield

Introdução

Medidas repetidas de um mesmo indivíduo têm sido analisadas sob diferentes aspectos metodológicos. As abordagens usuais têm utilizado os modelos de repetibilidade e os modelos de dimensão finita ou 'test-day models', em análises uni ou multicaracterísticas. Mais recentemente, os Modelos de Regressão Aleatória (MRA) têm sido propostos como alternativa na avaliação genética de dados

longitudinais (Henderson Jr., 1982), como a produção mensal de leite. Os MRA podem ser chamados também de modelos de dimensão infinita, uma vez que existem medidas infinitas por indivíduo e tais medidas são mais correlacionadas, quanto mais próximas entre si.

Os MRA permitem ajustar curvas de lactação aleatórias para cada indivíduo, como desvios de uma curva média da população, utilizando-se polinômios ordinários ou outras funções lineares. Além disso,

¹ Financiado pelo CNPq.

² Aluno de doutorado da FCAV/UNESP.

³ Pesquisador do Instituto de Zootecnia, SP. E.mail: lenira.ddd@apta.sp.gov.br

⁴ Prof. assistente doutor - FCAV/UNESP, Rod. Paulo Donato Castalani km 8, CEP: 14870-000. Jaboticabal - SP. E.mail: lgalb@fcav.unesp.br

⁵ Pesquisador do CNPq.

possibilitam a obtenção de estruturas de (co)variâncias entre os diferentes controles através de funções de covariância para os efeitos aleatórios de interesse. O uso de "test-day models" aplicando-se regressões aleatórias, além de permitir diferenciar uma curva para cada animal pode, no futuro, permitir a seleção de animais cujas curvas de lactação sejam mais persistentes (Ptak & Schaeffer, 1993; Schaeffer, 1996; Jamrozik & Schaeffer, 1997).

Um conceito que vem sendo abordado quando se trabalha com os MRA diz respeito às funções de covariância (FC). As FC foram propostas por Kirkpatrick et al. (1990) para a análise de dados longitudinais, por exemplo, medidas de crescimento ou de lactação e, segundo estes autores, são equivalentes às matrizes de covariância dos modelos multivariados de dimensão finita. As funções de covariância podem ser obtidas a partir das matrizes de variância e covariância dos modelos de dimensão finita, ou por meio das matrizes de (co)variância entre os coeficientes de regressão, estimadas pelos MRA (Van der Werf & Schaeffer, 1997; Meyer, 1998).

O grande interesse em relação às FC para a modelagem de dados longitudinais é que elas permitem descrever mudanças graduais das covariâncias com o passar do tempo, e prever variâncias e covariâncias para pontos ao longo de uma trajetória, mesmo que haja pouca ou nenhuma informação entre os pontos.

Os MRA/FC permitem a modelagem dos efeitos aleatórios, utilizando funções polinomiais lineares e, além disso, permitem considerar heterogeneidade de variâncias residuais em cada controle.

Estudos iniciais usando os MRA, como os de Jamrozik & Schaeffer (1997) e Jamrozik et al. (1997), desconsideravam estrutura heterogênea de variâncias para os resíduos, o que, aliado a problemas de modelagem do efeito de ambiente permanente, proporcionou superestimções das variâncias genéticas aditivas. Esses autores estimaram herdabilidades de até 0,59 para as produções de leite no dia do controle.

A modelagem considerando as variâncias residuais heterogêneas para cada classe pode melhorar a partição da variação total, mas proporciona um aumento no número de parâmetros a serem estimados no processo de maximização da função de verossimilhança. O número elevado de parâmetros dos modelos, a estrutura de covariâncias e o fato de as matrizes de incidência serem mais densas que para os modelos convencionais podem dificultar a convergência na

estimação dos componentes de variância, sendo entretanto para a utilização dessa metodologia para avaliações genéticas em conjuntos com grande volume de dados. Têm-se testado algumas alternativas para se obter MRA mais parcimoniosos, sem que haja perda na qualidade de ajuste dos efeitos aleatórios. Estudos têm proposto a utilização de estruturas de variâncias residuais distintas, com número reduzido de classes de variâncias, por intermédio do agrupamento de classes que contenham variações semelhantes ou por meio de funções de variância (FV) residuais, utilizando polinômios ortogonais de Legendre ou ordinários (Olori et al., 1999a; Meyer, 1999; Brotherstone et al., 2000; Albuquerque & Meyer, 2001). A utilização de funções de variância tem se mostrado uma boa alternativa, uma vez que, por meio de polinômios de baixa ordem, pode-se descrever a variação residual em todos os períodos da lactação.

O objetivo do presente estudo foi comparar diferentes estruturas de variâncias residuais para modelos de regressão aleatória, buscando minimizar o número de parâmetros a serem estimados e acelerar o processo de obtenção dos componentes de (co)variância, o que poderá viabilizar a adoção dessa metodologia em futuras avaliações genéticas em conjuntos de dados com grande volume de informação.

Material e Métodos

No presente estudo, utilizaram-se 2.155 primeiras lactações de vacas Caracu, iniciadas entre os anos de 1978 a 1998, pertencentes ao rebanho da fazenda Chiqueirão, Poços de Caldas, MG. Os controles foram registrados semanalmente, provenientes de ordenha manual, realizada duas vezes ao dia, com a presença do bezerro, sendo que o leite de um quarto de úbere é reservado à cria durante toda a lactação. Os animais foram alimentados basicamente a pasto, nativo ou cultivado, com suplementação no período das secas. A fazenda, quanto aos locais de ordenha, é dividida em 15 retiros e, durante a lactação, o animal passa por até três retiros, dependendo da fase da lactação e do nível de produção. As lactações foram truncadas aos 305 dias e apresentaram entre 6 e 43 controles leiteiros, sendo que acima de 94% das lactações contiveram 35 ou mais controles, totalizando 86.598 controles.

A produção de leite no dia do controle foi considerada em classes semanais de dia em lactação, variando de 1 até 43 classes, sendo que a classe 1

compreendeu os dias 5 a 11 de lactação, a classe 2, os dias 12 a 18 e assim, sucessivamente. As análises foram realizadas por meio de um modelo animal, unicaracterístico, de Regressão Aleatória. Foram estimadas funções de covariância para os componentes genético aditivos e de ambiente permanente, modeladas por meio de polinômios ortogonais numa escala de Legendre. O modelo usado conteve como efeitos fixos o grupo contemporâneo, os efeitos linear e quadrático da covariável idade da vaca ao parto. Além destes efeitos fixos, em modelos de regressão aleatória, considera-se também a regressão sobre o tempo (semana de lactação), o que gerou a curva média de lactação, ou a trajetória média da população (fixa). Assim, as curvas aleatórias para cada indivíduo foram estimadas como desvios da trajetória média. Como aleatórios, foram considerados os efeitos genético aditivo direto e de ambiente permanente, além do residual. O grupo contemporâneo foi definido pelas variáveis retiro, ano e mês do controle e apresentou 1.582 níveis. As subclasses de grupo contemporâneo foram formadas, inicialmente, pelas variáveis retiro e data do controle leiteiro, o que gerou um grande número de subclasses contendo poucas informações e problemas de convergência. A regressão fixa para a curva média de lactação, assim como as regressões aleatórias genético-aditivas e de ambiente permanente do animal foram modeladas por meio de polinômios ortogonais de Legendre, de ordem cúbica.

A ordem da função de regressão para descrever a trajetória média da população foi definida após testar polinômios ortogonais de Legendre cujas ordens variaram entre linear até sêxtupla, num modelo que considerou apenas os efeitos fixos já citados anteriormente. Os parâmetros pertencentes a esse modelo fixo foram estimados pelo método de quadrados mínimos ordinários. Com base no quadrado médio residual estimado para cada modelo, a trajetória média da população foi modelada por meio de um polinômio ortogonal de ordem cúbica, uma vez que os quadrados médios de resíduos obtidos entre as ordens cúbica até sêxtupla pouco diferiram.

No presente estudo foram testados quatorze modelos, sendo que eles diferiram quanto a maneira de ajustar as estruturas de variâncias residuais, considerando-as como classes contendo 1, 10, 15 ou 43 variâncias ou através de funções de variâncias (FV) de ordens quadrática até sêxtupla. As variâncias residuais, quando consideradas como classes, foram

agrupadas de acordo com o comportamento das mesmas no decorrer da lactação, tendo como base o modelo que conteve as 43 diferentes classes. Para o modelo Leg-10, as 10 classes foram agrupadas da seguinte maneira: semana 1, 2, 3, 4-7, 8-10, 11-13, 14-25, 26-35, 36-42 e 43. Para o modelo Leg-15, as classes foram agrupadas por: semana 1, 2, 3, 4-5, 6-7, 8-9, 10-11, 12-13, 14-16, 17-25, 26-30, 31-35, 36-38, 39-42, 43. As FV residuais foram modeladas por meio de polinômios ordinários ou polinômios ortogonais numa escala de Legendre.

Os modelos são referenciados como, por exemplo, Leg-1 (homogeneidade de variâncias residuais); Leg-10 (heterogeneidade de variâncias residuais agrupadas em 10 classes) Leg-2F (função de variância dos resíduos usando um polinômio ordinário quadrático); Leg-2L (função de variância dos resíduos usando um polinômio ortogonal de Legendre quadrático). O número de parâmetros estimados para cada modelo foi de $k(k+1)/2$ coeficientes para os efeitos genéticos aditivos e de ambiente permanente, sendo k o número de coeficientes da função de regressão usada, além das variâncias residuais. O modelo de regressão aleatória pode ser representado por:

$$y_{ij} = F + \sum_{m=0}^{k_b-1} \mathbf{b}_m \mathbf{f}_m(t_i) + \sum_{m=0}^{k_a-1} \mathbf{a}_{jm} \mathbf{f}_m(t_{ij}) + \sum_{m=0}^{k_{ap}-1} \mathbf{g}_{jm} \mathbf{f}_m(t_{ij}) + e_{ij}$$

em que: y_{ij} = produção de leite do i -ésimo controle, pertencente ao j -ésimo animal; Φ = conjunto de efeitos fixos, exceto as covariáveis semana de lactação sobre a produção; β_m = conjunto de m regressores fixos para a curva média da população; $\Phi_m(t_i)$ = função de regressão de ordem k_b , que descreve a curva média da população de acordo com a semana de lactação (t_i); $\Phi_m(t_{ij})$ = funções de regressão, que descrevem as trajetórias de cada indivíduo j , de acordo com a semana de lactação (t_i), para os efeitos aleatórios genético aditivos e de ambiente permanente; α_{jm} , γ_{jm} = são os m regressores aleatórios genético-aditivos e de ambiente permanente, respectivamente, para cada animal j ; k_b , k_a e k_{ap} = ordens dos polinômios ou das funções lineares utilizadas para os efeitos descritos acima e, $k_b = k_a = k_{ap} = 4$, indicando que uma função cúbica foi usada para descrever as trajetórias; e_{ij} = erro aleatório associado a cada controle i do animal j .

Na forma matricial o modelo pode ser representado por:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wap} + \mathbf{e}$$

em que: \mathbf{y} = vetor das N observações, medidas em Nd

animais; \mathbf{b} = vetor de efeitos fixos que inclui as soluções para GC e para as covariáveis, idade ao parto e semana de lactação; \mathbf{a} = vetor de soluções para os coeficientes de regressão aleatórios genético-aditivos; \mathbf{ap} = vetor de soluções para os coeficientes de regressão aleatórios de ambiente permanente; \mathbf{e} = vetor dos N diferentes resíduos; \mathbf{X} , \mathbf{Z} , \mathbf{W} = matrizes de incidência para os efeitos fixos, aleatórios genético-diretos e de ambiente permanente, respectivamente.

A dimensão do vetor \mathbf{a} é de $k \times N_a$ coeficientes, sendo que k_a representa o número de parâmetros da função de regressão, N_a , o número de animais na matriz de parentesco. O vetor \mathbf{ap} tem dimensão de $k_{ap} \times N_d$ coeficientes, sendo que k_{ap} representa o número de parâmetros da função de regressão para ambiente permanente e N_d , o número de animais com registros.

As pressuposições em relação aos componentes são:

$$E \begin{bmatrix} \mathbf{y} \\ \mathbf{a} \\ \mathbf{ap} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{Xb} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; \quad e \quad \begin{aligned} V(\mathbf{a}) &= \mathbf{K}_A \otimes \mathbf{A}; \\ V(\mathbf{ap}) &= \mathbf{K}_{AP} \otimes \mathbf{I}_{N_d}; \\ V(\mathbf{e}) &= \mathbf{R}, \end{aligned}$$

em que: \mathbf{K}_A e \mathbf{K}_{AP} são as matrizes de (co)variâncias entre os coeficientes de regressão aleatórios genético-aditivos e de ambiente permanente, respectivamente; \mathbf{A} é a matriz de parentesco entre os indivíduos; \mathbf{I}_{N_d} é uma matriz identidade de dimensão N_d ; \otimes é o produto de Kroenecker entre matrizes; \mathbf{R} representa uma matriz diagonal, contendo as variâncias residuais ($\sigma^2_{e_i}$). Assumiu-se independência entre as variâncias residuais.

Os componentes de variância foram estimados pelo método da Máxima Verossimilhança Restrita (REML), usando a opção DXMRR do pacote estatístico DFREML (Meyer, 1998).

Quando foram utilizadas funções de variância, as variâncias residuais foram estimadas por meio dos coeficientes de regressão das funções de variância (FV), representadas por:

$$V_{e_i}^2 = V_{e_0}^2 \left(1 + \sum_{r=1}^q \mathbf{b}_r t_{ij}^r \right)$$

em que: $V_{e_0}^2$ corresponde à variância do intercepto, \mathbf{b}_r são os q coeficientes de regressão da FV de ordem q e t_{ij} são as semanas de lactação.

A comparação entre os modelos aninhados foi feita por intermédio do teste da razão de verossimi-

lhança (LRT), que compara as diferenças entre as funções de máxima verossimilhança linearizadas de cada modelo, com o valor obtido de uma distribuição qui-quadrado com o número de graus de liberdade igual à diferença no número de parâmetros entre os modelos 1 e 2, dado por:

$$\{ [k_{a1}(k_{a1}+1)/2] + [k_{ap1}(k_{ap1}+1)/2] + e(i)1 \} - \{ [k_{a2}(k_{a2}+1)/2] + [k_{ap2}(k_{ap2}+1)/2] + e(i)2 \},$$

em que: $e(i)$ representa o número de variâncias residuais consideradas em cada modelo.

Além do LRT foram usados como referências, o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano de Schwarz (BIC). Ambos os testes permitem comparação entre modelos não aninhados e penalizam modelos com maior número de parâmetros, sendo que para BIC essa penalidade é mais rigorosa, tendendo a favorecer modelos mais parcimoniosos (Wolfinger, 1993; Nunez-Antón & Zimmerman, 2000). Para AIC o valor para comparação é dado por:

$$AIC = -2 \log L + 2p$$

e para BIC, a comparação é feita através de:

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

em que: p refere-se ao número de parâmetros do modelo, N é o número total de observações e r é o posto da matriz \mathbf{X} , que é a matriz de incidência para os efeitos fixos. Menores valores para AIC e BIC indicam melhor ajuste. Além dos testes citados, os parâmetros estimados para cada modelo foram comparados para averiguar se ocorreram mudanças em função da estrutura de variâncias residuais usada.

Resultados e Discussão

A curva média estimada por quadrados mínimos ordinários para um modelo contendo apenas os efeitos fixos, apresentou-se praticamente como uma reta (Figura 1). O polinômio ortogonal de ordem cúbica para a regressão fixa foi usado no presente estudo porque o quadrado médio de resíduos foi menor para essa função, quando comparado com os da linear e quadrática. Os parâmetros estimados a partir da regressão polinomial de Legendre para a curva de lactação média desse rebanho foram de 3,934 kg para o intercepto; -1,802 kg para o coeficiente de regressão linear; -0,043 kg para o coeficiente de regressão quadrático e 0,083 kg para o cúbico.

O resumo das análises quanto ao log da função de verossimilhança está apresentado na Tabela 1. Ape-

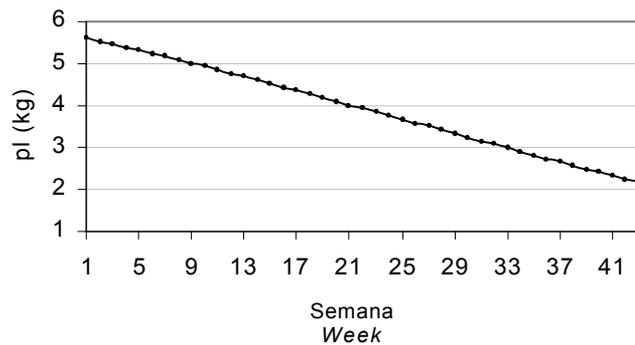


Figura 1 - Curva de lactação média estimada para a população, pelo método de quadrados mínimos ordinários.

Figure 1 - Average lactation curve estimated to the population by ordinary least square method.

nas modelos aninhados foram comparados através do LRT. Observaram-se mudanças significativas ($P < 0,01$) na função, quando houve aumento no número de parâmetros dos modelos, exceto quando se compararam Leg-15 e Leg-43.

Os modelos que usaram funções de variância residuais por meio de polinômios ordinários apresentaram convergências mais rápidas. Quando a comparação foi feita apenas entre os polinômios ordinários de diferentes ordens (Leg-2F, Leg-3F, Leg-4F, Leg-5F, Leg-6F), ocorreram mudanças significativas ($P < 0,01$)

na função de verossimilhança quanto maior o grau do polinômio, de acordo com o LRT. Dentre os cinco modelos, o polinômio ordinário de sexta ordem foi o mais adequado.

Os polinômios ortogonais de Legendre apresentaram dificuldades em relação à convergência do processo iterativo, sendo que a partir do modelo de ordem quártica (Leg-4L) não houve convergência mesmo quando tentou-se reiniciar as análises com diversos valores iniciais para os parâmetros ou com diferentes métodos de busca do máximo da função, usando Simplex, Powell ou AI-REML. Entre os modelos com polinômios ortogonais, o cúbico apresentou ajuste superior ao quadrático, de acordo com o LRT, AIC e BIC.

Dos modelos contendo as classes de variâncias residuais (Leg-1, Leg-10, Leg-15 e Leg-43), claramente o modelo considerando homogeneidade de variâncias residuais (Leg-1) mostrou-se inadequado. Os modelos contendo as classes de variâncias heterogêneas tenderam a ser superiores àqueles descritos através de funções de variâncias residuais, de acordo com os testes AIC e BIC. De acordo com o LRT, os modelos Leg-15 e Leg-43 não apresentaram diferenças quanto à função de máxima verossimilhança, dentre os modelos descrevendo as classes de variâncias heterogêneas. Entre todos os modelos estudados, o modelo Leg-15 foi o melhor, de acordo

Tabela 1 - Número de parâmetros estimados (p), função de máxima verossimilhança ($2 \text{ Log } L$), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC), teste da razão de verossimilhança (LRT) para os diferentes modelos

Table 1 - Number of estimated parameters (p), maximum likelihood function ($2 \text{ Log } L$), Akaike's information criterion (AIC), Bayesian information criterion (BIC), and likelihood ratio test (LRT) for different models

Modelo ¹ <i>Model</i>	e^2	p	$2 \text{ Log } L$	AIC	BIC	LRT ³
1) Leg-2L	2	23	-102115,49	102161,5	102377,0	(2-1) 2055,1 **
2) Leg-3L	3	24	-100060,35	100108,3	100333,2	-
3) Leg-2F	2	23	-99089,63	99135,6	99351,1	(4-3) 441,8 **
4) Leg-3F	3	24	-98647,87	98695,9	98920,7	(5-4) 21,5 **
5) Leg-4F	4	25	-98626,37	98676,4	98910,6	(6-5) 7,7 **
6) Leg-5F	5	26	-98618,70	98670,7	98914,3	(7-6) 67,4 **
7) Leg-6F	6	27	-98551,34	98605,3	98858,3	-
8) Leg-1	1	21	-106377,86	106419,9	106616,6	(9-8) 7784,6 **
9) Leg-10	10	30	-98593,23	98653,2	98934,3	(10-9) 177,7 **
10) Leg-15	15	35	-98415,48	98485,4	98813,4	(11-10) 36,1 ^{ns}
11) Leg-43	43	63	-98379,39	98505,4	99095,6	-

¹ Leg-2L função de variâncias com Legendre quadrático para as variâncias residuais; Leg-2F: função de variâncias com polinômio ordinário quadrático; Leg-10: classes com 10 variâncias heterogêneas; 2: número de variâncias residuais ou número de coeficientes da função de variâncias residuais; 3: teste da razão de verossimilhança entre os modelos aninhados; **: $P < 0,01$; ns: não significativo. Em itálico estão destacados os melhores modelos

¹ leg-2l: variance function with a quadratic legendre for residual variances; leg-2f: variance function with a quadratic polynomial; leg-10: classes with 10 heterogeneous variances; 2: number of residual variances or number of coefficients of variance function; 3: likelihood ratio test among hierarchic models; **: $p < 0,01$; ns: not significant. the best models are in italic.

com AIC e BIC. O segundo melhor modelo, de acordo com AIC seria o Leg-43, mas de acordo com BIC, foi o Leg-6F. Funções de variância usando polinômios ortogonais não seriam indicadas por problemas de convergência e por apresentarem ajustes inferiores às FV com polinômios ordinários e às classes.

As variâncias fenotípicas, genéticas, de ambiente permanente e residuais para os modelos Leg-6F, Leg-10, Leg-15 e Leg-43 são representadas na Figura 2. Pode-se observar a partir dos gráficos que, no decorrer da lactação, todas as variâncias apresentaram tendências parecidas para os quatro modelos destacados. As variâncias residuais e de ambiente permanente apresentaram valores muito próximos, sendo que a residual tendeu a ser menor a partir da 14ª semana. Para o modelo que usou função de variâncias residuais, o Leg-6F, as variâncias fenotípicas, residuais e de ambiente permanente apresentaram valores menores, principalmente no início da lactação. As variâncias genéticas, entretanto,

foram muito semelhantes em magnitude. Os três modelos contendo classes de variâncias residuais (Leg-10, Leg-15 e Leg-43) apresentaram valores muito próximos de suas variâncias.

Os maiores valores para as variâncias fenotípicas e de ambiente permanente ocorreram no início da lactação, estabilizando-se no meio e aumentando levemente no final da lactação. As variâncias residuais foram também maiores no início e decresceram até o final da lactação. As variâncias genéticas aditivas foram sempre menores que as demais e foram praticamente constantes por todo o período de lactação. Tais tendências têm sido observadas na literatura por outros autores, dentre eles, Jamrozik et al. (1997) e Olori et al. (1999a,b). Nestes estudos, as variâncias genéticas aditivas foram menores que as de ambiente permanente por toda a lactação, exceto em Olori et al. (1999b), que observaram mudanças nesses valores a partir da 28ª semana, tornando-se maiores, até o final da lactação.

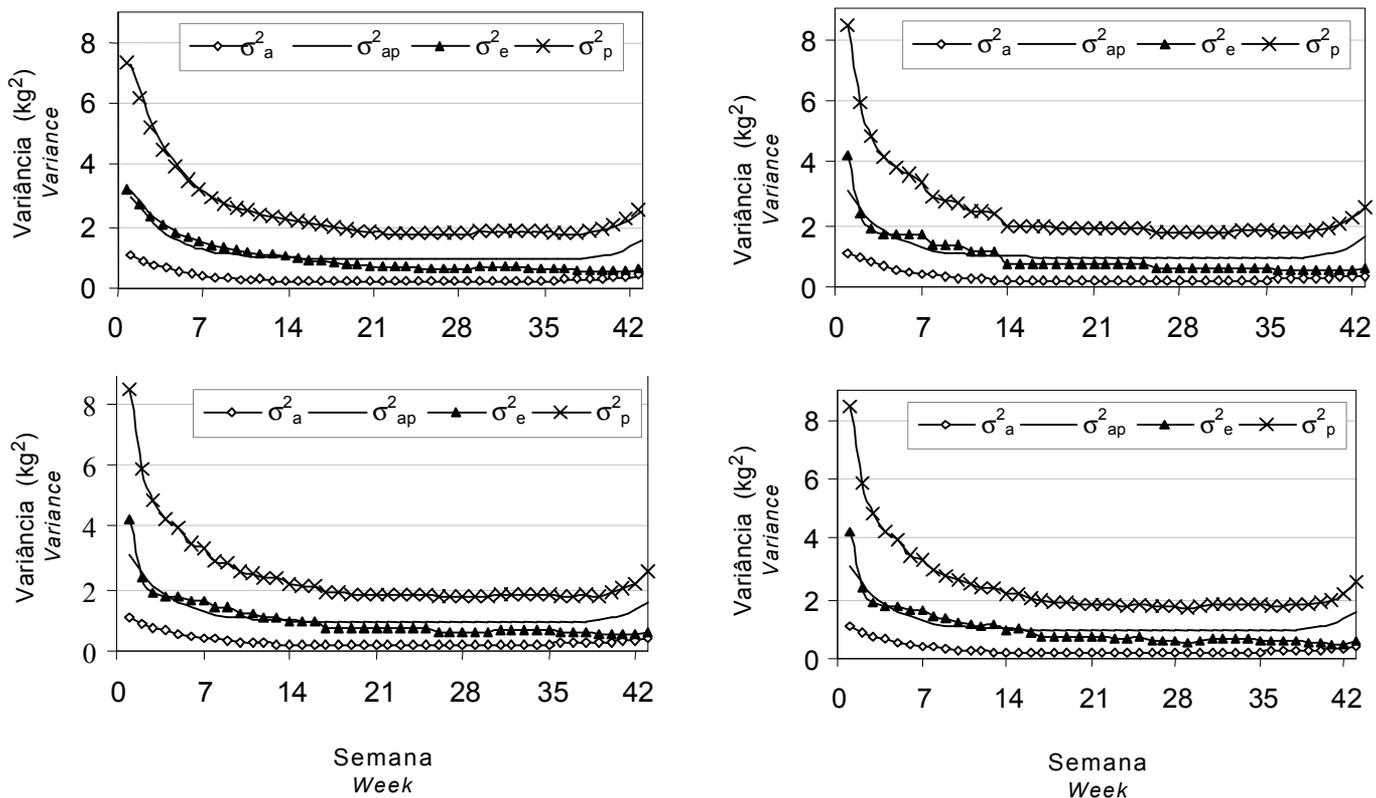


Figura 2 - Variâncias fenotípicas (σ^2_p), genéticas (σ^2_a), de ambiente permanente (σ^2_{ap}) e residuais (σ^2_e) para as produções de leite no dia do controle, de acordo com os modelos: Leg-6F (acima à esquerda), Leg-10 (acima à direita), Leg-15 (abaixo à esquerda) e Leg-43 (abaixo à direita).

Figure 2 - Phenotypic (σ^2_p), genetic (σ^2_a), permanent environmental (σ^2_{ap}) and residual (σ^2_e) variance for test day milk yield by Leg-6F (up, left), Leg-10 (up, right), Leg-15 (down, left) and Leg-43 (down, right).

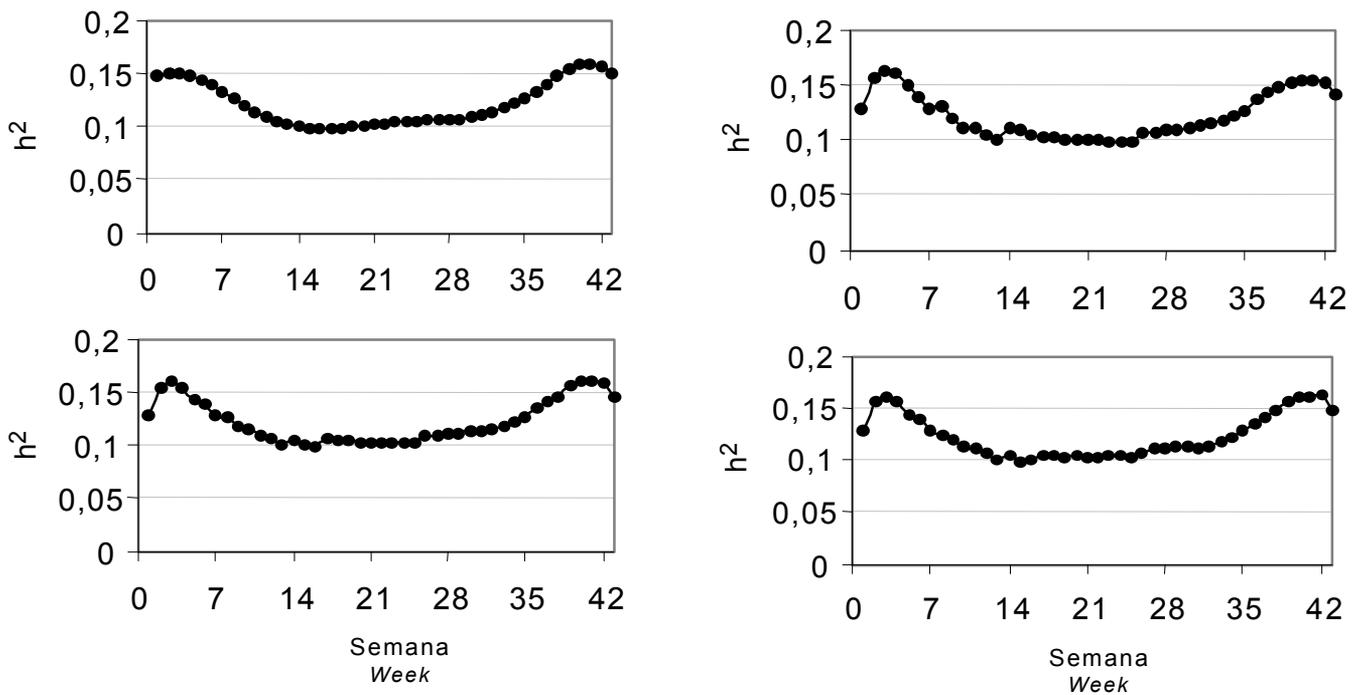


Figura 3 - Herdabilidades estimadas para as produções de leite no dia do controle, de acordo com os modelos: Leg-6F (acima à esquerda), Leg-10 (acima à direita), Leg-15 (abaixo à esquerda), Leg-43 (abaixo à direita)
 Figure 3 - Heritability estimated for test day milk yield by Leg-6F (up, left), Leg-10 (up, right), Leg-15 (down, left), Leg-43 (down, right)

Na Figura 3 são apresentadas as estimativas de h^2 para os quatro modelos. As estimativas de h^2 obtidas no presente estudo foram menores que as de estudos com a raça Holandesa (Jamrozik et al., 1997; Olori, 1997), com valores que variaram de 0,10 até 0,16. No meio da lactação os valores foram próximos a 0,10, menores que no início e no final. Para os quatro modelos os valores foram próximos, mas para os modelos contendo as classes de 10, 15 e 43 variâncias residuais ocorreram dois picos mais pronunciados, com valores de h^2 maiores entre a 2^a até a 5^a semana, valores estes próximos aos estimados para o final da lactação.

As covariâncias fenotípicas, genéticas e de ambiente permanente foram parecidas para os quatro melhores modelos, indicando que nesse caso não houve mudanças significativas nos parâmetros estudados, quando diferentes abordagens para a modelagem das variâncias residuais foram usadas.

As correlações genéticas (r_a) estimadas entre as produções nos controles, apresentadas apenas para o modelo Leg-15 (Figura 4), foram todas positivas e variaram entre 0,15 e 0,99. Como esperado, elas foram maiores quanto mais próximos os controles

leiteiros, diminuindo quando a distância entre eles aumentou. Entre os controles do meio da lactação estas r_a foram mais elevadas. Para os quatro modelos destacados (Leg-6F, Leg-10, Leg-15 e Leg-43) não houve muita diferença nas superfícies apresentadas, o que também foi constatado para as correlações fenotípicas (r_p) e de ambiente permanente (r_{ap}). As r_p e r_{ap} foram menores que as genéticas, e, ambas as superfícies apresentaram tendências semelhantes, embora para as r_p , a superfície tenha se mostrado mais achatada. As correlações de ambiente permanente e as genéticas foram elevadas para os controles mais próximos, caindo bastante quando a distância entre eles aumentou.

Assim como para os demais parâmetros usados como referência para medir as diferenças nos ajustes proporcionadas pelos modelos, a matriz de coeficientes de regressão aleatórios genético aditivos (K_A) mostrou estimativas de covariâncias e correlações entre os regressores muito próximas para todos os modelos estudados (Tabela 2). Pequenas diferenças foram observadas para o modelo Leg-10, para a variância do parâmetro a_3 e para as correlações deste com os demais parâmetros.

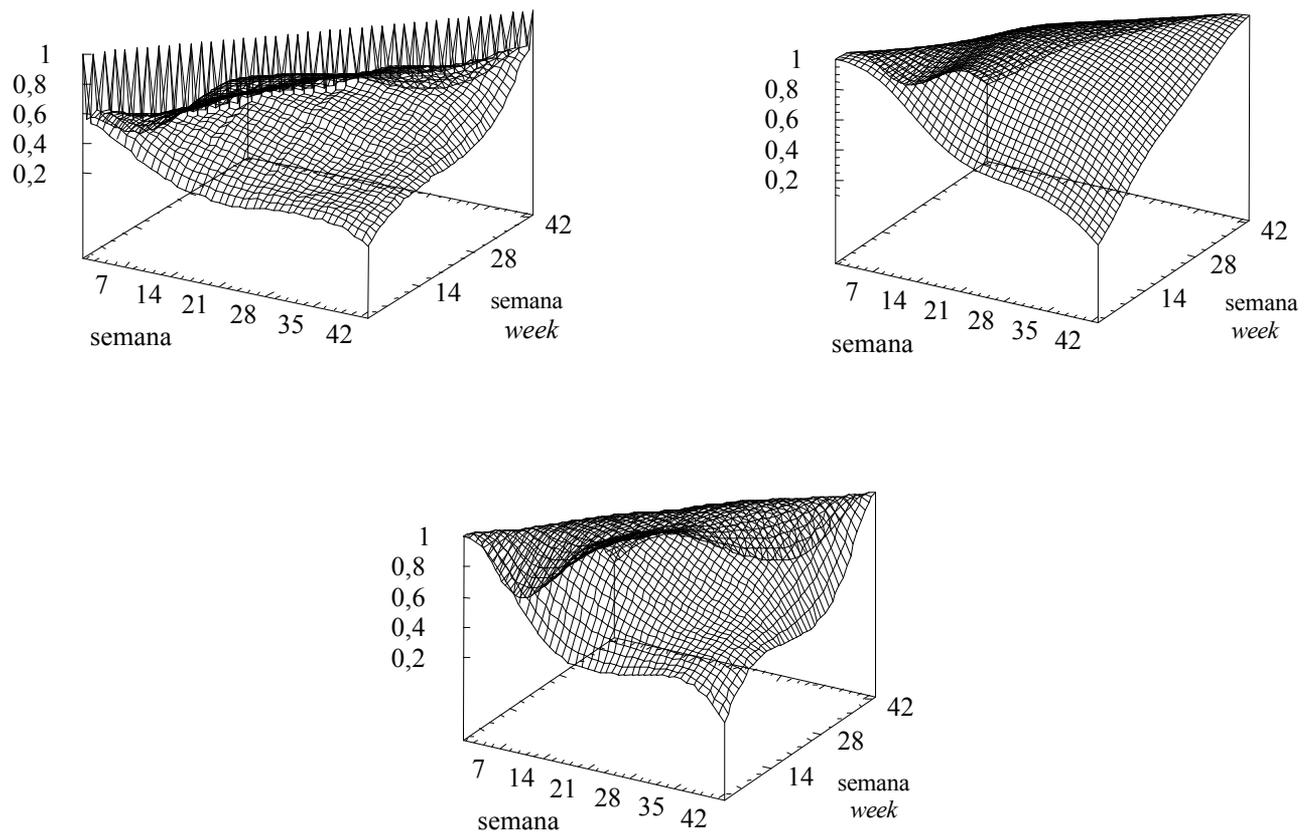


Figura 4 - Correlações fenotípicas (acima à esquerda), genéticas (acima à direita) e de ambiente permanente (abaixo) estimadas entre as produções de leite semanais, de acordo com o modelo Leg-15.

Figure 4 - Phenotypic (up, left), genetic (up, right) and permanent environmental (down), correlation estimated among test day milk yield by Leg-15.

Os resultados sugerem que não há necessidade de considerar todas as 43 variâncias residuais em diferentes classes, já que com 15 classes foi possível modelá-las. Analisando conjuntos de dados maiores, com maior número de efeitos fixos, com maior quantidade de níveis e, maior número de efeitos aleatórios no modelo, sua utilização poderia ser problemática, devido a dificuldades de convergência, tempo e demanda computacionais. Modelos alternativos como os modelos Leg-10, Leg-15 e Leg-6F proporcionariam ajustes mais parcimoniosos, principalmente se as ordens das funções de covariância para os efeitos aditivos diretos e de ambiente permanente fossem maiores. Além disso, tais modelos poderiam ser interessantes do ponto de vista de requerimentos computacionais, uma vez que para Leg-15 há uma redução de 28 parâmetros e de 36 parâmetros para

Leg-6F, em comparação com o modelo Leg-43.

Resultados relatados por Olori et al. (1999a, b) constatarem que para um modelo de ordem cúbica para os efeitos genéticos aditivos e de ambiente permanente, ocorreram poucas mudanças nas estimativas dos componentes de (co)variância quando a estrutura de variância residual foi modelada por meio de classes distintas, contendo 1, 4, 10 e 37 classes.

De maneira geral, o que se observou foi que as estimativas dos componentes de variância e dos demais parâmetros sofreram pouca interferência da maneira como foi modelada a estrutura residual, quando os quatro melhores modelos (Leg-10, Leg-15, Leg-43 e Leg-6F) foram considerados. Ocorreram diferenças nas estimativas de h^2 resultantes dos quatorze modelos estudados. Para o modelo Leg-1, por exemplo, as estimativas de h^2 foram maiores que

Tabela 2 - Estimativas de (co)variâncias (diagonal e acima) e correlações (abaixo da diagonal em itálico) genéticas entre os regressores aleatórios de acordo com quatro diferentes modelos de regressão aleatória

Table 2 - Genetic (co)variances (diagonal and up) and correlation (down of diagonal, italic) among random regression coefficients by different random regression models

Parâmetros Parameters	Leg-6F			
	a ₀	a ₁	a ₂	a ₃
a ₀	0,43476	-0,05720	0,04811	-0,01679
a ₁	-0,246	0,12411	-0,03508	0,02224
a ₂	0,389	-0,531	0,03510	-0,01173
a ₃	-0,275	0,681	-0,676	0,00858
Parâmetros Parameters	Leg-10			
	a ₀	a ₁	a ₂	a ₃
a ₀	0,44046	-0,05947	0,04875	-0,01689
a ₁	-0,251	0,12726	-0,03663	0,02185
a ₂	0,385	-0,538	0,03645	-0,01177
a ₃	-0,363	0,874	-0,880	0,004914
Parâmetros Parameters	Leg-15			
	a ₀	a ₁	a ₂	a ₃
a ₀	0,44006	-0,05725	0,04789	-0,01619
a ₁	-0,244	0,12545	-0,03549	0,02148
a ₂	0,382	-0,530	0,03574	-0,01138
a ₃	-0,276	0,687	-0,682	0,00780
Parâmetros Parameters	Leg-15			
	a ₀	a ₁	a ₂	a ₃
a ₀	0,43960	-0,05679	0,04786	-0,01642
a ₁	-0,242	0,12567	-0,03548	0,02190
a ₂	0,381	-0,528	0,03593	-0,01156
a ₃	-0,278	0,693	-0,684	0,00795

para os demais modelos, sugerindo que a partição dos componentes de variância foi afetada pelo ajuste deficiente da variação residual, subestimando as variâncias residuais no início da lactação e superestimando-as nos demais períodos. O mesmo pode ser constatado para os modelos que usaram funções de variância com polinômios ortogonais de Legendre. Os testes estatísticos, como LRT, BIC e AIC detectaram diferenças entre os modelos, indicando a dificuldade de se estabelecer algum desses critérios para a escolha de um modelo. Para os modelos contendo classes de variâncias residuais, os parâmetros estimados foram muito semelhantes, o que era esperado, uma vez que os agrupamentos das classes foram determinados após observação do comportamento das variâncias residuais para o modelo Leg-43.

R. Bras. Zootec., v.32, n.5, p.1104-1113, 2003

Conclusões

A utilização de heterogeneidade de variâncias residuais foi mais adequada para modelar os dados longitudinais de produção de leite desse conjunto de dados.

O modelo considerando 43 classes de variâncias heterogêneas aumenta consideravelmente o número de parâmetros do MRA e não apresenta resultados melhores que os obtidos no modelo com 15 classes, não havendo necessidade de adotá-lo em futuras análises.

O agrupamento das variâncias residuais em classes ou a utilização de funções de variâncias residuais podem viabilizar a adoção dos MRA para conjuntos com grande volume de dados, sem aumentar excessivamente o número de parâmetros do modelo.

De acordo com os testes estatísticos, a melhor descrição das variâncias residuais no decorrer da lactação foi a que considerou quinze classes heterogêneas. Entretanto, como poucas mudanças foram constatadas nos parâmetros genéticos estimados, as funções de variância com polinômios ordinários de sexta ordem poderiam ser usadas sem perda na precisão das estimativas, proporcionando boa redução no número de parâmetros estimados.

Literatura Citada

- ALBUQUERQUE, L.G.; MEYER, K. Estimates of covariance functions for growth from birth to 630 days of age in Nelore cattle. **Journal of Animal Science**, v.79, p.2776-2789, 2001.
- BROTHERSTONE, S.; WHITE, I.M.S.; MEYER, K. Genetic modelling of daily yield using orthogonal polynomials and parametric curves. **Journal of Animal Science**, v.70, p.407-415, 2000.
- HENDERSON JR., R. Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. **Biometrics**, v.38, p.623, 1982.
- JAMROZIK, J.; SCHAEFFER, L.R. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. **Journal of Dairy Science**, v.80, p.762-770, 1997.
- JAMROZIK, J.; KISTEMAKER, G.J.; DEKKERS, J.C.M et al. Comparison of possible covariates for use in a random regression model for analyses of test day yields. **Journal of Dairy Science**, v.80, p.2550-2556, 1997.
- KIRKPATRICK, M.; LOFSVOLD, D.; BULMER, M. Analysis of the inheritance, selection and evolution of growth trajectories. **Genetics**, v.124, p.979-993, 1990.
- MEYER, K. "DXMRR" – A program to estimate covariance functions for longitudinal data by REML. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 6., 1998, Armidale. **Proceedings...** Armidale: University of New England, 1998. CD ROM.

- MEYER, K. Estimates of genetic and phenotypic covariance functions for postweaning growth and mature weight of beef cows. **Journal of Animal Breeding and Genetics**, v.116, p.181-205, 1999.
- NUNEZ-ANTÓN, V.N.; ZIMMERMAN, D.L. Modelling nonstationary longitudinal data. **Biometrics**, v.56, p.699-705, 2000.
- OLORI, V.E. **Utilisation of daily milk records in genetic evaluation of dairy cattle**. Edinburgh, 1997. Thesis (PHD) – University of Edinburg, Edinburgh, 1997.
- OLORI, V.E.; HILL, W.G.; BROTHERSTONE, S. The structure of the residual error variance components for test day milk records by restricted maximum likelihood with a random regression models. In: COMPUTATIONAL CATTLE BREEDING WORKSHOP, 1999, Finland. **Proceedings...** Tuusula, Finland, 1999a.
- OLORI, V.E.; HILL, W.G.; MCGUIRK, B.J. et al. 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, 1999b.
- PTAK, E.; SCHAEFFER, L.R. Use of test day yields for genetic evaluation of dairy sires and cows. **Livestock Production Science**, v.34, p.23-34, 1993.
- SCHAEFFER, L.R. **Tópicos avançados em melhoramento animal: random regression models**. Jaboticabal: Universidade Estadual Paulista, 1996. p.25-33. (notas do curso)
- Van der WERF, J.H.J.; SCHAEFFER, L.R. **Random regression in animal breeding**. Guelph: University of Guelph, 1997. 58p. (course notes).
- WOLFINGEER, R. Covariance structure selection in general mixed models. **Community of Statistics – Simulation**, v.22, n.4, p.1079-1106, 1993.

Recebido em: 31/05/02

Aceito em: 10/03/03