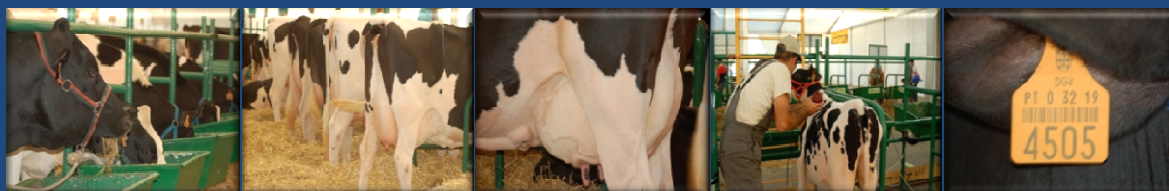


MODELOS AUTO-REGRESSIVOS PARA CONTRASTES NA AVALIAÇÃO GENÉTICA DE BOVINOS LEITEIROS: CONSEQUÊNCIAS DA SUA APLICAÇÃO



JOANA MEXIA DE ALMEIDA VASCONCELOS

Dissertação de doutoramento em Ciências Veterinárias

2007

JOANA MEXIA DE ALMEIDA VASCONCELOS

**MODELOS AUTO-REGRESSIVOS PARA CONTRASTES NA
AVALIAÇÃO GENÉTICA DE BOVINOS LEITEIROS:
CONSEQUÊNCIAS DA SUA APLICAÇÃO**

**Dissertação de Candidatura ao grau de
Doutor em Ciências Veterinárias
submetida ao Instituto de Ciências
Biomédicas de Abel Salazar da
Universidade do Porto.**

Orientador - Doutor Júlio Gil Vale
Carvalho, Professor Associado do Instituto
de Ciências Biomédicas Abel Salazar da
Universidade do Porto.

Co-orientador - Doutor Alessandro Bagnato,
Associate Professor in Università degli Studi
di Milano

DECLARAÇÃO

As conclusões apresentadas nesta dissertação são baseadas nos artigos científicos indicados:

Artigos publicados ou em publicação em revistas nacionais e internacionais com arbitragem científica:

Vasconcelos, J., A. Martins, M. F. Petim-Batista, J. Colaço, R. W. Blake, and J. Carvalheira. 2004. Prediction of Daily and Lactation Yields of Milk, Fat, and Protein Using an Autoregressive Repeatability Test Day Model. *J. Dairy Sci.* 87:2591-2598

Vasconcelos, J., A. Martins, A. Ferreira, S. Pinto, e J. Carvalheira. 2005. Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal. *Revista Portuguesa de Ciências Veterinárias* 100(553-554): 33-38.

Vasconcelos, J., A. Martins, A. Ferreira, e J. Carvalheira. 2006. Consequências da Eliminação de Rebanhos Pequenos da Avaliação Genética de Bovinos Leiteiros em Portugal. *Revista Portuguesa de Zootecnia*, ano XII nº2: 105-117.

Vasconcelos, J., F. Santos, A. Bagnato, and J. Carvalheira. 2008. Effects of Clustering Herds with Small Size Contemporary Groups in Dairy Cattle Genetic Evaluations. *J. Dairy Sci.* 91:377:384.

Vasconcelos, J., A. Bagnato, A. Pereira, and J. Carvalheira. Evaluation of the Impact of Clustering Herds with Small Size Contemporary Groups on Consecutive Genetic Evaluations of Dairy Cattle. Submitted.

Resumos em actas de Congressos:

Vasconcelos, J., A. Martins, M. F. Petim-Batista, A. Silvestre, A. Ferreira, J. Colaço, and J. Carvalheira. 2002. Estimation of lactation yields from test day yields of milk, fat and protein for Portuguese Holstein cattle using an autoregressive test day model. 7th World Congress on Genetics Applied to Livestock Production. 2002. Proc. 7th World Cong. Gen. Appl. Livest. Prod., CD-ROM communication n° 01-57.

Vasconcelos, J., A. Martins, M.F. Petim-Batista, A. Silvestre, J. Colaço, A. Ferreira, e J. Carvalheira. 2002. Validação do Modelo de Contrastes Auto-regressivo para a Predição de Lactações Padronizadas aos 305 dias, de Leite, Gordura e Proteína do Efectivo Leiteiro Nacional. XII Congresso de Zootecnia. Vila Real, 21 a 23 de Novembro de 2002, p.519- 524.

Vasconcelos, J., A. Martins, A. Ferreira, S. Pinto, e J. Carvalheira. 2003. Análise Preliminar dos Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal. XIII Congresso de Zootecnia. Évora, 1 a 4 de Outubro de 2003, p.52.

Vasconcelos, J., e J. Carvalheira. 2003. Monitoring herd performance in dairy farms using Test-Day models. EAAP – 54th Annual Meeting, Rome, Italy. 31 August – 3 September 2003, Book of Abstracts No. 9 (2003), 170.

Vasconcelos, J., e J. Carvalheira. 2004. Análise Preliminar da Conexão Genética entre Grupos de Maneio na Raça Holstein-Frísia em Portugal. XIV Congresso de Zootecnia. Angra do Heroísmo, 7 a 9 de Outubro de 2004, p.98-99.

Vasconcelos, J., e J. Carvalheira. 2005. Effect of Small-Herd Clustering on the Genetic Connectedness of the Portuguese Holstein Cattle Population. EAAP – 56th Annual Meeting, Uppsala – Suedden. 5 – 8 June 2005, Abstract no. 424.

Vasconcelos, J., A. Martins, A. Ferreira, e J. Carvalheira. 2005. Consequências da Eliminação de Pequenos Rebanhos da Avaliação Genética de Bovinos Leiteiros em Portugal. XV Congresso de Zootecnia. Vila Real, 3 a 5 de Novembro de 2005, p.160-163.

Vasconcelos, J., F. Santos, R. Barroso, A. Martins, A. Ferreira, and J. Carvalheira. 2006. Effects of Clustering Dairy Herds for Genetic Evaluations using Different Descriptors to Define Similarities between Production Environments. 8th World Congress on Genetics Applied to Livestock Production. Proc. 8th World Cong. Gen. Appl. Livest. Prod., CD-ROM communication n° 24-30.

Vasconcelos, J., and J. Carvalheira. 2007. Assessment of heterogeneity of residual variances in an autoregressive test-day multiple lactations model. EAAP – 58th Annual Meeting, Dublin – Ireland. 26 – 29 of August 2007, Session 10, Poster 39.

No cumprimento do disposto no nº2 do Artigo 8º do Decreto-lei nº388/70, como o autor desta dissertação, declaro que participei na concepção, execução e interpretação dos resultados dos trabalhos que estiveram na base destes artigos. Retenho todos os direitos de autor relativos a esta dissertação e o direito de a usar em trabalhos futuros (como artigos ou livros).

Este trabalho foi realizado com o apoio financeiro da Fundação para a Ciência e a Tecnologia (FCT) através da bolsa de doutoramento SRFH/BD/11397/2002.

Agradecimentos

A realização deste trabalho não seria possível sem a colaboração e apoio de todos aqueles a quem desde já agradeço. Um especial obrigado ao meu orientador Professor Doutor Júlio Carvalheira, por todo o apoio e incentivo ao longo do trabalho. Durante os últimos anos pacientemente transmitiu e ensinou-me a correcta metodologia de trabalho de um investigador, desde a simples organização de tarefas, à solução de problemas, que enfrenta sempre com boa disposição e sentido de humor. Agradeço também ao Professor Doutor Alessandro Bagnato, meu co-orientador, pelo incentivo inicial e total disponibilidade que permitiu a realização desta tese.

Uma palavra de grande apreço à Rosa, à Adriana e ao Filipe, do departamento de informática do CIBIO, por toda a ajuda e apoio na solução dos problemas técnicos durante estes anos e pela constante boa disposição que em muito contribuiu para a realização deste trabalho.

Um especial agradecimento também a todos os funcionários da ACAV pela sua total disponibilidade e pela forma carinhosa com que sempre me trataram, fazendo com que me sentisse no centro como em minha casa.

Ao Centro de Estudos em Biodiversidade e Recursos Genéticos, por me ter integrado no seu grupo de investigadores e por me ter criado as condições necessárias à realização deste estudo.

Gostaria de agradecer também à Associação Nacional para o Melhoramento de Bovinos Leiteiros, Associação Portuguesa de Criadores de Raça Frísia e Associação para o Apoio à Bovinicultura Leiteira do Norte pela cedência da base de dados de contastes nacional. Um especial agradecimento aos Engenheiros Lima Martins e Ferreira pela sua constante disponibilidade e rapidez na transferência da informação.

À School of Biological Sciences, Institute of Cell, Animal and Population Biology da Universidade de Edimburgo, por todos os conhecimentos que me transmitiram durante a minha estadia na Escócia, em especial à Doutora Judith McQueen, pela forma amistosa como me recebeu.

À Fundação para a Ciência e a Tecnologia que desempenhou um papel decisivo para a realização deste projecto através da atribuição de uma bolsa de doutoramento, graças à qual pude dedicar estes anos unicamente à investigação livre de preocupações de natureza material.

Agradeço a todos os meus amigos pelo apoio incondicional que me deram desde sempre, constituindo uma importante motivação para a concretização deste projecto. Quero, por fim, agradecer à minha família, por todo o apoio e compreensão demonstrados ao longo destes anos, ela que foi a mais sacrificada pelas minhas frequentes ausências e indisponibilidades.

A todos, o meu mais sincero obrigada.

ÍNDICE

Resumo	9
Abstract	13
Résumé	17
1. INTRODUÇÃO	21
1.1. Dados Indicativos do Rebanho Nacional de Bovinos Leiteiros	21
1.2. Estado Actual de Conhecimentos	23
1.2.1. Modelos de Avaliação Genética de Bovinos Leiteiros	23
1.2.2. Formação de Clusters de Grupos Contemporâneos	26
1.2.3. Rebanho-Ano-Mês: Efeito Fixo vs Aleatório	27
1.3. Avaliação Genética de Bovinos Leiteiros em Portugal	29
2. OBJECTIVOS	34
3. RESULTADOS	36
3.1. Prediction of Daily and Lactation Yields of Milk, Fat and Protein Using an Autoregressive Repeatability Test Day Mode	37
<i>Predição das Produções Diárias e de Lactações de Leite, Gordura e Proteína pelo Modelo de Contrastes Auto-Regressivo – resumo traduzido</i>	38
3.2. Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal	47
<i>Inbreeding and Inbreeding Depression in the Portuguese Dairy Cattle</i>	47
3.3. Consequências da Eliminação de Rebanhos Pequenos da Avaliação Genética de Bovinos Leiteiros em Portugal	53
<i>Effects of the Exclusion of Small Herds from the Genetic Evaluations of Dairy Cattle in Portugal</i>	53
3.4. Effects of Clustering Herds with Small Size Contemporary Groups in Dairy Cattle Genetic Evaluations	67
<i>Efeitos da Aglomeração de Rebanhos com Grupos Contemporâneos Pequenos na Avaliação Genética – resumo traduzido</i>	68

3.5. Evaluation of the Impact of Clustering Herds with Small Size Contemporary Groups on Consecutive Genetic Evaluations of Dairy Cattle	77
<i>Avaliação do Impacto da Aglomeração de Rebanhos com Grupos Contemporâneos Pequenos em Avaliações Genéticas Consecutivas –</i>	78
resumo traduzido	
4. CONCLUSÕES GERAIS	88
5. BIBLIOGRAFIA	95
6. ANEXOS	100

RESUMO

O modelo matemático para análise de contrastes (produção diária) auto-regressivo (**MCAR**), cuja aplicação esteve em estudo no presente trabalho, difere dos restantes modelos de contrastes por possuir uma estrutura de covariância não genética adicional entre as observações do animal. Ao assumir que a produção de leite resulta da expressão do mesmo grupo de genes durante a vida produtiva do animal, os sucessivos contrastes são considerados medidas repetidas dentro e entre as lactações. Como tal, são modelados como variáveis longitudinais por processos auto-regressivos de primeira ordem. Os MCAR têm demonstrado elevada precisão e rigor nas avaliações genéticas dos animais. No entanto, antes da sua sistematização no programa nacional, ficou clara a necessidade de estudar o impacto da sua aplicação.

Numa fase inicial da investigação foi feito um estudo comparativo entre modelos clássicos e o MCAR de forma a determinar as diferenças de precisão na estimativa da extensão e projecção de lactações incompletas para os 305 dias. A partir de lactações completas, foram eliminados contrastes de modo a simular lactações incompletas, reflectindo a frequência de lactações incompletas ou em curso observadas na base de dados nacional. As diferenças entre lactações padronizadas de 305 dias reais (estimadas a partir de lactações completas pelo método do intervalo entre testes) e as estimadas pelo MCAR a partir da simulação das lactações incompletas foram insignificantes, e menores do que as estimadas pelo método do intervalo entre testes. A correlação entre lactações padronizadas reais e projectadas foi progressivamente mais elevada quanto maior o número de contrastes disponível para a extensão das lactações (de 0,831 para 0,997, com 2 e 9 contrastes, respectivamente). Valores de correlação entre contrastes reais e estimados foram superiores a 0,50, mesmo nos casos de estimativa baseada em contrastes reais com 7 meses de intervalo, sendo que as correlações médias por lactação variaram entre 0,977 e 0,984. Estes valores de correlação revelam uma elevada capacidade de predição do MCAR, cuja aplicação pode contribuir significativamente para a diminuição do intervalo entre gerações, o que representará uma valiosa ferramenta para o melhoramento animal e para as decisões de manejo e gestão dos efectivos.

A investigação prosseguiu com um estudo cujo objectivo foi o de caracterizar a estrutura do pedigree nacional, consistindo na determinação dos níveis de endogamia na população dos bovinos leiteiros em Portugal incluindo o efeito depressivo da endogamia

na expressão dos principais caracteres produtivos da espécie: leite, gordura e proteína. A análise do pedigree nacional revelou uma proporção de animais endogâmicos de 8,6%, com um coeficiente de endogamia médio de 1,8%. O número de fundadores efectivos da população actual de fêmeas estimado foi 276,5, verificando-se uma tendência para a redução deste número nas últimas gerações, indicando uma importante perda na variabilidade genética na população. Os 10 touros geneticamente mais influentes contribuíram com 21% para o “pool” de genes da população actual. Apesar do efeito da depressão de endogamia obtido neste estudo não parecer causar reduções importantes na produção ($-6,6 \pm 1,5$ kg, $-0,17 \pm 0,052$ kg e $-0,18 \pm 0,042$ kg de leite, gordura e proteína, respectivamente, por cada 1% de aumento no coeficiente de endogamia), o aumento intencional da endogamia deve ser evitado, pois os efeitos negativos no animal são globais e cumulativos, acarretando perdas económicas significativas, especialmente ao nível dos caracteres do “fitness”. A importância da qualidade do ficheiro de pedigree foi assim demonstrada, sendo urgente a previsão dos níveis de endogamia e o valor genético da descendência em futuros programas de emparelhamento, instrumento fundamental de apoio aos programas de selecção e melhoramento.

Uma das limitações da aplicação de modelos de contrastes que definam os grupos contemporâneos como efeito fixo é a necessidade de formar “clusters” de explorações pequenas para garantir um número mínimo de observações por nível de rebanho-ano-mês. No entanto, a aplicação de técnicas de formação de “clusters” pode resultar numa associação sistemática entre touros e grupos de manejo e na consequente perda de conexão genética entre explorações, da qual depende o rigor de predição das avaliações. O nível produtivo e de manejo da exploração define, na maioria dos casos, a qualidade genética do sêmen utilizado, o que pode resultar na formação de grupos de explorações sem ligações genéticas entre si. Por este motivo, foram investigadas as consequências da aplicação de técnicas de aglomeração de explorações pequenas na conexão genética da base de dados nacional, a partir da qual foram produzidas 3 bases de dados: uma contendo todas as explorações, outra onde as explorações pequenas com grupos contemporâneos com menos de 4 observações foram aglomeradas com base na semelhança entre a média e desvio padrão na primeira lactação e uma última onde as explorações pequenas foram eliminadas. As correlações para os valores estimados da conexão genética entre as 3 bases de dados foram elevadas ($> 0,99$), sugerindo que a aglomeração de grupos contemporâneos não provocou alterações significativas na conexão genética, tornando-se assim numa solução viável para o problema da inclusão das explorações pequenas na avaliação genética.

Antes de avaliar em detalhe as consequências da aplicação de técnicas de formação de “clusters”, foi também estudado o impacto da eliminação de rebanhos pequenos nas avaliações genéticas. Os objectivos desta fase do estudo centraram-se na análise da variação do tamanho mínimo dos grupos contemporâneos, usando a informação nacional de contrastes para simular 3 bases de dados, com as pequenas explorações eliminadas progressivamente (mínimo de 3, 10 ou 15 observações por nível de efeitos fixos no modelo). Em média, verificou-se uma tendência decrescente no rigor de predição (0,868, 0,861 e 0,860) à medida que aumentou a restrição do número mínimo de observações por nível de efeitos fixos. As diferenças no ranking genético entre as bases de dados foram muito marcadas, com correlações baixas, especialmente entre touros (de 0,21 a 0,32). Concluiu-se que a exclusão de rebanhos pequenos é uma forma de tratamento preferencial de touros, o que inflaciona o seu valor genético e contribui para o aumento do enviesamento da avaliação genética. A exclusão de rebanhos deve ser limitada ao mínimo e o tamanho dos grupos contemporâneos deve ser garantido por processos de aglomeração de rebanhos.

Depois de confirmada a não influência das técnicas de agrupamento de explorações pequenas na conexão genética e após o estudo das consequências da eliminação de explorações (alternativa à aplicação da formação de “clusters”), a fase seguinte deste projecto consistiu na definição da técnica de aglomeração mais indicada para o efectivo nacional. Foram então avaliadas as consequências da inclusão de novas variáveis descritivas na definição de semelhanças entre explorações, assim como a introdução de novos parâmetros que permitissem um maior controlo do processo de aglomeração de explorações com número insuficiente de observações por nível de efeitos fixos no modelo. Foi comprovada a eficácia da análise de componentes principais na redução de 14 variáveis descritivas definidoras da semelhança entre explorações, para 5 vectores próprios que explicaram 88% da variância total. Comparativamente com a metodologia actualmente utilizada nas avaliações genéticas em Portugal, o procedimento que revelou o melhor ajustamento aos dados (“goodness-of-fit”) foi o que aglomerou apenas explorações com pelo menos um grupo contemporâneo com número inferior a 3 observações. Esta metodologia permitiu uma maior preservação da estrutura original dos dados, maximizando o número de grupos de comparação e revelando um maior progresso genético anual nos machos. As elevadas correlações entre os valores genéticos dos animais e ranking obtidos (0,97 e 0,97, respectivamente), assim como para variância do erro de predição e rigor de predição (0,85 e 0,82, respectivamente) sugeriram uma reduzida influência do processo de aglomeração na reordenação dos

animais, tendo sido confirmado pela elevada percentagem de animais em comum na lista do “top100” dos valores genéticos aditivos (86 machos e 89 fêmeas). No entanto, no final deste estudo ficou clara a necessidade de continuar a investigação no sentido de estudar as possíveis alterações no ranking dos valores genéticos aditivos dos animais de uma avaliação para a outra, assim como do respectivo rigor de predição.

Após ter sido definida a melhor metodologia a usar na aglomeração de explorações com o número de observações por grupo contemporâneo inferior ao requerido, foi estudado o impacto resultante das explorações poderem mudar de “cluster” entre avaliações consecutivas, quer a nível de alterações no ranking genético dos animais, quer do respectivo rigor de predição. A informação nacional de contrastes foi usada para construir bases de dados com o objectivo de representar 4 avaliações nacionais sucessivas, uma por quadrimestre. Após o processo de aglomeração, foram efectuadas avaliações genéticas nas 4 bases de dados usando o MCAR e os seus resultados comparados. As elevadas correlações obtidas entre os valores genéticos aditivos estimados em avaliações sucessivas, quer para o conjunto de todos os animais, quer separadamente entre sexos ($> 0,97$), indicou a inexistência de alterações importantes de uma avaliação para a outra. O “top 100” genético também não evidenciou grandes reordenações dos animais. Apesar da correlação entre as médias genéticas das explorações e respectivo ranking não ter sido muito alta ($> 0,64$), os seus valores não foram extremos, podendo ser justificados pelas práticas de manejo usualmente aplicadas nas explorações em Portugal. Desta forma, a inexistência de alterações significativas no ranking genético dos animais em avaliações sucessivas provocadas pela mudança da exploração de “cluster” torna viável a aplicação deste método de formação de “clusters” na avaliação genética dos caracteres produtivos e com interesse económico de bovinos leiteiros em Portugal.

Os estudos que compõem esta tese levaram à adopção do MCAR como metodologia de avaliação oficial por parte das estruturas nacionais e responsáveis pela avaliação genética das características de produção de bovinos leiteiros.

ABSTRACT

The auto-regressive test-day model (**ARTD**) presented on this work differs from the other test-day (**TD**) models for having an additional non-genetic covariance structure between the animal records. Assuming that milk production results from the expression of the same group of genes during the productive life of the animals, successive TD may be considered as repeated measurements within and between lactations and, therefore, modeled as longitudinal variables using first order auto-regressive processes. The application of ARTD models in genetic evaluations of dairy cattle revealed high accuracy of prediction and goodness-of-fit. However, before the application in the national evaluation program, it was necessary to study the impact of its use.

The first topic investigated in this thesis consisted of a comparison study to evaluate the difference in accuracy to predict cumulative 305-d lactation records between classic models and the ARTD model. From complete lactations, reduced datasets were created by randomly deleting TD records to mimic the pattern of missing TD measurements in the national database corresponding to incomplete or in-progress lactations. Differences between actual (estimates from complete lactations using the test interval method) and ARTD predicted 305-d yields were negligible and smaller than those predicted by the test interval method. The correlations between actual and predicted lactation yields increased with the number of available TD (from 0.831 to 0.997 with 2 and 9 TD records, respectively). Correlations between actual TD yields and corresponding predicted yields exceeded 0.50, even in the cases of prediction with a 7 month interval from the last actual available TD record. Average correlations by parity ranged between 0.977 and 0.984. These correlations values indicate a good predictive ability of the ARTD model whose application can contribute positively to the reduction of the generation intervals, accelerating the genetic improvement. Another important advantage concerns to the management information that may be available to the breeders.

The next study consisted of the characterization of the national pedigree file for the occurrence of inbreeding and inbreeding depression in the expression of production traits (milk, fat and protein) of the Portuguese Holstein dairy cattle. The proportion of inbred animals was 8.6% with an average inbreeding coefficient of 1.8%. The effective number of founders for the present female population was 276.5, with a decreasing trend in the last generations, indicating that there was an important loss in the genetic variability of this population. The 10 genetically most influential sires contributed with 21% for the gene pool of the present population. Although inbreeding depression was not enough to

cause large reductions in the expression of these traits (-6.6 ± 1.5 kg, -0.17 ± 0.052 kg and -0.18 ± 0.042 kg of milk, fat and protein, respectively, for each 1% increase in the inbreeding coefficient), intentional inbreeding should be avoided due to expected global negative and cumulative impact on the animals, inducing a significant economic loss, mainly for traits contributing for fitness. The importance of the quality of the pedigree file was hereby demonstrated, showing the urgency for computerized mating selection programs that include prediction of the inbreeding coefficient and breeding value of the progeny.

A limitation of the application of the TD models that define contemporary groups as fixed effects is the necessity to form clusters of small herds in order to assure a minimum number of observations by level of herd-TD. However, the application of clustering techniques can result in non-random associations between sires and management groups and may induce the loss of genetic connectedness between herds, which will influence negatively the accuracy of genetic evaluations. Management level of the herd may define the genetic quality of the semen used in A.I. and this may result in the formation of herd groups with no genetic connections. Therefore, it was a necessary to investigate the possible changes in genetic connectedness due to clustering small herds. Three datasets were produced using the national database: the first containing all the herds, the second, small herds (with contemporary groups having less than 4 observations) were clustered based on the simultaneous similarity between phenotypic mean and STD at first lactation. Finally, for the third dataset, all small herds were removed from the analysis. The high correlation obtained for the parameter studied in this work suggested that clustering will have a negligible effect on genetic connectedness. This study confirmed this procedure as a reliable method to joint herds with few observations, without significant changes on accuracy of prediction of breeding values.

Before studying in detail the consequences of the application of clustering techniques, it was necessary to evaluate the impact of eliminating small herds from genetic evaluations. The objectives of this study were centered on the analysis of the contemporary groups' size variation. The national database was used to simulate 3 datasets where small herds were progressively eliminated (minimum of 3, 10 or 15 TD records by level of fixed effects in the model). On average, there was a negative trend on accuracy (0.868, 0.861 and 0.860) according to the increasing level of restriction imposed on the three datasets. There were significant differences in the genetic ranking

between datasets, with low correlation values, especially among sires (from 0.21 to 0.32). Eliminating small herds from the genetic evaluations was considered a form of preferential treatment of sires, increasing the bias on the prediction of breeding values. The elimination of small herds should be avoided, emphasizing the need to evaluate clustering techniques to assemble contemporary groups in TD models.

After confirming the absence of the influence of clustering small herds on genetic connectedness and studying the consequences of eliminating small herds as an alternative to clusterization, the next step of this thesis consisted of the definition of the most appropriated clustering technique to apply to the national database. Principal components analysis was successfully used to summarize 14 descriptive variables to measure similarities among herds in 5 eigenvectors that explained 88% of the total variation. Compared with the current procedure used in the Portuguese genetic evaluations, clustering only those herds without a minimum of 3 TD records in at least one contemporary group was the one that provided the better goodness-of-fit for the data. This procedure was also the one that better preserved the original herd structure of the database, maximizing the number of herd groups and revealing the higher annual genetic progress of sires. The high correlations between estimated breeding values, rank, prediction error variance and accuracies of prediction (0.97, 0.97, 0.85 and 0.82, respectively), suggested that no major re-ranking is to be expected with the application of this clustering procedure. The high percentage of elite animals found in common among the datasets (86 males and 89 females in the top 100) also confirmed that conclusion.

Following the definition of the better clustering procedure for the Portuguese database, it was necessary to study the impact that herds changing clusters between consecutive evaluations might have on the animals breeding value and respective ranking. The national database was used to build 4 datasets representing 4 consecutive genetic evaluations, taken 4 months apart. After the clustering process, an ARTD model was applied to all datasets and the resulting genetic parameters were compared. The high correlations obtained for the estimated breeding values for all animals and within sex (> 0.97) between datasets indicated that the clustering procedure did not induce major changes between consecutive evaluations. Furthermore, the list of the top 100 animals did not evidence the existence of major re-rankings. Although correlations between herd genetic means and respective rank were lower than those among animals (> 0.64), values were not extreme and may be explained by the normal farm management dynamics occurring in the Portuguese dairy cattle population. These results show that

this agglomeration methodology may be applied in future dairy cattle genetic evaluation of yield traits in Portugal.

The studies which are the basis of this thesis lead to the adoption of the ARTD model as the official methodology for the genetic evaluations of the Portuguese dairy cattle by the Holstein breeders associations.

RÉSUMÉ

Le modèle de contrôles laitiers autorégressif (**MCLAR**) présenté sur ce travail diffère des autres modèles par sa structure non génétique additionnelle de covariance entre les productions mensuelles de l'animal. Supposant que la production de lait résulte de l'expression du même groupe de gènes pendant la vie productive de l'animal, les successifs contrôles laitiers (**CL**) mensuelles sont considérés des mesures répétées à l'intérieur et entre les lactations. En conséquence, ils sont modélisés comme des variables longitudinales par des processus autorégressifs de premier ordre. Les MCLAR ont démontré une précision et une rigueur élevées dans les évaluations génétiques des animaux. Cependant, avant son application dans le programme national d'évaluation, il était nécessaire d'étudier l'impact de son utilisation.

Dans une phase initiale de la recherche, une étude comparative entre des modèles classiques et MCLAR a été faite pour déterminer les différences de précision pour prévoir l'extension et la projection de lactations incomplètes pour les 305 jours. À partir de lactations complètes, les CL ont été éliminés afin de simuler des lactations incomplètes en reflétant la fréquence de lactations incomplètes ou en cours, observées dans la base de données nationale. Les différences entre des lactations standardisées de 305 jours réels (évaluations des lactations complètes en utilisant la méthode d'intervalle de CL, utilisée dans la dernière évaluation génétique publiée au Portugal) et estimées par le MCLAR à partir de la simulation des lactations incomplètes, ont été négligeables et inférieures aux estimées par la méthode de l'intervalle entre CL. La corrélation entre des lactations standardisées réelles et projetées a augmenté avec le nombre de CL disponibles (de 0,831 à 0,997 avec 2 et 9 CL, respectivement). Des valeurs de corrélation entre des CL réels et estimées ont été supérieures à 0,50, même dans les cas d'estimation basée sur des CL réels avec 7 mois d'intervalle. Les corrélations moyennes par lactation ont varié entre 0,977 et 0,984. Ces chiffres de corrélation indiquent une capacité élevée de prévision de MCLAR, dont l'application peut contribuer significativement à la diminution des intervalles entre générations, ce qui représentera un outil précieux pour l'amélioration génétique. Une autre source importante c'est l'avantage apporté par l'information de gestion qui peut être disponible aux sélectionneurs.

La recherche a continué avec une étude dont l'objectif était de caractériser la structure du fichier du pedigree national pour déterminer l'occurrence des niveaux d'endogamie chez les bovins laitiers portugais incluant l'effet dépressif de l'endogamie dans les traits

de production les plus importants de l'espèce (lait, graisse et protéine). L'analyse du pedigree national a révélé une proportion d'animaux endogamiques de 8,6%, avec un coefficient d'endogame moyen de 1,8%. Le nombre de fondateurs effectifs de la population actuelle de femelles a été de 276,5, avec une tendance décroissante dans les générations plus récentes, indiquant une perte importante de la variabilité génétique dans la population. Les 10 taureaux génétiquement plus influents ont contribué avec 21% au pool de gènes de la population actuelle. Malgré que l'effet de la dépression d'endogamie obtenu dans cet étude n'ait pas été assez pour causer de grandes réductions de l'expression de ces traits ($6,6 \pm 1,5$ kg, $0,17 \pm 0,052$ kg et $0,18 \pm 0,042$ kg de lait, graisse et protéine, respectivement, par chaque 1% d'augmentation dans le coefficient d'endogamie), l'augmentation intentionnelle de l'endogamie doit être évitée. Ses effets négatifs dans l'animal sont globaux et cumulatifs, causant des pertes économiques significatives, particulièrement au niveau des caractères de "fitness". L'importance de la qualité du fichier de pedigree a été ainsi démontrée, étant urgente la prévision des niveaux d'endogamie et la valeur génétique de la descendance dans de futurs programmes automatisés de choix d'appariement, qui sont des instruments d'aide fondamentaux aux programmes de sélection et l'amélioration.

Une des limitations de l'application de modèles de CL qui définissent les groupes contemporains comme un effet fixe c'est la nécessité de former des grappes de petits troupeaux afin d'assurer un nombre minime d'animaux par niveau de troupeaux -CL. Cependant, l'application de techniques de formation de grappes peut résulter dans une association systématique entre des taureaux et des groupes de maniement et aussi dans la conséquente perte de connexion génétique entre des troupeaux, de laquelle dépend la fiabilité de prévision des évaluations. Le niveau productif et de maniement des troupeaux définissent, dans la plupart la formation de groupes des troupeaux sans liaisons génétiques. Pour cette raison, il fallait étudier les conséquences de l'application de techniques d'agglomération de petits troupeaux dans la connexion génétique de la base de données nationale. A partir des informations des CL nationales on a créé 3 bases de données: La première contenant tous les troupeaux, une autre où les petits troupeaux (avec des groupes contemporains ayant moins de 4 CL) ont été agglomérés ayant comme base la similitude entre la moyenne et le détournement phénotypique dans la première lactation, et une troisième où les petits troupeaux ont été éliminés de l'analyse. Les corrélations obtenues pour les valeurs de la connexion génétique entre les 3 bases de données ont été élevées (0,99), suggérant que l'agglomération de groupes contemporains n'ait pas provoqué de modifications significatives dans la connexion

génétique, confirmant ce procédé comme une méthode fiable pour joindre des petits troupeaux dans l'évaluation génétique.

Avant d'étudier en détail les conséquences de l'application de techniques de formation de grappes, il était nécessaire d'évaluer l'impact d'éliminer de petits troupeaux des évaluations génétiques. Les objectifs de cette phase de l'étude étaient centrés dans l'analyse de la variation de la dimension minimale des groupes des contemporains, utilisant les informations nationales de CL pour simuler 3 bases de données, avec les petits troupeaux éliminées progressivement (minimum de 3, 10 ou 15 CL par niveau d'effets fixes dans le modèle). En moyenne, on a vérifié une tendance négative sur l'exactitude de prévision (0,868, 0,861 et 0,860) autant que la restriction du nombre minimale de CL par niveau d'effets fixes augmentait. Les différences dans la rangée génétique entre les bases de données ont été beaucoup marquées, par de basses valeurs de corrélation, surtout parmi des taureaux (de 0,21 à 0,32). On est arrivé à la conclusion que l'exclusion de petits troupeaux était une forme préférentielle de traitement de taureaux ce qui fait monter la valeur génétique et contribue à l'augmentation du biais de l'évaluation génétique. L'exclusion de troupeaux doit être limitée au minimum et la dimension des groupes contemporains doit être garantie par des processus d'agglomération de troupeaux.

Après avoir confirmé que les techniques d'agroupement des petits troupeaux n'influençaient pas la connexion génétique et après l'étude des conséquences de l'élimination des troupeaux (alternative à l'application de la formation de grappes), on a passé à la phase suivante de ce projet qui consistait à définir la technique d'agroupement la plus appropriée à l'ensemble de données nationales. Alors, les conséquences de l'inclusion de nouvelles variables descriptives dans la définition de similitude parmi les troupeaux ont été évaluées, ainsi comme l'introduction de nouveaux paramètres qui permettaient un plus grand contrôle du processus d'agglomération des troupeaux avec un nombre insuffisant de CL par niveau d'effets fixes dans le modèle. L'analyse de composants principaux a été employée avec succès pour réduire 14 variables descriptives pour 5 vecteurs propres qui ont expliqué 88% de la variation totale. Comparativement à la méthodologie utilisée actuellement dans les évaluations génétiques portugaises, celle qui a révélée la meilleure qualité d'ajustement aux données a été celle qui agglomérait seulement des troupeaux ayant au moins un groupe contemporain avec un nombre inférieur à 3 CL. Cette méthodologie a permis une plus haute conservation de la structure originelle des données, maximisant le nombre de groupes de comparaison et révélant un progrès génétique annuel supérieur pour les

mâles. Les corrélations élevées entre les valeurs génétiques des animaux, les rangées, les variances d'erreur de prévision et l'exactitude de prévision obtenus (0,97, 0,97, 0,85 et 0,82, respectivement) ont suggéré une influence réduite de la méthodologie d'agglomérations dans la réordination des animaux. Ceci a été confirmé par le grand pourcentage d'animaux en commun dans les listes de la "top100" des valeurs génétiques additives (86 mâles et 89 femelles). Cependant, cette étude a confirmé la nécessité de continuer la recherche ayant comme but l'étude des modifications possibles dans le ranger des valeurs génétiques additives des animaux entre évaluations, ainsi que de la respective exactitude de prévision.

Après la définition de la meilleure méthodologie d'agglomération des petits troupeaux sans nombre minimum exigé de CL par groupe contemporain, on a analysé quel était l'impact résultant des changements de cluster par les troupeaux, entre des évaluations consécutives, soit au niveau des modifications du rangé génétique des animaux, soit au niveau de la rigueur de prévision. Les informations nationales de CL ont été utilisées pour construire des bases de données avec l'objectif de représenter 4 évaluations nationales successives une par quadrimestre. Après le processus d'agglomération, des évaluations génétiques ont été effectuées dans les 4 bases de données en utilisant MCLAR et leurs résultats comparés. Les élevées corrélations obtenues entre les valeurs génétiques additives estimées dans des évaluations consécutives, soit pour l'ensemble de tous les animaux, soit séparément entre des sexes (0,97), a indiqué l'inexistence de modifications importantes. Le "Top 100" génétique n'a même pas prouvé de grandes réordinations des animaux. Malgré que la corrélation entre les moyennes génétiques des explorations et leurs rangés n'aie pas été très haute (0,64), ces chiffres ne peuvent pas être considérés extrêmes et sont justifiés par la dynamique de gestion usuellement appliquée aux troupeaux portugais. De cette forme, l'absence de modifications significatives dans le rangé génétique des animaux entre évaluations successives provoquées par le changement de troupeaux en grappe, suggère que cette méthode de formation de grappes pourrait être appliquée dans de futures évaluations génétiques des caractères productifs et avec intérêt économique de bovins laitier aux Portugal.

Les études qui composent cette thèse ont emmené à l'adoption de MCLAR comme méthodologie d'évaluation officielle de la part des structures nationales et responsables par l'évaluation génétique des caractéristiques de production de bovines laitières.

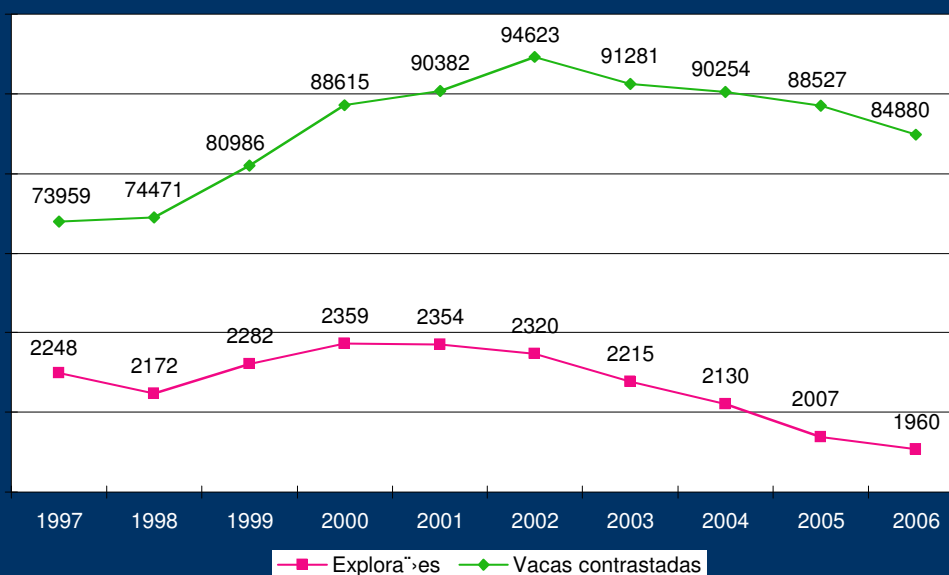
1. INTRODUÇÃO

1.1. Dados Indicativos do Rebanho Nacional de Bovinos Leiteiros

Em Portugal existem aproximadamente 1.412.000 registos no ficheiro de pedigree dos animais da raça Holstein (em Junho 2007), dos quais cerca de 328.000 são vacas leiteiras com valores de produção (GPPAA, Anuário Pecuário, 2005). O sector leiteiro representa 10,2% do total da produção agrícola nacional, cujo peso é de 2,7% no VABpb (Valor Acrescentado Bruto a preços de base) da economia Portuguesa (GPPAA, Agricultura Portuguesa – Principais Indicadores, 2006).

Nos últimos 10 anos o sector leiteiro sofreu uma profunda reestruturação com consequências importantes no efectivo. O número de animais com contrastes válidos aumentou entre 1997 e 2002, decrescendo desde então (Figura 1). Factores como a implementação do sistema de cotas leiteiras, a introdução de medidas mais rigorosas de qualidade higio-sanitária do leite (Directiva CEE 2597/97), o surto de encefalopatia espongiforme bovina (foram registados em Portugal, no período de 1990 a 2006, 1002 casos, dos quais 52 no ano de 2005; Direcção Geral de Veterinária, Boletim Estatístico nº11, 2005) e o aumento dos valores do custo de produção sem o respectivo acompanhamento do preço por quilograma de leite pago aos produtores, foram responsáveis pela diminuição da margem de lucro das explorações. Consequentemente, foi inevitável a falência das explorações menos competitivas e de menor dimensão e as restantes, para se tornarem economicamente viáveis, maximizaram a utilização dos seus recursos aumentando os seus efectivos.

Figura 1. Evolução do contraste leiteiro nos últimos 10 anos (ANABLE, publicação de resultados, 2006)



Na Figura 1 está representado o número de explorações contrastadas em Portugal continental, verificando-se a extinção de aproximadamente 300 explorações de 1997 a 2006, verificando-se, no entanto, um aumento do número médio de vacas por exploração contrastada no mesmo período (ANABLE, Publicação de Resultados, 2006). Nesse sentido, o aumento da rentabilidade produtiva do efectivo leiteiro resultou num acréscimo da produção nacional de leite de vaca, que no período de 1998 a 2004 passou de 1.850×10^3 para 2.010×10^3 toneladas (ANABLE, Publicação de Resultados, 2006).

Na União Europeia (25 Estados Membros), entre 2000 e 2004, registou-se igualmente uma tendência para a redução do número de produtores (de 1.703.286 em 2000 para 1.626.306 em 2004) e para o aumento do rendimento produtivo por vaca [5.843kg de leite/vaca/ano em 2003 para 5.970 kg de leite/vaca/ano em 2004 (GPPAA, Anuário Pecuário, 2005)]. A União Europeia registou uma diminuição do efectivo leiteiro de 2,4% no período de 2003 a 2004, tendo sido a Alemanha, França, Espanha, Reino Unido e Dinamarca os países que mais contribuíram para esta redução (GPPAA, Anuário Pecuário, 2005).

1.2. Estado Actual de Conhecimentos

1.2.1. Modelos de Avaliação Genética de Bovinos Leiteiros.

Num programa de selecção, para que seja feita a escolha dos animais geneticamente superiores em relação aos caracteres objectivo do melhoramento genético, é necessário prever o mérito genético dos indivíduos a partir da informação fenotípica disponível, quer do próprio animal, quer dos seus familiares. Para tal é necessário considerar os efeitos ambientais a que estão sujeitos os candidatos à selecção. Na avaliação genética dos caracteres quantitativos de animais domésticos são usados modelos matemáticos que pretendem representar as interacções entre o seu genótipo e o ambiente que os rodeia. Os modelos que melhor reflectem esta interacção, revelam componentes residuais ambientais mais pequenos (Carvalheira et al., 1998; Gengler et al., 1999; Gengler et al., 2001; Carvalheira et al., 2002a), maximizando as diferenças genéticas entre os animais. Uma avaliação mais rigorosa do mérito genético dos candidatos à selecção impulsiona, por seu lado, um maior avanço no progresso genético entre gerações.

Os modelos clássicos de avaliação do mérito genético da produção leiteira utilizam a produção acumulada aos 305 dias de lactação. No entanto, foram demonstradas vantagens na utilização do valor diário da produção no dia do contraste (uma vez por mês), em detrimento da produção acumulada aos 305 dias (Ptak e Schaeffer, 1993; Swalve, 1995; Jamrozik e Schaeffer, 1997; Jamrozik et al., 1997; Wiggans e Goddard, 1997; Carvalheira et al., 1998; Carvalheira et al., 2002a, entre outros). Sendo a produção acumulada aos 305 dias uma função linear dos contrastes (Carvalheira, 1996; Norman et al., 1999; Schaeffer et al., 2000), a sua estimativa implica a perda de graus de liberdade e consequentemente uma menor precisão na estimativa dos valores genéticos. Os modelos de contrastes permitem também uma definição mais exacta dos grupos contemporâneos (**GC**), ou seja de animais sujeitos a efeitos ambientais semelhantes (Carvalheira et al., 1998; Macciotta et al., 1999; Rekaya et al., 1999; Ptak e Zarnecki, 2000; Carvalheira et al., 2002a), pois a sua associação pode ser feita em função do ano e do mês de contraste em cada exploração [rebanho-ano-mês (**RAM**)], em vez do clássico rebanho-ano-estação em que se realizou o parto do animal. No entanto, enquanto os modelos clássicos são relativamente simples e com poucos parâmetros, os modelos de contrastes são bastante mais complexos e podem ter centenas de parâmetros (Misztal et al., 2000), limitando a sua aplicação à capacidade computacional disponível. O modelo de contrastes que considera cada dia da lactação como um carácter independente seria o modelo mais completo. Para três lactações,

cada uma com 305 dias, este modelo super parametrizado, seria constituído por um total de 915 parâmetros, número este que poderia triplicar ao incluir, para além da produção de leite, gordura e proteína como caracteres. Para além do elevado custo computacional, este modelo apresenta como desvantagem a formação de GC, que neste caso seriam formados apenas pelos contrastes independentes, com um reduzido número de animais. É portanto necessária a utilização de modelos simplificados, que consigam atingir um equilíbrio entre a complexidade e o rigor de predição dos valores genéticos dos animais (Misztal et al., 2000).

Os modelos de contrastes utilizam como base a informação dos contrastes cuja estrutura de (co)variância varia com os dias-em-lactação (**DEL**). Os vários modelos de contrastes desenvolvidos (“random regression”, “autoregressive repeatability”, “orthogonal polynomials”, “character processing models”) diferem na forma como modelam a curva de lactação e nas estruturas de covariância assumidas para os efeitos genéticos e ambientais. Um modelo de regressão aleatória (“random regression model”) assume a existência de variâncias e covariâncias funcionais relativamente aos DEL. Nestes modelos todas as covariâncias variam com o DEL e a sua forma é especificada pela função usada na regressão aleatória. Uma característica importante nestes modelos é o facto de atribuírem variâncias funcionais a todos os efeitos, inclusivamente ao resíduo, baseando-se no facto de que assumir variância constante a uma variável aleatória poderá introduzir enviesamento na estimativa da heritabilidade (Misztal et al., 2000). O modelo de repetibilidade assume a inexistência de (co)variâncias entre os efeitos genéticos aditivos e os DEL. Este difere dos restantes também pelo facto de possuir uma estrutura de covariância não genética adicional entre as observações repetidas do mesmo animal. Assumindo que a produção de leite durante uma lactação está dependente do mesmo grupo de genes, os sucessivos contrastes são medidas repetidas dentro e entre as lactações do animal, sendo essencial distinguir efeitos ambientais associados à performance da vaca que se repetem ao longo do tempo. Um modelo de repetibilidade tem em consideração covariâncias ambientais dentro e entre lactações (Carvalho, 1996).

Segue-se o exemplo de um modelo de repetibilidade que modela lactações múltiplas e respectivos contrastes em cada animal, onde as covariâncias entre as medições repetidas por cada animal são modeladas como variáveis longitudinais e apresentadas como processos auto-regressivos de primeira ordem dentro e entre lactações. As estruturas de auto-correlação impostas neste modelo permitem ajustar para os efeitos temporais (correlações entre as medidas sucessivas dos contrastes em cada vaca) que

de outra forma provocariam o enviesamento do componente genético (Carvalheira, 1996, Carvalheira et al., 1998; Carvalheira et al., 2002a):

$$\mathbf{y}_{ijkLmn} = \mathbf{RAM}_i + \mathbf{I(R)}_j + \mathbf{DEL(R)}_{k(L)} + \mathbf{a}_m + \mathbf{p}_{m(L)} + \mathbf{t}_{n(mL)} + \mathbf{e}_{ijkLmn}$$

Onde, $\mathbf{y}_{ijkLmn}$ representa o vector das produções (leite, gordura e proteína) no dia do contraste, $\mathbf{RAM}_i$ o efeito fixo do rebanho-ano-mês do contraste, $\mathbf{I(R)}_j$ o efeito fixo da idade (classificada em j classes) no parto, por rebanho, $\mathbf{DEL(R)}_{k(L)}$ o efeito fixo dos DEL por rebanho e ordem de lactação, $\mathbf{a}_m$ representa o efeito aleatório genético do animal, $\mathbf{p}_{m(L)}$ o efeito aleatório do ambiente a longo prazo, tomando em consideração as correlações entre lactações para cada vaca, $\mathbf{t}_{n(mL)}$ representa o efeito aleatório do ambiente a curto prazo, tomando em consideração as correlações entre contrastes para cada lactação e $\mathbf{e}_{ijkLmn}$ representa o efeito aleatório do resíduo no modelo. Ambos, $\mathbf{p}$ e $\mathbf{t}$ são variáveis que seguem processos auto-regressivos de primeira ordem.

A aplicação dos modelos de contrastes auto-regressivos (**MCAR**) tem demonstrado maior precisão e rigor nas avaliações genéticas dos animais quando comparados com outras avaliações, nomeadamente as obtidas com modelos clássicos de lactações acumuladas aos 305 dias (Vasconcelos et al., 2002a; Vasconcelos et al., 2002b; Vasconcelos et al., 2004). No entanto, a possibilidade dos modelos de contrastes particionarem os efeitos ambientais sistemáticos em períodos temporais de curta duração (*e.g.* um mês), o que se reflecte em menores variâncias residuais, pode representar um problema no caso de existência de rebanhos pequenos na base de dados. Em explorações pequenas o número de observações pode ser insuficiente para definir GC de animais, o que poderá limitar a utilização do RAM como efeito fixo no modelo.

1.2.2. Formação de Clusters de Grupos Contemporâneos

A formação de agrupamentos ou “clusters” de GC apresentada neste projecto é uma das possíveis soluções para o problema do número insuficiente de observações por classe de efeitos fixos no caso de explorações recentes ou de dimensão reduzida. Existe, no entanto, uma grande controvérsia na definição ideal dos GC. Se a sua definição abrange um longo período de tempo, o número de animais contemporâneos em cada classe é elevado, aumentando o rigor de predição da avaliação. No entanto, considerar animais sujeitos a diferentes ambientes e diferentes condições de manejo como pertencendo ao mesmo GC irá enviesar a avaliação genética dos mesmos (Crump et al., 1997; Carabaño et al., 2004). Torna-se assim necessário conseguir um equilíbrio entre o rigor de predição e o enviesamento das avaliações genéticas (Van Bebber et al., 1997).

No início da década de noventa foi proposto por Schmitz et al. (1991) a formação de “clusters” de animais com datas de parto próximas, na mesma estação e na própria exploração. Com o objectivo de obter um mínimo de 15 observações por classe de efeitos ambientais, agruparam os animais tendo em vista a distância entre contrastes, variando o intervalo de tempo entre um mês e o limite máximo de um ano. Na Polónia, Strabel e Szwaczkowski (1999), sugeriram a formação de “clusters” associando classes de RAM consecutivas dentro da mesma exploração ou entre explorações semelhantes. As classes de RAM, classificadas pela produção média da exploração e data do contraste, foram agrupadas tendo em consideração três parâmetros: distância máxima em dias entre os grupos, diferença máxima da média de produção entre grupos e, para evitar a junção de dois grandes GC, pelo menos um dos grupos teria menos de 5 observações. Mais tarde, Carabaño et al. (2004) forma “clusters” agrupando GC dentro da mesma exploração com o objectivo de obter o mínimo de 5 observações por classe de RAM. Em Portugal, Carvalheira (2000) defende um sistema de formação de “clusters” que permite conservar as 12 classes mensais por ano em cada rebanho, mantendo as associações temporais entre os efeitos fixos e as observações. Assumindo que a maior fonte de heterogeneidade de variância está associada à performance fenotípica das explorações, formou “clusters” de rebanhos, baseando-se na semelhança simultânea entre médias e desvios padrão fenotípicos à primeira lactação, usando o método das variâncias mínimas. Explorações no mesmo “cluster” teriam níveis de manejo semelhantes. A aplicação destas técnicas permite assim definir o número mínimo de observações por classe de efeitos fixos.

1.2.3. Rebanho-Ano-Mês: Efeito Fixo vs Aleatório

Assumir os GC como efeitos fixos ou aleatórios nos modelos tem gerado bastante polémica. Ao considerar este efeito como fixo, todos os animais são comparados com a média do grupo, assumindo que não existem diferenças genéticas entre GC e que as existentes são devidas a factores ambientais. É no cálculo desta média que a questão do número mínimo de observações por GC se torna relevante. Em alternativa, considerar o efeito dos GC como aleatório, assume que parte da diferença é devida a factores genéticos e parte a factores ambientais. Os animais são comparados com a média geral da raça Holstein e não com a média do GC. Autores como Henderson (1973; 1975), Meyer (1987), Schaeffer (1987), Van Vleck (1987), Visscher e Goddard (1993), entre outros, defendem a utilização dos GC como efeito fixo, pois tomam em consideração os efeitos não aleatórios, gerados por tendências sistemáticas de origem ambiental e genética. A sistematização de efeitos genéticos por “cluster”, isto é, a possibilidade de associação não aleatória entre touros e explorações, como por exemplo o tratamento preferencial (Meyer, 1987; Van Vleck, 1987; Tierney e Schaeffer, 1994) é uma situação frequente em Portugal, onde a dimensão da exploração ou o seu nível de manejo e produção determinam sistematicamente a qualidade de sêmen usado. Outra questão não menos importante relativa à aplicação dos GC como efeito fixo é a possibilidade de diminuir o número de componentes de variância a estimar (Wade, 1990).

Ao considerar o GC como efeito aleatório não é necessária a existência de um número mínimo de observações para efectuar o cálculo da média do grupo. Schaeffer e Sullivan (1994) sugeriram a utilização do RAM como efeito aleatório na análise da produção de leite em cabras devido, não só ao reduzido número de efectivo, mas também ao tempo insuficiente para o desenvolvimento de associações aleatórias importantes entre o mérito genético dos reprodutores e os efeitos ambientais do rebanho. Particionar as variáveis ambientais em efeitos fixos e aleatórios simultaneamente no modelo é uma alternativa que poderá ser viável, pois possibilita a utilização de pequenos GC. Em certas circunstâncias poderá tomar em consideração a introdução de enviesamento na estimativa dos valores genéticos, como nos casos de existência de associações não aleatórias entre touros e GC (Van Bebber et al., 1997). Carabaño et al. (2004) compara a aplicação de diversos modelos onde o RAM é representado como efeito fixo ou aleatório, com ou sem a introdução de um efeito fixo rebanho-ano, mas o seu estudo não foi conclusivo. A decisão da sua utilização como efeito fixo, aleatório ou ambos

simultaneamente é uma questão de balanço entre o rigor de precisão e o enviesamento na estimativa dos valores genéticos (Van Vleck, 1987; Visscher e Goddard, 1993; Van Bebber et al., 1997).

A aplicação dos GC como efeito fixo no modelo implica, como já foi dito, a existência de um número mínimo de observações por nível de efeitos fixos para maximizar o número efectivo de observações e, simultaneamente, minimizar o erro residual e a variância do erro de predição. Na Alemanha, Swalve (1995) define os GC com base no RAM e impõe um limite mínimo de observações anuais por exploração (60 lactações por ano). Esta restrição elimina as explorações com reduzido número de observações por GC. A aplicação deste limite implica a perda de parte da informação, que poderá ser significativa quando existe um elevado número de explorações de dimensões reduzidas, como é o caso da base de dados Portuguesa (Vasconcelos et al., 2008). A não utilização de parte da informação terá consequências negativas quer para o progresso genético (Swalve, 1995; Vasconcelos et al., 2005a; Vasconcelos et al., 2005c), quer para os programas de selecção individuais das explorações (Carabaño et al., 2004). Assumir o efeito dos GC como fixo implica encontrar uma solução técnica que permita utilizar o maior número possível de explorações e animais, minimizando confundimentos genéticos que possam enviesar a sua avaliação.

1.3. Avaliação Genética de Bovinos Leiteiros em Portugal

A selecção dos animais com base nos resultados da avaliação genética permite tomar decisões sobre a escolha dos melhores indivíduos que irão ser os reprodutores da próxima geração. Deles depende a futura performance da população (VanRaden e Smith, 1999), pelo que assume capital importância a correcta definição dos critérios de selecção no sentido de saber quais as características com maior interesse económico no futuro.

A Figura 2 representa o progresso genético para a produção de leite (a), gordura (b), proteína (c) e células somáticas (d) no efectivo português de vacas Holstein desde 1994 (efectuado no CIBIO em Março de 2007). Verifica-se um progresso genético anual médio de 18,63 kg (11,57 kg), 0,42 kg (0,25 kg), 0,61 kg (0,34 kg) e -0,0009 SCS (-0,0006 SCS) para a produção de leite, gordura, proteína e células somáticas para os machos (fêmeas), respectivamente. Apesar de serem valores elevados, especialmente se considerarmos apenas o período de 1990 a 2004, durante o qual se verificou um maior progresso genético (24,171 Kg e 16,255 kg para a produção de leite em machos e fêmeas, respectivamente), quando comparados com os progressos médios anuais de outros países, [e.g. 131 kg de leite por ano nos USA (AILP-USDA, 2007)] pode-se depreender que ainda existe um grande potencial genético por explorar no efectivo nacional.

Em Portugal a última avaliação genética do efectivo leiteiro nacional (Direcção Geral de Veterinária, 2003) foi feita utilizando modelos clássicos de análise do registo da produção acumulada aos 305 dias, apenas de lactações terminadas. As lactações com menos de 210 dias foram rejeitadas e as restantes projectadas para 305 dias, utilizando factores pré-calculados de ajustamento (Harris e Johnson, 1998). As restrições impostas por estes critérios (somente lactações terminadas e com mais de 210 dias) implicaram a não utilização de grande parte da informação disponível [15,74% das lactações terminadas com menos de 210 dias nos últimos 10 anos (ANABLE, 2006)]. A acrescentar a este número é ainda necessário considerar as lactações em curso.

Figura 2a. Progresso genético anual de fêmeas e machos Holstein nacionais para a produção de leite, por ano de nascimento (Carvalho, comunicação pessoal)

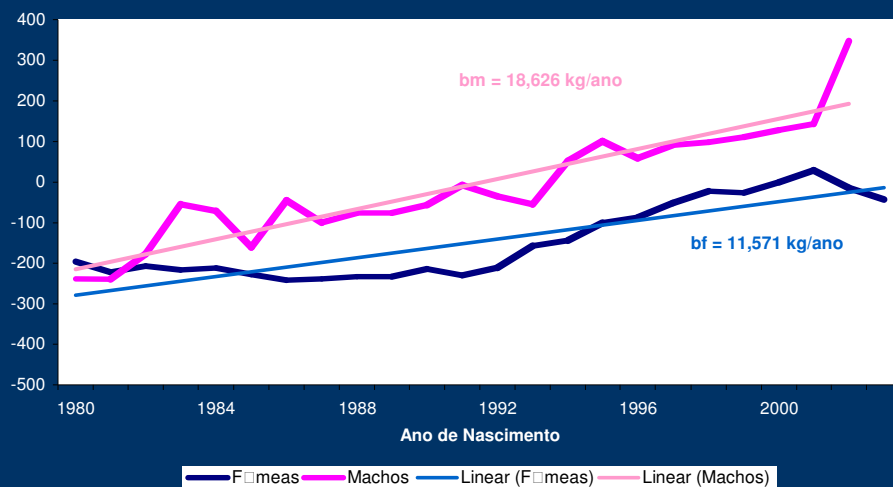


Figura 2b. Progresso genético anual de fêmeas e machos Holstein nacionais para a produção de proteína, por ano de nascimento (Carvalho, comunicação pessoal)

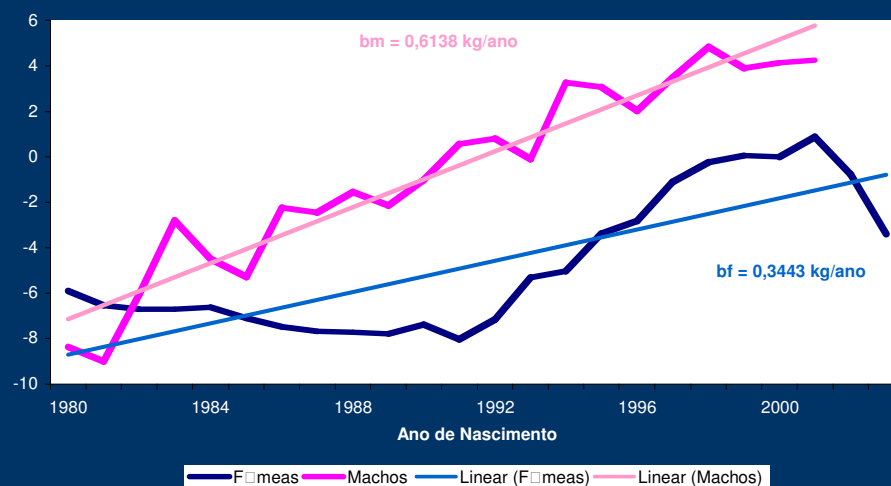


Figura 2c. Progresso genético anual de fêmeas e machos Holstein nacionais para a produção de gordura, por ano de nascimento (Carvalho, comunicação pessoal)

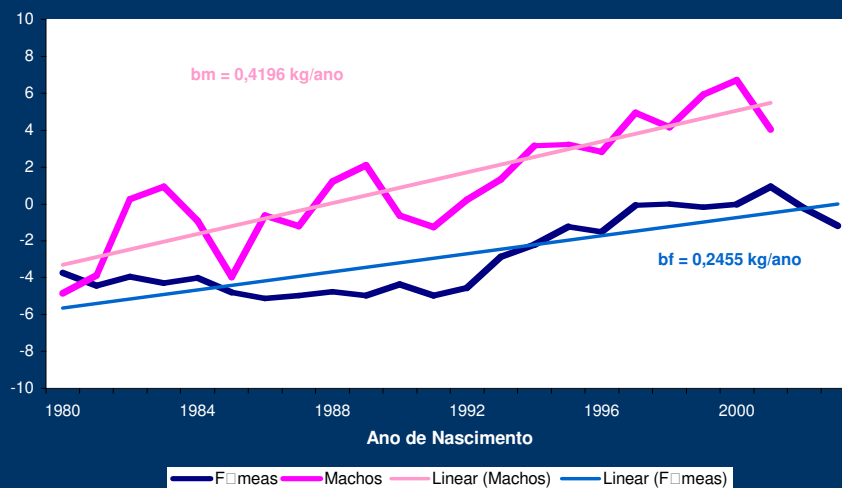
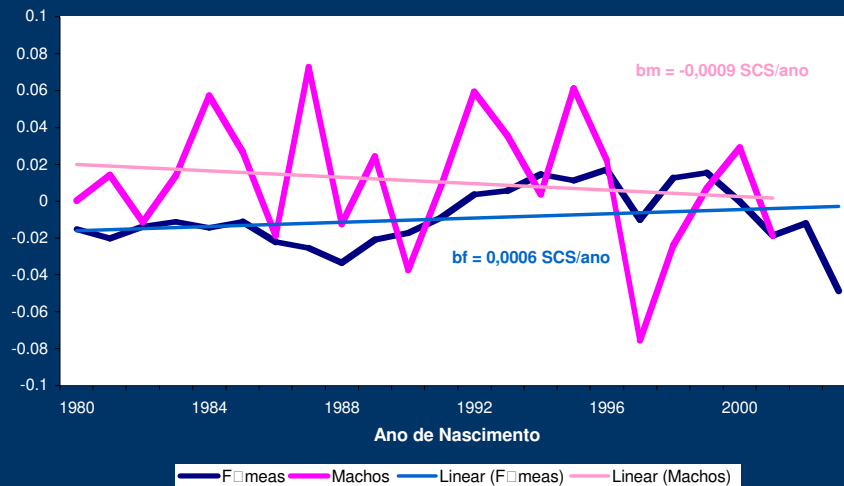


Figura 2d. Progresso genético anual de fêmeas e machos Holstein nacionais para a produção de células somáticas, por ano de nascimento (Carvalho, comunicação pessoal)



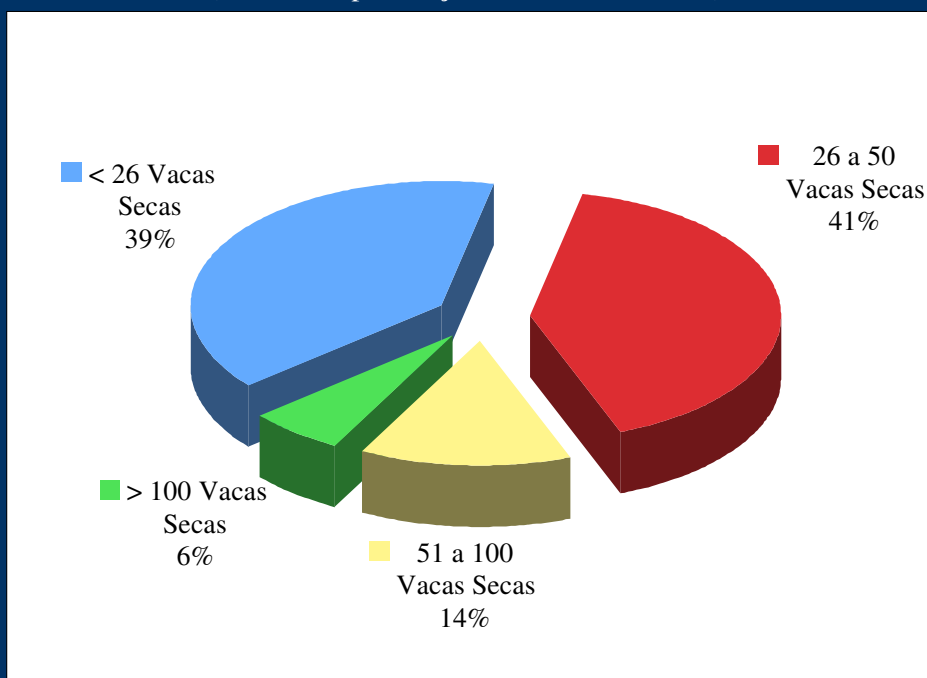
A adopção de modelos de contrastes elimina estas restrições, pois define o contraste como carácter de interesse e, como tal, todas as observações são usadas nas avaliações dos animais. Desta forma é possível dar oportunidade a um maior número de animais de contribuírem para a caracterização genética da população. A maximização do progresso genético da capacidade produtiva da população nacional de bovinos leiteiros é uma das questões críticas para o aumento da competitividade e da viabilização económica da indústria do sector. Os modelos de contrastes poderão contribuir muito positivamente para este fim.

O tratamento preferencial é uma situação frequente em Portugal. O maior uso de touros de pior qualidade genética em explorações de baixa média fenotípica e de touros de elevado potencial genético em explorações de elevada média fenotípica, produz associações não aleatórias entre touros e GC (Henderson, 1973; Schaeffer, 1991), uma propriedade não biológica dos dados que representa uma forma de selecção (Van Bebber et al., 1997). Assim, ao considerar o RAM como um efeito fixo no modelo, são consideradas tendências sistemáticas de origem ambiental e genéticas, minimizando o enviesamento nas avaliações genéticas em Portugal. No entanto, surge a necessidade da existência de um número mínimo de observações por nível de classes de efeitos fixos para definir GC de animais. Este facto poderá representar uma limitação no caso da base de dados nacional, cuja percentagem de explorações de reduzida dimensão é elevada (Figura 3). A percentagem de explorações com CG com menos de 15 ou 3 observações é de respectivamente 97% e 40%, (Vasconcelos et al., 2008). Se estas

explorações não forem incluídas nas avaliações genéticas, a estimativa dos valores genéticos dos touros será enviesada devido à perda de informação das suas filhas, com consequências negativas para o progresso genético (Vasconcelos et al., 2005a; Vasconcelos et al., 2005c).

Assumir o efeito dos GC como fixo implica a aplicação de uma solução técnica que permita utilizar o maior número possível de explorações e animais, minimizando confundimentos genéticos que possam enviesar a sua avaliação. A formação de “clusters” permite definir o número mínimo de observações por classe de efeitos fixos e desta forma solucionar o problema das explorações de pequena dimensão.

Figura 3. Distribuição das explorações em função das suas dimensões
(ANABLE, publicação de resultados, 2006)



No entanto, antes da sua aplicação e sistematização no programa de avaliações nacional, foi necessário estudar o impacto que a eliminação de pequenos rebanhos teria na estimativa dos valores genéticos dos animais (Vasconcelos et al., 2005a; Vasconcelos et al., 2005b). Importou ainda determinar os parâmetros para a formação de “clusters” que melhor se adaptaram à realidade nacional, nomeadamente a influência do número de observações por GC e a definição mais precisa dos critérios de semelhança a serem usados entre explorações a juntar (Vasconcelos et al., 2006; Vasconcelos et al., 2008). Em avaliações genéticas anteriores, as explorações

pequenas foram aglomeradas mantendo as 12 classes de RAM anuais. Apesar disso, apenas foi usado como critério para a aglomeração, a média e o desvio padrão fenotípico à primeira lactação, alegando que fenótipos semelhantes equivalem a níveis similares de manejo das explorações. O processo de aglomeração parava apenas quando todos os “clusters” formados tivessem o número mínimo de observações requerido. Parâmetros como o tamanho ou o número máximo de observações por “cluster” não eram controlados, estando apenas dependentes da árvore formada pela matriz das distâncias (Carvalheira, 2000). Tornou-se então necessário o estudo das consequências da inclusão de novas variáveis descritivas na definição de semelhanças entre explorações. Foi igualmente necessária a introdução de novos parâmetros que permitissem um maior controlo do processo de aglomeração, minimizando o número de explorações por “cluster”, maximizando o número de GC final, mas que assegurassem simultaneamente o mínimo requerido observações por nível de efeitos fixos (Vasconcelos et al., 2006; Vasconcelos et al., 2008). Importou ainda estudar o impacto em avaliações consecutivas que as mudanças de explorações de um “cluster” para o outro teriam ao nível de alterações no ranking genético dos animais e do respectivo rigor de predição no programa de selecção nacional (Vasconcelos, dados não publicados). Tornou-se, portanto, imperativa a realização de um estudo mais exaustivo das consequências da utilização de técnicas de formação de “clusters” antes da sua aplicação regular em avaliações genéticas.

Os artigos científicos que compõem esta tese resultaram da necessidade de tentar responder a estas questões. São igualmente anexados a esta dissertação, os trabalhos de investigação preliminares, apresentados sob a forma de resumos em actas de congressos, que deram origem aos artigos publicados.

2. OBJECTIVOS

1º Objectivo – Este projecto incluiu, numa fase inicial da investigação, o estudo comparativo entre os modelos clássicos e de contrastes auto-regressivos para determinar as diferenças de precisão no cálculo de lactações padronizadas de 305 dias. Pretendeu-se ainda avaliar a capacidade preditiva do MCAR para contrastes em falta em lactações incompletas e a respectiva extensão e projecção para os 305 dias. O objectivo da primeira fase do projecto consistiu assim no estudo do impacto da aplicação de MCAR na maximização do progresso genético, nomeadamente a sua influência no intervalo entre gerações e no incremento da precisão de predição dos valores genéticos dos reprodutores.

2º Objectivo – Numa segunda fase da investigação, foi efectuado um estudo com o objectivo de determinar os níveis de endogamia na população dos bovinos leiteiros em Portugal, analisar o impacto que os touros geneticamente mais influentes têm no “pool” genético desta população e determinar o efeito depressivo da endogamia na expressão dos principais caracteres produtivos da espécie: leite, gordura e proteína.

3º Objectivo – Antes de avaliar em detalhe as consequências da aplicação de técnicas de formação de “clusters”, foi efectuado um estudo com o objectivo de analisar as consequências da eliminação de pequenos rebanhos da avaliação genética de bovinos leiteiros em Portugal, usando o modelo de contrastes. O impacto da perda de informação nas avaliações genéticas foi analisado através da variação de alguns parâmetros genéticos da população Holstein nacional.

4º Objectivo – Na fase seguinte estudaram-se os critérios e os efeitos da aplicação e sistematização da utilização de várias técnicas de formação de GC nas avaliações genéticas do efectivo Holstein nacional. Foram avaliadas as consequências da inclusão de novas variáveis descritivas, recorrendo à análise de componentes principais na definição das semelhanças entre explorações. Para um maior controlo do processo de formação de “clusters”, foram também introduzidos novos parâmetros com o objectivo de minimizar o número de explorações por “cluster”, maximizando o número de GC final.

Foi ainda determinado o tamanho mínimo dos GC de modo a minimizar as mudanças na estrutura original da base de dados.

5º Objectivo – Na fase final deste projecto foi estudado o impacto das mudanças de explorações de um “cluster” para o outro em avaliações consecutivas, quer a nível de alterações no ranking genético dos animais como do respectivo rigor de predição no programa de selecção nacional.

3. RESULTADOS

Artigo 1*

Prediction of Daily and Lactation Yields of Milk, Fat and Protein Using an
Autoregressive Repeatability Test Day Model

*Predição das Produções Diárias e de Lactações de Leite, Gordura e Proteína
pelo Modelo de Contrastes Auto-Regressivo*

*Vasconcelos, J., A. Martins, M. F. Petim-Batista, J. Colaço, R. W. Blake, and J. Carnevalheira (2004) Prediction of Daily and Lactation Yields of Milk, Fat, and Protein Using an Autoregressive Repeatability Test Day Model. J. Dairy Sci. 87:2591-2598

Predição das Produções Diárias e de Lactações de Leite, Gordura e Proteína pelo Modelo de Contrastes Auto-Regressivo

RESUMO

Foi avaliada a fiabilidade de um modelo para contrastes auto-regressivo (**MCAR**) de múltiplas lactações para a predição de contrastes em falta nas produções de leite, gordura e proteína de modo a obter estimativas das produções acumuladas aos 305 dias em vacas com lactações incompletas ou ainda em progresso. Foi utilizada a base de dados nacional, composta por mais de um milhão de contrastes, produzidos nas primeiras 3 lactações por cerca de 75.000 vacas Holstein. As diferenças entre lactações padronizadas de 305 dias reais (estimadas a partir de lactações completas pelo método do intervalo entre testes - **MIT**) e estimadas pelo MCAR foram insignificantes e menores do que as estimadas pelo MIT. Verificou-se uma ligeira tendência para a subestimativa das lactações padronizadas aos 305 dias pelo MCAR, enquanto o MIT, pelo contrário, sobrestimou-as largamente. As menores diferenças obtidas pelo MCAR resultaram em estimativas de lactações padronizadas menos enviesadas, logo, com um maior rigor de predição. Como esperado, a correlação entre lactações padronizadas reais e projectadas foi progressivamente maior quanto maior o número de contrastes disponíveis (de 0,831 para 0,997). As correlações médias por lactação variaram entre 0,977 e 0,984. Valores de correlação entre contrastes reais e estimados foram superiores a 0,50, mesmo nos casos de estimativa baseada em contrastes com 7 meses de intervalo. Estes valores de correlação revelam uma elevada capacidade de predição do MCAR. Do ponto de vista dos criadores, pelo facto de a maioria das decisões sobre selecção serem baseadas na capacidade produtiva das vacas, a aplicação do MCAR representa uma valiosa ferramenta informativa para as decisões de manejo e gestão do efectivo, adicionado ao facto de a sua aplicação não requerer grandes recursos informáticos.

Prediction of Daily and Lactation Yields of Milk, Fat, and Protein Using an Autoregressive Repeatability Test Day Model

J. Vasconcelos,¹ A. Martins,² M. F. Petim-Batista,³
J. Colaço,³ R. W. Blake,⁴ and J. Carvalheira^{1,5}

¹CIBIO/ICETA, Centro de Investigação em Biodiversidade e Recursos Genéticos,
Universidade do Porto, Portugal

²ABLN, Associação para o apoio à Bovinicultura Leiteira do Norte, Portugal

³UTAD, Universidade de Trás-os-Montes e Alto Douro, Portugal

⁴Department of Animal Science, Cornell University, Ithaca, NY 14853

⁵ICBAS, Instituto de Ciências Biomédicas Abel Salazar,
Universidade do Porto, Portugal

ABSTRACT

We evaluated the accuracy of an autoregressive multiple-lactation test day (ATD) model to predict missing test day yields of milk, fat, and protein to obtain cumulative 305-d records for cows with incomplete or in-progress lactations. The data consisted of more than one million observations of daily yields on test days in the first 3 lactations of over 75,000 Portuguese Holstein cows. Differences between actual (estimates from complete lactations using the test interval method) and ATD-predicted 305-d yields were negligible and smaller than those predicted by the test interval method. The ATD procedure tended to slightly underestimate cumulative lactation yields, whereas the test interval method substantially overestimated them. Smaller differences obtained by the ATD procedure resulted in less biased estimates of lactation yield, which also implies greater accuracy. As expected, the correlations between actual and predicted lactation yields increased with the number of test days from 0.831 to 0.997. Average correlations (by parity) between actual and ATD-predicted yields ranged from 0.977 to 0.984. Correlations between actual test day yields and corresponding predicted yields exceeded 0.5 for up to 7 time-intervals from the last test day yield used to predict cumulative yield of projected lactations. These correlations indicate the good predictive ability of the ATD method. From a producer's viewpoint, these advantages underwrite management because most on-farm selection decisions are based on the producing abilities of cows. Implementation of ATD methodology does not require special computing capability and is easily transferable to the farm level.

(**Key words:** test day model, autoregression, dairy cattle, prediction of test day record)

Abbreviation key: ATD = autoregressive multiple lactations test day model, LTE = long-term environmental effect, STE = short-term environmental effect, TD = test day, TIM = test interval method.

INTRODUCTION

Daily yield on a test day (TD) is a potent source of information for dairy cattle genetic evaluation and herd management. Nonetheless, there are many situations in which estimates of cumulative 305-d yields are needed, especially for cow culling based on producing ability. Another reason for using estimates of cumulative yield is the limited capacity to process large amounts of TD data (Bormann et al., 2002; Wiggans et al., 2002), which may preclude direct use of TD records in genetic evaluation. For these cases, a classical 2-step approach of first estimating 305-d lactation yields of cows followed by genetic evaluation of sires and daughters is most practical because more individuals are considered (Bormann et al., 2002; Wiggans et al., 2002).

Although genetic evaluation based on TD records is under investigation in Portugal (Carvalheira et al., 2001, 2002), the official method to evaluate the dairy cattle population is based on cumulative 305-d lactation yields of milk and milk components. The Portuguese Dairy Cattle Breeders Association utilizes the test interval method (TIM) to estimate them from TD measurements, following the rules defined by the International Committee for Animal Recording (ICAR, 1995). Criteria for records to be used in genetic evaluation are restrictive. For example, the Portuguese national dairy database (primarily from the Holstein breed) annually accumulates about 90,000 valid 305-d lactation records from approximately 104,000 cows with dry dates. Completed lactations must be at least

Received February 5, 2004.

Accepted March 29, 2004.

Corresponding author: J. Carvalheira; e-mail: jgc3@mail.icav.up.pt.

Table 1. Variance components (kg²) and autocorrelations for daily milk, fat, and protein yields for the Portuguese Holstein population.

Component	Milk	Fat	Protein
Error variance	2.967	0.010	0.003
LTE ¹ variance	7.995	0.015	0.006
LTE autocorrelation	0.133	0.185	0.153
STE ² variance (L1) ³	12.960	0.014	0.009
STE autocorrelation (L1)	0.830	0.687	0.766
STE variance (L2) ³	24.756	0.031	0.018
STE autocorrelation (L2)	0.828	0.639	0.792
STE variance (L3) ³	28.758	0.037	0.022
STE autocorrelation (L3)	0.818	0.628	0.796

¹LTE = Long-term environmental effects.²STE = Short-term environmental effects.³L1 = 1st lactation; L2 = 2nd lactation; L3 = 3rd lactation.

305 d long or in-progress records must have at least 210 DIM to be projected to a 305-d basis (Gama et al., 2000). To illustrate information losses for genetic evaluations, about 23% of all lactations from 1994 to 2000 were ignored because they were too short or still in progress (<210 DIM). Yet more information is lost because multiparous cows without recorded first lactation yields are also ignored. For small national data sets, like the case of Portugal, recovery of lost information is vital to genetic evaluation.

Several reports indicated that TIM might be less accurate than other methods to estimate cumulative 305-d yields (Schaeffer and Jamrozik, 1996; VanRaden, 1997; Norman et al., 1999; Pool and Meuwissen, 1999). Schaeffer and Jamrozik (1996) predicted lactation yields for milk, fat, and protein using a multiple-trait procedure that incorporated standard lactation curves following Wood (1967) equation. This procedure permitted more accurate predictions of 305-d yields compared with the TIM regardless of the number of missing TD measurements. Pool and Meuwissen (1999) found that a 5th-order Legendre polynomial random regression TD model yielded the best fit and more accurate 305-d lactation predictions than TIM.

An alternative approach for estimating cumulative lactation yield is by a TD model with autoregressive structure (Carvalho et al., 1998, 2002; Macciotta et al., 2002; Vasconcelos et al., 2002). Autoregressive multiple-lactation test day (ATD) models may predict cumulative yield within herd and parity, adjusted for herd management and age of cow, by accounting for long-term (correlations between parities) and short-term (correlations between TD within a lactation) environmental effects (LTE and STE, respectively) governing repeated milking performance of a cow. Advantages of this method are that all TD residuals from incomplete or in-progress lactations can be predicted and cumulative yields estimated without need for projection factors by summing the appropriate within

herd means. Thus, using the classic genetic evaluations, recuperating “lost” records would be expected to result in less biased genetic evaluations by permitting culled animals to contribute to their sire’s evaluation, including the possibility for earlier evaluation of young sires (Henderson et al., 1959; Wiggans and Van Vleck, 1979; Norman et al., 1999). Therefore, the objective of this study was to evaluate the accuracy of the ATD model to predict missing TD yields of milk, fat, and protein and to estimate 305-d cumulative yields compared with standard TIM procedures.

MATERIALS AND METHODS

Data

Data provided by the Portuguese Dairy Cattle Breeders Association consisted of TD yields of milk, fat, and protein from Portuguese Holstein cows from June 1994 to December 2000. Monthly records (morning plus afternoon yields on the same day) from cows in the first 3 lactations in supervised recording plans were used. Lactations were required to have at least 2 TD records per cow. Further edits followed ICAR (1995) rules for TD recording. Briefly, TD records corresponded to the range of 5 to 305 DIM, first reported TD could not exceed 75 DIM, and time interval between consecutive TD has to be between 26 and 33 d. Edited files contained 2,101,098, 2,093,524, and 2,103,506 TD yields of milk, fat, and protein, respectively, from which data subsets were extracted for this study.

Model

The ATD model developed and applied by Carvalho et al. (1998, 2002) was adapted to include a random cow effect partitioned into 2 autoregressive processes. One process accounted for correlations among TD within each lactation: STE. Long-term environmental effects were accounted by the correlations between lactations (parities). The TD yields within a lactation were treated as independent random variables with heterogeneous (co)variance components describing STE effects in each lactation. Genetic relationships among cows were assumed to be nil. However, this simplification of our model (Carvalho et al., 1998, 2002) permits confounding of the genetic component with the LTE and STE effects. The ATD “cow” model used to analyze the TD data was as follows:

$$Y_{ijkLmn} = \text{HTD}_i + \text{Age}(\text{H})_j + \text{DIM}(\text{H})_{k(L)} + p_{m(L)} + t_{n(mL)} + e_{ijkLmn},$$

Table 2. Frequency (%) of complete (with 10 test day records) and incomplete or in-progress (2 to 9 test day records) milk lactations (L) obtained from the Portuguese dairy cattle database (from June 1994 to December 2000).¹

Test days	Frequency (%)		
	L1	L2	L3
2	4.0	4.2	4.6
3	4.0	4.6	4.8
4	4.0	4.6	4.7
5	3.5	4.3	4.8
6	3.6	4.4	4.5
7	4.1	5.4	5.6
8	6.9	8.6	8.4
9	25.9	25.0	24.9
10	44.0	38.9	37.7

¹Frequencies for fat and protein were similar.

where y is the TD yield, HTD is the fixed effect of herd-test-date, Age(H) is the fixed effect of age at calving (grouped in 16 classes from 18 to 76 mo of age) nested within herd (H), DIM(H) is the fixed effect of DIM nested with herd and lactation number (L), p is the random effect of LTE accounting for the correlations generated by the cow across lactations, t is the random effect of STE accounting for the correlations due to the cow between TD within a lactation, and e is the random residual effect. The effects of p and t are fitted with first-order autocorrelation structures. Variance components and correlations were obtained using derivative free REML methodology (Smith and Graser, 1986) to estimate 9 parameters (5 variance components and 4 autocorrelations, Table 1).

Data subsets containing only complete lactations (i.e., those with 10 TD measurements) were extracted from the edited database for each trait to evaluate the predictive capacity of the ATD model. These complete datasets contained 1,048,730 (76,633), 1,024,020 (75,252), and 1,048,150 (76,553) TD records (cows) for milk, fat, and protein, respectively. Reduced datasets were created for each trait by randomly deleting TD records to mimic the pattern (frequency) of missing

TD measurements in the national database corresponding to incomplete or in-progress lactations (Table 2).

Cumulative 305-d lactation yields were estimated for both the complete and reduced datasets using ATD and TIM (based on the rules by ICAR [1995] methods). Projected yields by the TIM were for lactations with at least 210 DIM (≥ 7 TD) and were based only on the last TD yield. The ATD method was applied to lactations with at least 2 TD measurements. Goodness of prediction was assessed by comparing cumulative 305-d lactations records obtained by these 2 methods for complete and reduced datasets. Measures of accuracy were correlations, means, and standard deviation of differences between ATD- and TIM-estimated yields.

RESULTS AND DISCUSSION

Complete Lactations with 10 Test Days

Means, standard deviations, and correlations between estimated lactation yields obtained by TIM and ATD for complete lactations with 10 daily yields on TD are given in Table 3. Estimated yields did not differ significantly for any trait ($P > 0.5$). Standard deviations were slightly smaller for the ATD method. Large correlations (≥ 0.995) indicated that these methods have equal predictive value for cumulative yields when daily yields on all TD are available. Hereafter, TIM estimates of lactation yields from complete lactations are the basis for comparisons with yields projected from fewer TD measurements by either TIM or ATD.

Incomplete and In-Progress Lactations

Means and standard deviations of the difference between complete and projected lactation yields in each parity and correlations between them are given in Table 4 for milk, fat, and protein. Differences between complete and projected lactation yields by the ATD

Table 3. Means and standard deviations¹ of 305-d lactation (L) records of milk, fat, and protein (kg) obtained from the complete dataset (all 10 test day records available) using the test interval method (TIM) and autoregressive repeatability test day model (ATD), and their correlations.

	Milk			Fat			Protein		
	1st L	2nd L	3rd L	1st L	2nd L	3rd L	1st L	2nd L	3rd L
TIM	7071.7 (1426.8)	7825.0 (1669.3)	7985.7 (1641.6)	252.6 (55.2)	277.3 (63.7)	283.7 (63.5)	223.5 (45.8)	248 (52.1)	250.2 (51.1)
ATD	7057.9 (1401.7)	7804.9 (1639.7)	7965.4 (1613.1)	252.5 (53.8)	277.4 (62.1)	283.6 (61.8)	223.2 (45.1)	248 (51.4)	250.4 (50.4)
Correlation	0.998	0.997	0.995	0.998	0.997	0.995	0.998	0.997	0.995

¹Standard deviations are in parenthesis.

Table 4. Means and SD of differences between actual¹ and predicted 305-d milk, fat, and protein yields (kg), and correlations between actual and predicted lactation yields by autoregressive repeatability test day model (ATD) and test interval method (TIM).

Tests ²	Milk						Fat						Protein					
	Mean		SD		Correlation		Mean		SD		Correlation		Mean		SD		Correlation	
	ATD	TIM	ATD	TIM	ATD	TIM	ATD	TIM	ATD	TIM	ATD	TIM	ATD	TIM	ATD	TIM	ATD	TIM
Lactation 1																		
2	-4.5	—	702.24	—	0.860	—	0.27	—	29.41	—	0.856	—	0.31	—	23.45	—	0.861	—
3	-21.7	—	583.86	—	0.913	—	0.82	—	24.34	—	0.898	—	-0.47	—	19.46	—	0.907	—
4	-2.5	—	485.03	—	0.940	—	0.97	—	20.47	—	0.928	—	0.13	—	16.24	—	0.934	—
5	-27.5	—	399.13	—	0.957	—	0.05	—	17.27	—	0.949	—	0.15	—	13.04	—	0.957	—
6	-1.9	—	318.58	—	0.976	—	-0.09	—	14.00	—	0.967	—	-0.52	—	10.39	—	0.974	—
7	-22.8	171.0	236.98	267.29	0.987	0.985	0.04	3.13	10.44	12.92	0.982	0.975	-0.10	3.69	8.30	9.78	0.983	0.978
8	-18.4	90.5	172.03	175.71	0.993	0.993	-0.09	1.90	8.06	9.22	0.990	0.987	-0.19	2.31	5.74	6.18	0.992	0.991
9	-13.4	37.3	115.14	93.39	0.997	0.998	-0.01	0.82	5.21	4.64	0.996	0.997	-0.12	0.98	3.75	3.29	0.997	0.997
Lactation 2																		
2	-3.6	—	799.97	—	0.877	—	-0.22	—	3.45	—	0.848	—	-0.50	—	26.67	—	0.858	—
3	4.4	—	688.23	—	0.912	—	1.04	—	28.81	—	0.887	—	1.28	—	21.77	—	0.909	—
4	-0.6	—	569.28	—	0.942	—	0.74	—	24.07	—	0.925	—	-0.03	—	18.62	—	0.929	—
5	-14.4	—	453.76	—	0.961	—	0.54	—	20.70	—	0.945	—	0.58	—	15.89	—	0.953	—
6	-20.6	—	376.88	—	0.975	—	0.88	—	16.43	—	0.966	—	0.39	—	12.61	—	0.973	—
7	-17.6	337.7	287.93	327.40	0.984	0.983	0.38	8.80	13.01	15.81	0.979	0.973	0.08	8.55	9.61	10.85	0.983	0.980
8	-20.0	175.3	213.75	204.86	0.992	0.993	0.14	4.66	9.40	9.94	0.989	0.989	0.01	4.60	7.00	7.10	0.991	0.991
9	-18.6	61.9	148.43	102.78	0.996	0.998	0.05	1.76	6.20	4.80	0.995	0.997	-0.01	1.70	4.58	3.54	0.996	0.998
Lactation 3																		
2	-27.2	—	822.83	—	0.858	—	0.35	—	33.24	—	0.841	—	-0.43	—	27.25	—	0.831	—
3	-9.6	—	684.57	—	0.909	—	-0.69	—	30.15	—	0.889	—	-0.22	—	23.36	—	0.892	—
4	-74.9	—	576.35	—	0.937	—	1.68	—	25.31	—	0.919	—	-0.55	—	19.96	—	0.920	—
5	-1.7	—	478.00	—	0.958	—	0.46	—	21.21	—	0.945	—	1.65	—	16.05	—	0.945	—
6	-25.1	—	373.01	—	0.973	—	-0.54	—	16.42	—	0.963	—	-0.23	—	12.90	—	0.969	—
7	-14.0	365.9	299.43	340.14	0.983	0.982	0.20	10.14	13.67	16.12	0.976	0.971	0.23	9.56	9.62	10.96	0.982	0.980
8	-19.9	192.8	221.77	217.40	0.991	0.992	-0.47	5.11	9.87	9.89	0.988	0.988	0.13	5.26	7.18	7.08	0.990	0.991
9	-14.9	65.9	150.77	99.02	0.996	0.998	0.04	1.95	6.79	4.85	0.995	0.997	0.14	1.77	4.89	3.50	0.996	0.998

¹Actual 305-d milk yields corresponds to TIM estimates with all 10 test day records.²Number of test days used for prediction of lactation yields.

Table 5. Correlations between observed and ATD¹-predicted test days (TD) for milk, fat, and protein yields in first lactation.

TD	Number of TD used for prediction of lactation yields							
	2	3	4	5	6	7	8	9
Milk								
3	0.827	—	—	—	—	—	—	—
4	0.771	0.834	—	—	—	—	—	—
5	0.744	0.815	0.873	—	—	—	—	—
6	0.724	0.775	0.809	0.846	—	—	—	—
7	0.683	0.759	0.782	0.802	0.865	—	—	—
8	0.678	0.719	0.742	0.767	0.824	0.871	—	—
9	0.633	0.696	0.726	0.737	0.782	0.811	0.870	—
10	0.561	0.645	0.650	0.657	0.723	0.742	0.799	0.841
Fat								
3	0.721	—	—	—	—	—	—	—
4	0.677	0.760	—	—	—	—	—	—
5	0.674	0.724	0.761	—	—	—	—	—
6	0.661	0.690	0.736	0.765	—	—	—	—
7	0.638	0.666	0.713	0.736	0.782	—	—	—
8	0.620	0.636	0.654	0.697	0.731	0.792	—	—
9	0.589	0.616	0.644	0.654	0.695	0.744	0.779	—
10	0.551	0.579	0.605	0.632	0.634	0.673	0.695	0.775
Protein								
3	0.801	—	—	—	—	—	—	—
4	0.764	0.835	—	—	—	—	—	—
5	0.738	0.802	0.831	—	—	—	—	—
6	0.710	0.753	0.794	0.848	—	—	—	—
7	0.698	0.733	0.760	0.800	0.848	—	—	—
8	0.675	0.717	0.730	0.776	0.816	0.850	—	—
9	0.626	0.689	0.698	0.722	0.773	0.798	0.848	—
10	0.590	0.624	0.662	0.690	0.718	0.726	0.779	0.832

¹Autoregressive repeatability test day model.

method were negligible and much smaller for all traits than those obtained using the TIM, which agrees with other reports (Schaeffer and Jamrozik, 1996; Norman et al., 1999; Macciotta et al., 2002). The available number of TD did not produce a discernible pattern in the mean differences. However, predictions by the TIM resulted in large differences in means and standard deviations of differences of actual complete and predicted 305-d yields with 7 TD measurements. These differences decreased with additional TD observations. Generally, the ATD method slightly underestimated (underprojected) lactation yields, especially for milk, whereas the TIM substantially overestimated them. In absolute terms, the smaller differences obtained with the ATD procedure are expected to result in less biased predictions of lactation yield, which signifies greater accuracy.

The standard deviations of the differences decreased with the available number of TD in all traits for both procedures, similar to other findings (Schaeffer and Jamrozik, 1996; Norman et al., 1999). The ATD method always obtained the smallest standard deviation with 7 TD measurements. Similar to the estimates obtained by best prediction procedures (VanRaden, 1997; Norman et al., 1999), lactation yields predicted

by the ATD method had relatively smaller standard deviation because predicted TD were regressed toward the parity means within herds. A companion study of variance components estimated from ATD-projected lactation records (141,681 cows, sired by 8,189 bulls, unpublished data) resulted in heritabilities that were 5 to 7 percentage points larger for yields of milk, fat, and protein than TIM-projected records, agreeing with other reports (Swalve, 2000; Jensen, 2001; Bormann et al., 2002; Wiggins et al., 2002).

The correlations between yields from complete and predicted lactations increased from 0.831 to 0.997 with the available number of TD measurements. These coefficients were similar to the ones found by Macciotta et al. (2002), but are slightly larger than those reported by Schaeffer and Jamrozik (1996), especially for yields predicted with 2, 3, or 4 TD. Similar to previous findings (e.g., Norman et al., 1999), the average correlations by parity between complete lactation yields and ATD-predicted yields were 0.977 to 0.984 for all traits.

Correlations between actual daily yields on TD and ATD-predicted daily yields in first parity for milk and fat are given in Table 5. As expected, the largest correlations were found nearest to TD and decayed with the time interval to more distant TD toward the end

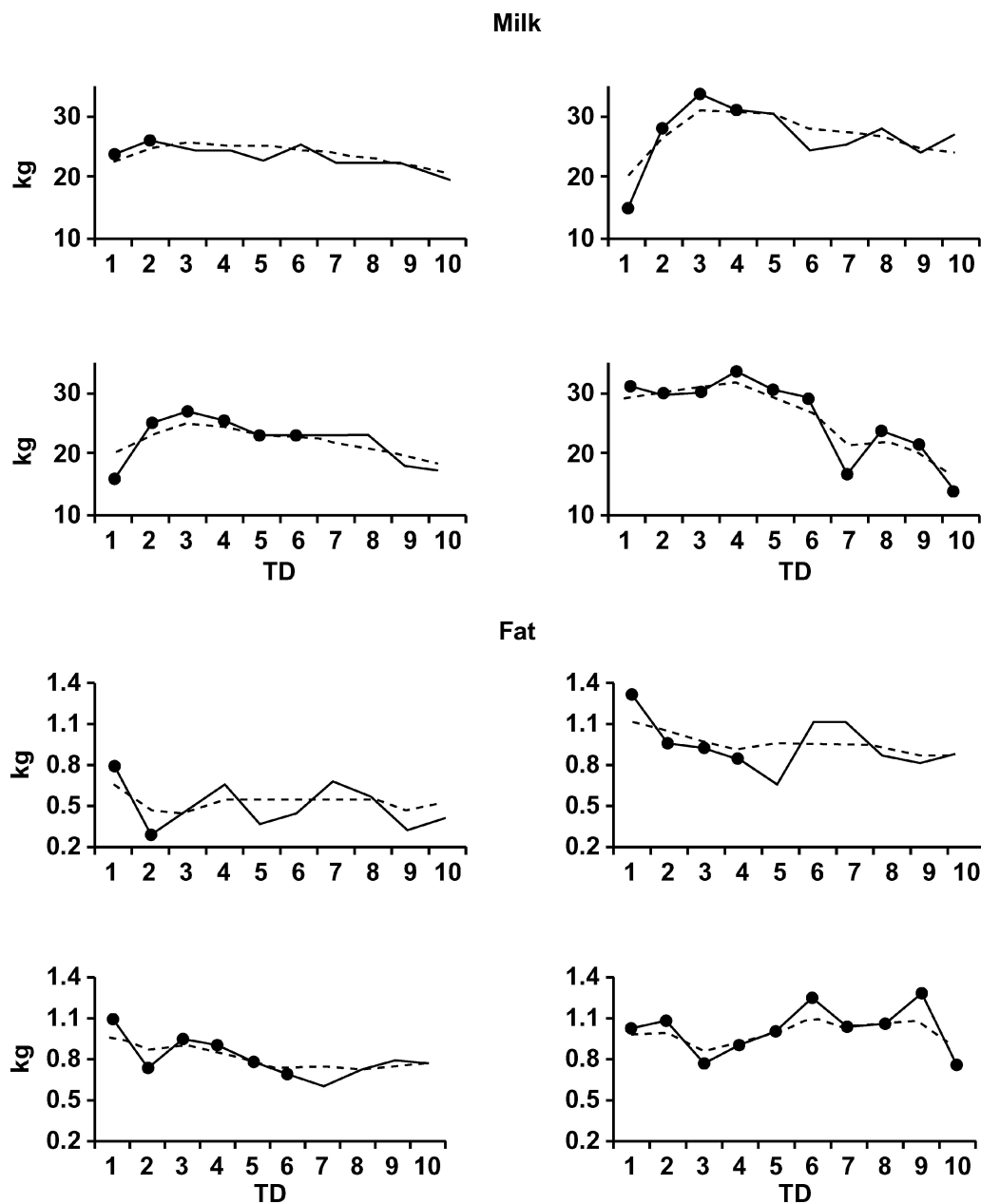


Figure 1. First lactation curves for milk and fat yields. Actual (—) curve based on all 10 test day (TD) records, predicted (-----) curve by autoregressive repeatability TD model based on 2, 4, 6, and 10 (•) observed TD.

of lactation. Similar results were obtained for the 2nd and 3rd parities for all traits. Average correlations within parity for the first predicted TD yields were 0.84 to 0.86 for milk, 0.73 to 0.77 for fat, and 0.82 to 0.84 for protein, which were larger than correlations reported by Macciotta et al. (2002). Correlations between actual daily yields and predicted TD yields exceeded 0.5 up to 7 time intervals from the last actual

TD used in the prediction of the extended lactation. These relatively large correlations are an indirect measure of the predictive power of the ATD method.

Farmers need reliable information for management decisions, especially for cow culling. Shapes of lactation curves (e.g., presence or absence of detectable ascent to peak daily milk yield; Raffrenato et al., 2003) and information about lactation persistency of individ-

ual cows are valuable management inputs. Besides cumulative 305-d lactation yields, accurate prediction of missing and expected TD yields permits graphical depiction of lactation curves for individuals and groups of cows, which are illustrated in Figure 1 for milk and fat in first lactation. These curves are from incomplete lactations that were projected based on varying numbers of available TD measurements. As VanRaden (1997) pointed out, the TIM is intended for simple interpolation, which may lead to poor prediction of distant TD yields. On the other hand, ATD-predicted lactation curves are adaptable to a wide range of data patterns. Regression of ATD-predicted TD yields toward the parity means within herd at each DIM helps ensure a good fit, even with only 2 TD available.

CONCLUSIONS

Dairy cattle evaluation systems based on 305-d cumulative lactation yields may benefit from TD methodology to predict complete lactations. This is especially important for situations similar to those in Portugal where incomplete and in-progress lactations are discarded from evaluations.

Differences between actual and predicted lactation yields by ATD were negligible and much smaller than those predicted by TIM for all 3 traits. In absolute terms, the smaller differences obtained from the ATD procedure resulted in less biased prediction, which also implies greater accuracy with this method. For these Portuguese data, it will be possible to recuperate all records from incomplete and in-progress lactations (between 23 and 30%) that the traditional TIM procedure would have discarded for a genetic evaluation. Furthermore, genetic analysis performed on ATD-predicted 305-d lactation yields using the Portuguese national dataset resulted in heritabilities that were 5 to 7 percentage points larger than TIM-predicted records for all traits. The possibility of double counting when using ATD-predicted 305-d lactation records for genetic evaluations with the standard repeatability animal model was empirically tested by comparing the pattern and magnitude of the correlations between all combinations of real and ATD-predicted lactation records. No significant differences were found. However, additional model checking is required to assure this outcome. Our next step will be to use the animal model version of the ATD to fully describe the genetic component (Carvalho et al., 2002). It is expected that this procedure will further increase the accuracy of predictions. The implementation of the ATD methodology does not require special computer power and would be easily transferred to the farm level.

ACKNOWLEDGMENTS

The authors thank A. Silvestre from Trás-os-Montes e Alto Douro University and A. Ferreira from the Portuguese Dairy Cattle Breeders Association for their assistance. This work contributes to project POCTI/33162/CVT/2000 and was partially funded by the Fundação para a Ciência e a Tecnologia (FCT) and FEDER (EU).

REFERENCES

- Bormann, J., G. R. Wiggans, T. Druet, and N. Genglar. 2002. Estimating effects of permanent environment, lactation stage, age, and pregnancy on test-day yield. Available: <http://www.adsa.org/jds/abs/2002/d0210263.htm>. Accessed Dec. 2003.
- Carvalho, J., R. W. Blake, E. J. Pollak, R. L. Quaas, and C. V. Duran-Castro. 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for dairy milk yield in a tropical herd of Lucerna cattle and in US Holstein herds. *J. Dairy Sci.* 81:2738–2751.
- Carvalho, J., E. J. Pollak, R. L. Quaas, and R. W. Blake. 2001. Fitting multiple-lactation test-day records with an autoregressive repeatability animal model. Page 6 in *Proc. 52nd Annu. Meeting of the Eur. Assoc. for Anim. Prod.*, Budapest, Hungary.
- Carvalho, J., E. J. Pollak, R. L. Quaas, and R. W. Blake. 2002. An autoregressive repeatability animal model for test-day records in multiple lactations. *J. Dairy Sci.* 85:2040–2045.
- Gama, L. T., I. Carolino, and N. Carolino. 2000. Correlações genéticas e fenotípicas entre a produção de leite e gordura em bovinos da raça Frísia. Page 104 in *Proc. Congresso de Zootecnia, Estação Zootécnica Nacional, Vale de Santarém*.
- Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15:192–218.
- ICAR. 1995. Recording Guidelines. Appendices to the information agreement of recording practices (milk recording). Int. Committee for Anim. Rec., Rome, Italy.
- Jensen, J. 2001. Genetic evaluation of dairy cattle using test-day models. *J. Dairy Sci.* 84:2803–2812.
- Macciotta, N. P. P., D. Vicario, G. Pulina, and A. Cappio-Borlino. 2002. Test day and lactation yield predictions in Italian Simmental cows by ARMA methods. *J. Dairy Sci.* 85:3107–3114.
- Norman, H. D., P. M. VanRaden, J. R. Wright, and J. S. Clay. 1999. Comparison of test interval and best prediction methods for estimation of lactation yield from monthly, a.m.-p.m., and tri-monthly testing. *J. Dairy Sci.* 82:438–444.
- Pool, M. H., and T. Meuwissen. 1999. Prediction of daily milk yields from a limited number of test days using test day model. *J. Dairy Sci.* 82:1555–1564.
- Raffrenato, E., R. W. Blake, P. A. Oltenacu, J. Carvalho, and G. Licitra. 2003. Genotype by environment interaction for yield and somatic cell score with alternative environmental definitions. *J. Dairy Sci.* 86:2470–2479.
- Schaeffer, L. R., and J. Jamrozik. 1996. Multiple-trait prediction of lactation yields for dairy cows. *J. Dairy Sci.* 79:2044–2055.
- Smith, S. P., and H. U. Graser. 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.* 69:1156–1165.
- Swalve, H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. *J. Dairy Sci.* 83:1115–1124.
- VanRaden, P. M. 1997. Lactation yields and accuracies computed from test day yields and (co)variances by best prediction. *J. Dairy Sci.* 80:3015–3022.

- Vasconcelos, J., A. Martins, M. F. Petim-Batista, A. Silvestre, A. Ferreira, J. Colaço, and J. Carvalheira. 2002. Estimation of lactation yields from test day yields of milk, fat, and protein for Portuguese Holstein cattle using an autoregressive test day model. Communication no. 01-57 in Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. CD-ROM.
- Wiggans, G. R., and L. D. Van Vleck. 1979. Extending partial lactation milk and fat records with a function of last-sample production. *J. Dairy Sci.* 62:316–325.
- Wiggans, G. R., P. M. VanRaden, J. Bormann, J. C. Philpot, T. Druet, and N. Gengler. 2002. Deriving lactation yields from test-day yields adjusted for lactation stage, age, pregnancy, and herd test date. <http://www.adsa.org/jds/abs/2002/d0210264.htm>. Accessed Dec. 2003.
- Wood, P. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216:164–165.

Artigo 2*

Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro
em Portugal

Inbreeding and Inbreeding Depression in Portuguese Dairy Cattle

*Vasconcelos, J., A. Martins, A. Ferreira, S. Pinto, e J. Carvalheira (2005) Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal. Revista Portuguesa de Ciências Veterinárias 100(553-554): 33-38

Níveis de endogamia e depressão de endogamia no gado bovino leiteiro em Portugal

Inbreeding and inbreeding depression in portuguese dairy cattle

J. Vasconcelos^{1,2}, A. Martins³, A. Ferreira³, S. Pinto⁴ e J. Carvalheira^{1,2,*}

¹Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-ICETA) - Universidade do Porto, Rua Padre Armando Quintas-Crasto, 4485-661 Vairão, Portugal

²Instituto de Ciências Biomédicas de Abel Salazar (ICBAS) - Universidade do Porto, Largo Prof. Abel Salazar, n.º 2, 4099-003 Porto, Portugal

³Associação para o Apoio à Bovinicultura Leiteira do Norte (ABLN), Av. D. António Bento Martins Júnior, 86, 4480-461 Vila do Conde, Portugal

⁴Associação Portuguesa dos Criadores da Raça Frísia (APCRF), Av. Egas Moniz, n.º 14-2º, 2135-232 Samora Correia, Portugal

Resumo: O ficheiro do Pedigree Nacional de bovinos da raça Holstein, com 1.197.384 registos, foi analisado para avaliar a ocorrência de endogamia. Cerca de 40% da população tem pelo menos um progenitor desconhecido. A proporção de animais endogâmicos foi de 8,6%, com um coeficiente de endogamia (CE) médio de 1,8%. O CE médio na população em geral tem vindo a crescer a uma taxa de 0,03% ao ano. A moda da distribuição do número máximo de gerações conhecidas por indivíduo foi de apenas 3. O número de Fundadores efectivos (NFe) da população actual de fêmeas foi de 276,5. Verificou-se uma diminuição relativa de -18% no NFe entre as 2 últimas gerações de fêmeas do efectivo nacional, indicando que houve uma importante perda na variabilidade genética na população. Os 10 touros geneticamente mais influentes contribuíram com 21% para o pool de genes da população actual. O efeito de endogamia foi analisado em 141.134 primeiras lactações. Houve uma diminuição de $-6,6 \pm 1,5$ kg, $-0,17 \pm 0,052$ kg e $-0,18 \pm 0,042$ kg de leite, gordura e proteína, respectivamente, por cada 1% de aumento no CE. Apesar deste efeito não parecer causar reduções importantes nestes caracteres, o aumento intencional da endogamia deve ser evitado, pois os efeitos negativos no animal são globais e cumulativos. É urgente incluir a identificação dos touros disponíveis nos programas informáticos de emparelhamento de modo a prevenir os níveis de endogamia e o valor genético da descendência, instrumento fundamental de apoio aos programas de selecção e melhoramento.

Palavras chave: Endogamia, Depressão de endogamia, Bovinos leiteiros, Holstein

Summary: The Holstein national pedigree file with 1,197,384 animals was investigated for the occurrence of inbreeding. Forty percent of the population in this pedigree file had at least, one missing parent. The proportion of inbred animals was 8.6% with an average inbreeding coefficient (CE) of 1.8%. The number of inbred animals increased at a rate of 2.3% per year. Average CE on the general population was also increasing (0.03% per year). The mode for the distribution of pedigrees according to the maximum number of known generations per individual was only 3. The effective number of Founders (NFe) for the present female population was 276.5. There was a relative decrease of -18% of NFe between the last 2 generations of females, indicating that there was a dramatic change in the genetic variability of this population. The 10 genetically, most influential sires contrib-

uted with 21% for the gene pool of the present cow population. First lactation milk, fat and protein records for 141,134 cows were analysed for inbreeding depression. For each 1% increase in CE, there was a loss of -6.6 ± 1.5 kg, -0.17 ± 0.052 kg and -0.18 ± 0.042 kg of milk, fat and protein, respectively. Inbreeding depression was not enough to cause large reductions in the expression of these traits. Nevertheless, intentional inbreeding should be avoided because of expected global negative impact in animals. It is urgent that computerized mate selection programs include genetic values and extensive pedigree information in order to optimise selection of potential mates.

Key words: Inbreeding, Inbreeding depression, Dairy cattle, Holstein

Introdução

O aumento da homozigotia devido a ascendentes comuns nascidos desde o ano base (endogamia), é medido pelo Coeficiente de Endogamia (CE). A sua importância nas espécies domésticas está relacionada com a probabilidade acrescida de uma diminuição na performance de caracteres com importância económica, e.g., viabilidade e produção. Este efeito, conhecido por Depressão de Endogamia (DE), é o resultado duma diminuição na heterozigotia e de um aumento na frequência de genes deletérios recessivos na população. Outro aspecto importante resultante do aumento da homozigotia é a perda de variabilidade genética e consequente diminuição das oportunidades para selecção (Swalve et al., 2003).

A endogamia está a tornar-se num problema crescente nos programas de melhoramento de bovinos leiteiros. Os programas de selecção genética permitem identificar indivíduos e famílias com um potencial geneticamente superior. As actuais tecnologias de reprodução são bastantes eficientes na passagem desses genótipos para a população (Weigel e Lin, 2000), os quais tendem a concentrar-se em poucos indivíduos (Smith et al., 1998). A população Holstein, apesar de grande em números absolutos, é relativamente peque-

*correspondência: e-mail: jgc3@mail.icav.up.pt ;
Tel. 252 660400; Fax: 252 661780

na em tamanho efectivo (Swalve et al., 2003). A maior parte da endogamia tem origem na utilização intensiva de touros de Inseminação Artificial onde a intensidade de selecção é particularmente elevada (Swalve et al., 2003). Uma das consequências deste processo é fazer com que um número relativamente pequeno de touros, a maior parte com um forte relacionamento familiar, seja responsável pela manutenção de quase todo o efectivo (Miglior et al., 1992).

Em vacas, a endogamia diminui a sobrevivência, a produção de leite, o desempenho reprodutivo e pode ser responsável pelo aumento do número de células somáticas no leite (Smith et al., 1998). No que diz respeito aos caracteres de tipo, a literatura refere um pequeno efeito não significativo (Smith et al., 1998). O interesse e preocupação em relação à endogamia continua a crescer, com os cientistas a procurar formas de evitar as perdas económicas associadas à DE e a um progresso genético potencialmente mais lento (VanRaden e Smith, 1999). Para maximizar o progresso genético e manter a diversidade genética, é necessário identificar os touros que tenham o menor relacionamento possível com as vacas ou novilhas a serem inseminadas (VanRaden e Smith, 1999). Os objectivos deste estudo foram determinar os níveis de endogamia na população dos bovinos leiteiros em Portugal, analisar o impacto que os touros geneticamente mais influentes têm no pool genético desta população e determinar o efeito depressivo da endogamia na expressão dos principais caracteres produtivos da espécie: leite, gordura e proteína.

Material e métodos

O ficheiro do Pedigree Nacional da raça Holstein, com 258.965 machos e 938.419 fêmeas (1.197.384 registados até Setembro de 2003, inclusive), foi analisado para a ocorrência de endogamia. Para o cálculo dos CE foi utilizado o algoritmo desenvolvido por Meuwissen e Luo (1992) e foram incluídas todas as relações conhecidas entre os indivíduos cadastrados relativamente a uma população base, definida em 1960. Assumiu-se que os progenitores desta população base não eram relacionados nem endogâmicos. As médias dos CE (gerais, por ano de nascimento e sexo) foram também calculadas com o objectivo de analisar tendências e possíveis repercussões na população. Na análise genealógica foi utilizado um pacote de programas informáticos (PEDKIT) desenvolvido por Boichard et al. (1997).

Para determinar o efeito da DE na produção de leite, gordura e proteína, foram utilizadas as lactações ajustadas aos 305 dias de 141.134 primíparas, paridas entre 1994 e 2002. Estudos preliminares demonstraram uma marcada sazonalidade na produção ao longo do ano, existindo claramente 2 estações, de Março a Agosto e de Setembro a Fevereiro. Vários estudos (e.g., Miglior et al., 1994; Smith et al., 1998) concluíram que não há grandes diferenças entre os valores da DE estima-

dos por modelos fixos ou modelos animais. Tendo em conta os objectivos deste estudo (avaliar tendências ao longo do tempo), optou-se pela utilização de um modelo fixo cujos efeitos principais incluíram o rebanho/ano/estação do parto, a idade ao parto (categorizada em 8 classes) e o efeito linear dos CE como covariável:

$$y_{ijk} = RAE_i + I_j + b(CE_k) + e_{ijk},$$

onde y_{ijk} é o vector das produções (leite, gordura ou proteína), RAE_i é o efeito hierarquizado do rebanho/ano/estação de parto, I_j é a idade ao parto, b é o coeficiente de regressão linear, CE_k é o vector com os coeficiente de endogamia para cada animal e e_{ijk} representa o vector de resíduos que se assume ter uma distribuição $N(0, \sigma^2)$.

Resultados e discussão

Endogamia

A proporção de animais endogâmicos foi de 8,6% (102.547 animais dos quais 33.938 eram machos e 68.609 fêmeas) com um CE médio de 1,8%, variando entre 0,0% e 37,5% (Figura 1). Comparativamente com a Alemanha, onde a proporção de animais endogâmicos foi de 80,3% (nascidos entre 1994 e 1997) com um CE médio de 1,93% (Swalve et al., 2003), a proporção de animais endogâmicos neste estudo foi relativamente pequena (a grande maioria tem um CE inferior a 5%), mas a tendência verificada foi de um crescimento contínuo desde 1981 (1,3% por ano). Se considerarmos apenas os últimos 10 anos, este crescimento aumenta

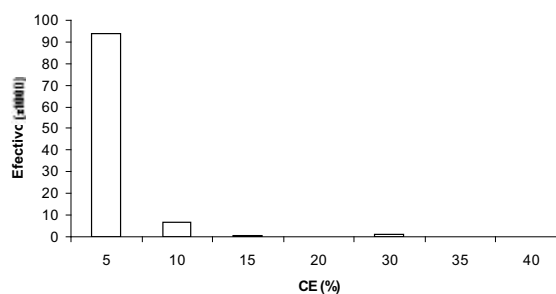


Figura 1 – Distribuição dos Coeficientes de Endogamia (CE) na população Holstein portuguesa.

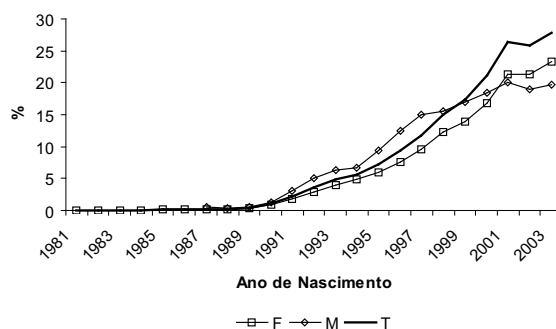


Figura 2 – Percentagem de animais endogâmicos por ano de nascimento na população Holstein portuguesa (F = fêmeas, M = machos e T = total).

para 2,3% ao ano, valor preocupante em certas condições (Figura 2). A tendência do CE médio na população em geral (entre 1991 e 2002) tem sido também de crescimento contínuo a uma taxa de 0,03% ao ano (0,02% em machos e 0,03% em fêmeas). A média dos CE dos animais nascidos e registados em 2002 foi relativamente baixa (1,4%), com uma pequena variação entre sexos (1,4% em fêmeas e 1,3% em machos). Em geral, os valores aqui apresentados foram inferiores aos obtidos em outros estudos, quer Americanos (Miglior et al., 1992; Wiggans e VanRaden, 1995; Young e Seykora, 1996), quer Europeus (Boichard et al., 1997; Swalve et al., 2003).

No que diz respeito aos animais endogâmicos, apesar de em número estarem a aumentar rapidamente, a média dos seus CE baixou muito entre 1985 e 1991 (-3,1% por ano), tendo praticamente estabilizado após esse período (-0,2% ao ano, Figura 3). Esta aparente contradição entre as tendências verificadas nos últimos 10 anos (1,34% de aumento anual de animais endogâmicos e 0,2% de diminuição anual na média dos seus CE), estará muito provavelmente, relacionada com problemas na estrutura do ficheiro de pedigree: o pedigree nacional encontra-se bastante incompleto; 40,1% da população tem pelo menos um progenitor desconhecido (27,5% com os dois e 12,6% com um), implicando que os CE calculados para os indivíduos desta população estão subestimados. Por outro lado, com menor impacto, poderá ser que a utilização mais ou menos generalizada dos programas de emparelhamento que, apesar de não evitarem totalmente o acasalamento entre animais relacionados sem afectar o valor genético da descendência, podem mesmo assim ter minimizado

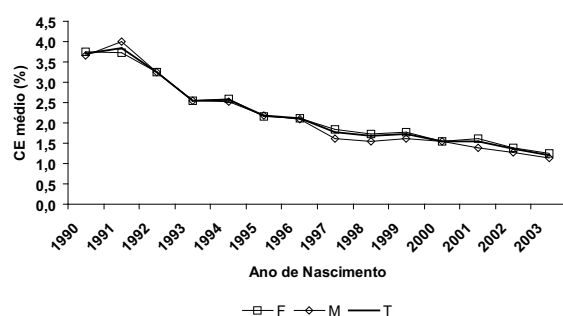


Figura 3 – Evolução do Coeficiente de Endogamia (CE) médio em animais endogâmicos por ano de nascimento na população Holstein portuguesa (F = fêmeas, M = machos e T = total).

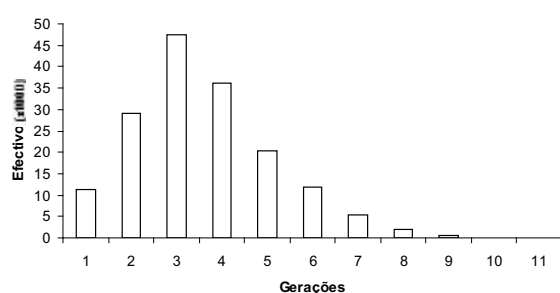


Figura 4 – Distribuição do número máximo de gerações existentes em fêmeas nascidas entre 1999 e 2003 (164.130 animais).

o incremento do CE médio.

A Figura 4 mostra a distribuição de pedigrees de acordo com o número máximo de gerações que foram encontradas em 164.130 fêmeas nascidas entre 1999 e 2003, inclusive. A moda foi de apenas 3 gerações conhecidas sendo mais um indicador de um pedigree incompleto, com uma fraca estrutura familiar. Este é um dado importante na análise da estrutura genealógica desta população. Na Alemanha a moda foi de 8 gerações conhecidas e mesmo assim referiram a existência de uma alta frequência de progenitores desconhecidos no seu pedigree (Swalve et al., 2003). Por outro lado, um facto positivo que se pôde constatar, relaciona-se com a forma consistente com que o número médio de ancestrais conhecidos por indivíduo e ano de nascimento tem vindo a crescer. O mesmo acontece à tendência do número equivalente de gerações conhecidas por indivíduo e ano de nascimento. Estes valores atingiram em 2002 os 25,8 ancestrais conhecidos e 2,75 gerações conhecidas por indivíduo.

Outro aspecto importante na análise genealógica de uma população diz respeito ao estudo da probabilidade de perda de genes entre gerações (quanto maior for a intensidade de selecção, maior será a homozigotia) e consequente diminuição da variabilidade genética. A Tabela 1 contém alguns parâmetros genealógicos da população Holstein portuguesa que podem ajudar na sua caracterização. O conhecimento do tamanho efectivo da população (N_e) ao longo do tempo permite determinar as alterações na variabilidade genética, pois é o reflexo do número e da escolha de reprodutores em gerações anteriores. No entanto, o N_e é um parâmetro extremamente sensível à qualidade do ficheiro de pedigree e faz mais sentido em análises de longo prazo. Geralmente, o N_e é calculado como função da taxa de incremento da endogamia por geração. Com base nessa fórmula, o N_e da população Holstein Portuguesa nos últimos 10 anos corresponde a 1.798,6 reprodutores efectivos. A diferença entre estes e o número de Fundadores (Tabela 1) reflecte claramente os problemas estruturais existentes no pedigree. Em pedigrees com

Tabela 1 – Parâmetros genealógicos da população Holstein portuguesa.

N.º de animais no Pedigree	1.197.384
N.º animais sem descendência	753.851
N.º de Fundadores ^a	93.765
N.º de Fundadores efectivos ^b	276,5
N.º de animais na população de referência ^c	141.355
N.º médio de gerações conhecidas ^c	2,6
N.º médio de ancestrais por indivíduo ^c	22,4
N.º de Ancestrais efectivos ^d	[136,1 ; 162,0]

^aAnimais com descendência mas sem pais conhecidos.

^bDefinido no texto.

^cFêmeas nascidas entre 1999 e 2003 inclusive, com ambos pais conhecidos.

^dLimites mínimo e máximo estimados com base nos 50 ancestrais geneticamente mais influentes na população de referência (ver texto).

uma taxa tão elevada de pais desconhecidos, não faz sentido usar estes parâmetros na sua análise.

Por outro lado, Boichard *et al.* (1997) demonstrou que parâmetros derivados da probabilidade de origem dos genes na população, como o número de Fundadores efectivos (NFe) ou o número de Ancestrais efectivos (NAe), são mais úteis para descrever a estrutura de uma população, especialmente em análises com poucas gerações de intervalo. Estes parâmetros são menos sensíveis aos problemas gerados por pedigrees incompletos, pois tomam em consideração o efeito de selecção (i.e., a probabilidade dos indivíduos serem ou não reprodutores) e a variação no tamanho das famílias. A principal diferença entre o NFe e o NAe reside no facto de o NAe ajustar também para afunilamentos no pedigree i.e., quando poucas famílias contribuem para a maior parte da diversidade genética. Tal pode suceder porque os ancestrais (fundadores ou não) no NAe são escolhidos na base do valor esperado da sua contribuição para o pool genético dos indivíduos que constituem a população de referência (e.g., animais nascidos num determinado período). Para este estudo definiu-se como população de referência o grupo de fêmeas nascidas entre 1999 e 2003 inclusive, com ambos os progenitores identificados. A grande desvantagem do NAe consiste na dificuldade computacional para o estimar. Normalmente opta-se por calcular os seus limites (superior e inferior) com base num número predeterminado de ancestrais (os mais influentes na população de referência).

O número relativamente pequeno de NFe (Tabela 1) reflectiu o uso intensivo da inseminação artificial e mostra que o número efectivo de machos fundadores foi mais determinante do que o número de fêmeas. A quociente entre NFe/NAe (valor médio) é aproximadamente 2, indicando provavelmente que, devido à importação de fêmeas em grande número, existem muitos grupos familiares a contribuir geneticamente, resultando em poucos afunilamentos no pedigree. O aumento do CE médio entre gerações (diminuição do Ne) dá uma indicação da diminuição da variabilidade

genética (perda de genes devido à homozigotia), mas a sua fiabilidade depende do nível de organização dos pedigrees dessas populações. Boichard *et al.* (1997) propõe que se avalie a perda de genes comparando a diminuição percentual relativa do NFe ou do NAe entre gerações. Neste estudo, definiram-se dois grupos de animais, um nascido entre 1994 e 1998 inclusive, e o outro nascido entre 1999 e 2003, constituindo assim grupos geracionais sucessivos. O último grupo apresenta uma diminuição relativa de -18% de NFe, o que significa uma importante perda na variabilidade genética da população. Esta perda pode ser explicada na sua grande parte devido à deriva genética e à alta intensidade de selecção, principalmente de touros (reduzido NFe de machos).

A Tabela 2 mostra os 10 touros geneticamente mais influentes na população de referência. Estes 10 touros contribuíram com cerca de 21% para o pool de genes da população actual das vacas Holstein portuguesas. O aumento da endogamia foi apenas um aspecto da utilização intensiva destes touros. Outra questão, também importante, será o aumento significativo de animais relacionados na população, pois diminui a possibilidade de encontrar emparelhamentos onde a descendência não seja endogâmica. Actualmente, o uso de programas informáticos que, simultaneamente, minimizem a endogamia e maximizem o mérito genético da descendência (Weigel e Lin, 2000), deixou de ser uma opção e passou a ser uma necessidade para todos criadores, extensionistas e consultores que queiram contrariar os efeitos do aumento da endogamia na população.

A lista dos 10 touros geneticamente mais influentes (Tabela 2) coincide, em grande parte, com os touros que mais contribuíram para outras populações de vacas Holstein, e.g., Estados Unidos e Alemanha (Young e Seykora, 1996; Swalve *et al.*, 2003). O uso intensivo de alguns touros na população mundial de Holsteins é a principal razão do incremento da endogamia nesta população, podendo já estar a contribuir seriamente para uma diminuição da variabilidade genética com repercussões ao nível reprodutivo e produtivo, com especial

Tabela 2 – Lista dos 10 touros geneticamente mais influentes na população de referência*.

Nome	Ano de Nascimento	Contribuição genética (%)	
		Marginal	Acumulada
To-Mar Blackstar-Et	1983	3,9	3,9
Hanoverhill Starbuck	1979	3,8	7,7
SWD Valiant	1973	2,6	10,3
Rothrock Tradition Leadman	1985	2,4	12,8
Carlin-M Ivanhoe Bell	1974	1,8	14,6
Walkway Chief Mark	1978	1,7	16,2
Singing-Brook N-B Mascot-Et	1986	1,6	17,8
Bis-May Tradition Cleitus	1981	1,4	19,1
Arlinda Rotate	1975	1,0	20,1
Skalsumer Sunny Boy	1985	0,9	20,9

*Fêmeas nascidas entre 1999 e 2003 inclusive, com ambos pais conhecidos.

ênfase no que diz respeito ao número de anos de vida produtiva. Smith et al. (1998) chama a atenção para o facto de serem necessários 1 a 3 anos de produção para que o criador recupere os custos da recria e comece a ter lucros com esse animal. Oltenacu (comunicação pessoal, 2003), refere uma tendência crescente de muitos criadores, nos Estados Unidos, optarem por cruzamentos para resolverem problemas recorrentes de baixa performance reprodutiva, coadjuvado com uma diminuição da vida produtiva dos seus animais.

Depressão de Endogamia

Todos os efeitos incluídos no modelo foram muito significativos ($P < 0,0001$), explicando 75,8%, 82,6% e 82,2% da variação total para a produção de leite, gordura e proteína, respectivamente. O efeito de endogamia provocou uma depressão de $-6,6 \pm 1,5$ kg, $-0,17 \pm 0,052$ kg e $-0,18 \pm 0,042$ kg de leite, gordura e proteína, respectivamente, produzido na 1ª lactação, por cada 1% de aumento no CE. Estes valores são inferiores aos publicados por outros autores. Miglior et al. (1992) em vacas Jersey, apresentam reduções de -10,10 kg e -0,62 kg de leite e gordura, respectivamente por incrementos de 1% no CE. Em vacas Holstein, Wiggans e VanRaden (1995), Smith et al. (1998) e mais tarde também Thompson et al. (2000), indicaram reduções ainda maiores, entre -18,7 kg e -35 kg para o leite, -0,66 kg e -1,08 kg para a gordura e -0,59 kg e -0,97 kg para a proteína, por cada 1% de aumento no CE. Comparativamente com estes estudos, os resultados obtidos não parecem ser suficientes para causar uma grande apreensão em termos produtivos. No entanto, a alta frequência de animais com ascendência desconhecida (40,1%), bem como a presença de erros de identificação no pedigree nacional, são factores importantes que podem enviesar estes resultados, explicando os valores relativamente baixos obtidos neste estudo.

Conclusões

A importância da qualidade do ficheiro de pedigree foi claramente demonstrada. A validade de alguns parâmetros genealógicos depende directamente deste factor. Em estudos futuros poderá ser vantajoso criar grupos de pais desconhecidos com um CE correspondente aos seus contemporâneos (nascidos no mesmo ano). Apesar do efeito depressivo obtido neste estudo não parecer causar reduções importantes na produção de leite, gordura e proteína em vacas com CE médio, o aumento intencional da endogamia deve ser evitado, pois os efeitos negativos no animal são globais e cumulativos. As perdas económicas que podem resultar da

diminuição da produção (especialmente em termos de vida produtiva) serão incrementadas, principalmente devido a factores relacionados com a diminuição da expressão dos caracteres ligados ao “fitness” (reprodução e sobrevivência). Nos programas informáticos de emparelhamento, torna-se urgente introduzir a identificação dos touros disponíveis, de modo a prever (e assim minimizar ou maximizar) os níveis de endogamia e o valor genético da descendência futura, instrumento fundamental de apoio aos programas de selecção e melhoramento.

Agradecimentos

Os autores agradecem aos Professores Jorge Colaço, António Rocha e Gertrude Thompson pela revisão do manuscrito. Este estudo insere-se num projecto aprovado pela FCT e pelo POCTI, participado pelo fundo comunitário europeu FEDER. Primeiro autor financiado por uma bolsa de investigação da FCT com a refª SFRH/BD/11397/2002.

Bibliografia

- Boichard, D., Maignel, L. e Verrier, É. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genet Sel Evol*, 29, 5-23.
- Meuwissen, T. H. E. e Luo, Z. (1992). Computing inbreeding coefficients in large populations. *Genet Sel Evol*, 24, 305-313.
- Miglior F., Szkotnicki, B. e Burnside, E. B. (1992). Analysis of levels of inbreeding and inbreeding depression in Jersey cattle. *J Dairy Sci*, 75, 1112-1118.
- Miglior, F., E. B. Burnside, and J.C.M. Dekkers. (1994). Nonadditive genetic effects and inbreeding depression for somatic cell counts of Holstein cattle. *J. Dairy Sci.* 78, 1168-1173.
- Smith, L. A., Cassell, B. G. e Pearson, R. E. (1998). The effects of inbreeding on the lifetime performance of dairy cattle. *J Dairy Sci*, 81, 2729-2737.
- Swalve, H. H., Rosner, F. e Wemheuer, W. (2003). Inbreeding in the German Holstein cow population. *Proc. EAAP – 54th Annual Meeting*, Rome, Italy, 2003, 17.
- Thompson J. R., Everett, R. W. e Hammerschmidt, N. L. (2000). Effects of inbreeding on production and survival in Holsteins. *J Dairy Sci*, 83, 1856-1864.
- VanRaden, P. M. e Smith, L. A. (1999). Selection and mating considering expected inbreeding of future progeny. *J Dairy Sci*, 82, 2771-2778.
- Weigel, K. A. e Lin, S. W. (2000). Use of computerized mate selection programs to control inbreeding of Holstein and Jersey cattle in the next generation. *J Dairy Sci*, 83, 822-828.
- Wiggans, G. R. e VanRaden, P. M. (1995). Calculation and use of inbreeding coefficients for genetic evaluation of United States dairy cattle. *J Dairy Sci*, 78, 1584-1590.
- Young, C. W. e Seykora, A. J. (1996). Estimates of inbreeding and relationship among registered Holstein females in the United States. *J Dairy Sci*, 79, 502-505.

Artigo 3*

Consequências da Eliminação de Rebanhos Pequenos da Avaliação
Genética de Bovinos Leiteiros em Portugal

*Effects of the Exclusion of Small Herds from the Genetic Evaluations of Dairy
Cattle in Portugal*

*Vasconcelos, J., A. Martins, A. Ferreira, e J. Carvalheira (2006) Consequências da Eliminação de Rebanhos Pequenos da Avaliação Genética de Bovinos Leiteiros em Portugal. Revista Portuguesa de Zootecnia, ano XII nº2: 105-117

EFFECTS OF THE EXCLUSION OF SMALL HERDS FROM THE GENETIC EVALUATIONS OF DAIRY CATTLE IN PORTUGAL

J. VASCONCELOS^{1,2}, A. MARTINS³, A. FERREIRA³ e J. CARVALHEIRA^{1,2}

¹CIBIO-ICETA, Centro de Investigação em Biodiversidade e Recursos genéticos - Universidade do Porto. Campus Agrário de Vairão, Rua Padre Armando Quintas, 4485-661 Vairão.

²ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto

³ABLN, Associação para o Apoio à Bovinicultura Leiteira do Norte
.joanav@mail.icav.up.pt.

(Recepção: 15 de Outubro 2005; Aceitação: 11 de Novembro de 2005)

ABSTRACT

Small herds may limit the use of herd-test-day as a fixed effect in test-day models due to insufficient number of observations per contemporary group. Some authors recommend the exclusion of those farms but do not discuss the implications that the loss of information may cause to the genetic evaluations. This study simulated 3 data subsets using the National test-day database (**DB**) by gradually eliminating small herds in order to have at least a minimum of 3, 10 and 15 observations per level of fixed effects in the model. The objective was to study trends and the impact of this restriction on some population parameters, such as variance components (**VC**), accuracy of prediction of breeding values (**ACC**) and changes in the genetic rankings of the animals. There was no significant difference ($p>0.3$) between estimated VC, implying that the 3 DB were originated from the same population. On average, there was a negative trend on ACC (0.868, 0.861 and 0.860) according to the level of restriction imposed on the 3 DB (from 3 to 15 observations per contemporary group), implying a diminished opportunity for sires to compete. The correlation between the genetic rankings of the animals was low, especially between sires (from 0.21 to 0.32). Eliminating small herds from the genetic evaluations increased the bias on the prediction of breeding values and was considered a form of preferential treatment of sires. In a relatively small data set like the Portuguese test day DB, the elimination of small herds is not recommended. This emphasizes the need to evaluate other approaches of assembling contemporary groups for test day models.

Keywords: Accuracy of prediction, contemporary group, herd-test-day effect, small herds, test day model

CONSEQUÊNCIAS DA ELIMINAÇÃO DE REBANHOS PEQUENOS DA AVALIAÇÃO GENÉTICA DE BOVINOS LEITEIROS EM PORTUGAL

RESUMO

A existência de rebanhos com um número insuficiente de observações por grupo contemporâneo põe em causa alguns pressupostos estatísticos dos modelos de contrastes, impondo limites na utilização do rebanho-ano-mês como efeito fixo. Alguns autores optam pela exclusão destes rebanhos sem, no entanto, discutirem as consequências desta perda de informação para as avaliações genéticas. Este estudo utilizou a informação Nacional de contrastes para simular 3 bases de dados (BD), eliminando progressivamente as pequenas explorações (mínimo de 3, 10 ou 15 observações por nível de efeitos fixos no modelo), de modo a evidenciar variações e possíveis tendências nos componentes de variância (CV), rigor de predição (RP) e alterações de ranking dos valores genéticos dos reprodutores. Não houve diferenças significativas ($p>0,3$) entre os CV estimados, assegurando que o processo de amostragem foi aleatório, no sentido de que as 3 BD representam a mesma população. Em média, verificou-se uma tendência decrescente no RP (0,868, 0,861 e 0,860) à medida que aumentou a restrição do número mínimo de observações por nível de efeitos fixos, o que impede alguns reprodutores de competirem em pé de igualdade com os seus pares. As diferenças no ranking genético entre as 3 BD foram muito marcadas, com correlações baixas, especialmente entre touros (de 0,21 a 0,32). Concluiu-se que a exclusão de rebanhos pequenos é uma forma de tratamento preferencial de touros, o que inflaciona o seu valor genético e contribui para o aumento do enviesamento da avaliação genética.

Palavras chave: Avaliação genética, efeito rebanho-ano-mês, grupos contemporâneos, modelo de contrastes, rebanhos pequenos

INTRODUÇÃO

As vantagens dos modelos de contrastes (MC) sobre os modelos clássicos de produção acumulada aos 305 dias na avaliação genética dos bovinos leiteiros, estão bem estabelecidas e já não oferecem discussão (Ptak e Schaeffer, 1993; Swalve, 1995; Jamrozik e Schaeffer, 1997; Jamrozik *et al.*, 1997; Wiggans e Goddard, 1997; Carvalheira *et al.*, 1998, 2002a). Existem, porém, alguns constrangimentos identificados nos MC, nomeadamente em relação aos pequenos rebanhos. Se por um lado, a partição dos efeitos ambientais sistemáticos em períodos temporais de curta duração (*e.g.* um mês), conseguem diminuições drásticas nas variâncias ambientais permitindo uma maior expressão da

componente genética, por outro lado, o número de observações em cada grupo contemporâneo (**GC**) torna-se crítico, especialmente em explorações de pequena dimensão. Nestes casos, o número de observações pode ser insuficiente para definir os GC, limitando a possibilidade de utilização do rebanho-ano-mês (**RAM**) como efeito fixo no modelo.

Considerar o RAM como mais um efeito aleatório no modelo seria uma forma de contornar o problema. Esta via foi estudada em detalhe por Schaeffer e Sullivan (1994), apresentando no entanto, algumas desvantagens. A necessidade de estimar mais componentes de variância (**CV**) em modelos que já por si são de grande complexidade, pode tornar o processo demasiado exigente e inviável em termos computacionais. Outros autores apontam para a extrema importância de ajustar os dados para o enviesamento provocado pela utilização não aleatória de touros por exploração (Henderson, 1973; Meyer, 1987; Schaeffer, 1987; Van Vleck, 1987; Visscher e Goddard, 1993), o qual só é possível quando os efeitos sistemáticos ambientais são assumidos como efeitos fixos no modelo.

Há, portanto, vantagens e simplificações importantes em se assumir o efeito dos GC como fixo. Das medidas já propostas, incluem-se técnicas de formação de “clusters”, quer dentro do rebanho (e.g., Schmitz *et al.*, 1991; Carabaño *et al.*, 2004), quer entre rebanhos (e.g., Carvalheira, 2000), ou mesmo uma combinação de ambos os conceitos (e.g., Strabel e Szwaczkowski, 1999). Com o mesmo objectivo, na Alemanha, Swalve (1995) definiu os GC com base no RAM, mas impondo um limite mínimo de observações anuais por exploração (60 lactações por ano), o que garantiria pelo menos 3 contrastes por GC. Esta restrição elimina explorações com reduzido número de observações por GC. A aplicação desta restrição implica a perda de parte da informação, perda esta que pode ser significativa no caso da base de dados portuguesa, cuja percentagem de explorações com número insuficiente de observações por GC, é muito elevada.

Torna-se assim importante, antes de avaliar em detalhe as consequências da aplicação de técnicas de formação de “clusters”, estudar o impacto que a eliminação de pequenos rebanhos terá na avaliação genética Nacional usando o MC. Os objectivos do presente estudo centram-se portanto, na análise da variação de alguns parâmetros genéticos da população Holstein Nacional, de acordo com a aplicação de diferentes critérios para a determinação do tamanho mínimo dos GC.

MATERIAL E MÉTODOS

A capacidade de particionar os efeitos ambientais sistemáticos em intervalos mensais, permite aos MC tomarem em consideração fontes de variação ambiental que, em outros modelos, contribuem para o enviesamento das avaliações genéticas. Dos vários MC, o auto-regressivo difere dos restantes por assumir que a expressão da produção de leite, ao longo da vida produtiva da fêmea, corresponde a uma única característica quantitativa controlada pelo mesmo grupo de genes.

Este modelo, ao assumir uma única característica e por consequência, um único componente de variância genético, não significa que a heterogeneidade da variância ambiental não seja considerada. Pelo contrário, as diferenças ambientais entre lactações e entre contrastes são tomadas em conta (efeito aleatório do ambiente a longo e curto prazo), resultando em diferentes heritabilidades para cada lactação. Este pressuposto implica que o modelo de análise possua uma estrutura de covariância não genética, onde as medições repetidas são consideradas como variáveis longitudinais representadas como processos auto-regressivos de primeira ordem dentro e entre lactações (Carvalho *et al.*, 1996, 1998, 2002a). O modelo adoptado neste estudo, pode ser descrito da seguinte forma:

$$y_{ijkLmn} = RAM_i + Idade(R)_j + DEL(R)_{k(L)} + a_m + p_{m(L)} + t_{n(mL)} + r_{ijkLmn},$$

onde,

y_{ijkLmn} representa o valor do contraste;

RAM_i mede o efeito fixo do rebanho-ano-mês em que se observou o contraste;

$Idade(R)_j$ mede o efeito fixo hierarquizado por rebanho correspondente à j ésima classe de idade ao parto;

$DEL(R)_{k(L)}$ mede o efeito fixo devido aos animais testados na mesma classe de dias-em-lactação (DEL) para cada rebanho e lactação (curvas de lactação);

a_m representa o efeito aleatório do animal;

$p_{m(L)}$ representa o efeito aleatório do ambiente a longo prazo, medindo a correlação criada pela vaca ao produzir em mais do que uma lactação, e para o qual se assume um processo auto-regressivo de primeira ordem;

$t_{n(mL)}$ representa o efeito aleatório do ambiente a curto prazo, que mede a correlação entre os contrastes da mesma vaca para cada

lactação e que segue também um processo auto-regressivo de primeira ordem. Assume-se que os contrastes são independentes entre lactações;

r_{ijkLmn} é o termo residual aleatório.

A base de dados (BD) de contrastes do efectivo leiteiro Nacional, com produções entre Junho de 1994 e Dezembro de 2004, foi cedida pela Associação Nacional para o Melhoramento dos Bovinos Leiteiros. A edição dos dados foi baseada nos pressupostos assumidos pelo modelo e descrita detalhadamente em Carvalheira *et al.* (1998) e Vasconcelos *et al.* (2001). Em resumo, da BD inicial conservaram-se apenas as observações pertencentes às primeiras três lactações de cada animal e os contrastes do tipo A4 (manhã e tarde). Foram também eliminadas todas as observações com menos de 5 ou mais de 310 DEL e garantiu-se que todas as fêmeas teriam pelo menos 2 contrastes consecutivos. No final deste processo foram validadas 3.537.305 observações distribuídas por 1.310 explorações, com base na produção de 214.618 vacas, filhas de 12.154 touros.

Para estudar as consequências da eliminação de rebanhos de pequena dimensão, foram comparados parâmetros populacionais obtidos a partir de 3 BD criadas com base nos seguintes critérios: BD3 - onde explorações com menos de 3 observações em pelo menos um nível de efeitos fixos no modelo foram eliminadas (2.956.763 observações distribuídas por 788 explorações, produzidas por 175.487 vacas, filhas de 10.831 touros); BD10 – onde o limite mínimo foram 10 observações por nível de efeitos fixos (1.054.775 observações distribuídas por 147 explorações, produzidas por 62.672 vacas, filhas de 5.521 touros) e; BD15 - contendo apenas explorações com pelo menos 15 observações em todos os níveis de efeitos fixos (290.797 observações distribuídas por 31 explorações, produzidas por 16.784 vacas, filhas de 2.042 touros).

Ao eliminar de forma progressiva as explorações mais pequenas, pretendeu-se evidenciar variações e possíveis tendências nestes parâmetros que permitissem concluir sobre a viabilidade da utilização destas BD truncadas nas avaliações genéticas Nacionais. Os parâmetros avaliados neste estudo incluíram comparações entre CV e autocorrelações, heritabilidade, rigor de predição dos valores genéticos e alterações no ranking genético dos animais presentes. Os CV e autocorrelações foram estimados pelo método DFREML-SIMPLEX (Smith e Graser, 1986) recorrendo a um processo de amostragem aleatória estratificada com 5 réplicas, de modo a garantir a representação de explorações em todas as regiões do País.

RESULTADOS E DISCUSSÃO

O Quadro I contém os valores estimados para os CV, autocorrelações e heritabilidade nas 3 BD em estudo. Os valores obtidos neste trabalho são similares aos previamente estimados em Portugal (Carvalho et al., 2002b), usando o mesmo modelo animal e estão dentro dos limites descritos na literatura (Carvalho et al., 2002a; Druet et al., 2005; Melo et al., 2005). Verificou-se uma tendência muito ligeira para o aumento nos CV associada positivamente ao tamanho dos GC que definem as 3 BD. Estas diferenças podem também ser observadas entre os valores de heritabilidade obtidos. A ausência de significância estatística entre estas diferenças ($p>0,3$), valida o processo aleatório de amostragem, no sentido das 3 BD representarem a mesma população. Por este motivo e para efeitos deste trabalho, foi utilizada a média aritmética dos 3 estimadores para o cálculo dos restantes parâmetros analisados.

QUADRO I - COMPONENTES DE VARIÂNCIA (KG^2) E AUTOCORRELAÇÕES PARA A PRODUÇÃO DIÁRIA DE LEITE NO EFECTIVO HOLSTEIN NACIONAL PARA AS PRIMEIRAS 3 LACTAÇÕES (L_1 , L_2 e L_3) NAS 3 BASES DE DADOS (BD3, BD10 e BD15).

Parâmetro	BD3 ^a	BD10	BD15
Variância genética	7,9363	8,2256	8,7843
Variância ambiental	3,7009	3,9786	4,4478
Variância ALP ^b	<0,0001	<0,0001	<0,0001
Autocorrelação ALP	<0,0001	<0,0001	<0,0001
Variância ACP ^c (L_1)	12,5742	13,5512	13,0997
Autocorrelação ACP (L_1)	0,8412	0,8250	0,8294
Variância ACP (L_2)	24,7861	25,8795	26,2600
Autocorrelação ACP (L_2)	0,8327	0,8164	0,8200
Variância ACP (L_3)	29,3067	30,4231	30,5559
Autocorrelação ACP (L_3)	0,8241	0,8044	0,8003
Heritabilidade (L_1)	0,3278	0,3194	0,3336
Heritabilidade (L_2)	0,2179	0,2160	0,2224

^aBD3 = 3 ou mais observações por grupo contemporâneo; BD10 = 10 ou mais observações por grupo contemporâneo; BD15 = 15 ou mais observações por grupo contemporâneo. ^bALP = Efeito Ambiental a Longo Prazo; ^cACP = Efeito Ambiental a Curto Prazo.

O rigor de predição (RP), definido como a correlação teórica entre os valores genéticos aditivos (VGA) verdadeiros (nunca observados) e os VGA estimados, é um critério fundamental de selecção num programa de melhoramento animal (Pollak e Quaas, 1983). Este parâmetro é função da variância genética aditiva e da variância do erro de predição, a qual depende em grande medida do número de observações que cada animal contribui para a sua avaliação (Henderson 1975; Swalve, 1995). No que diz respeito aos touros, a diminuição do número de explorações entre BD e, por consequência, do número de filhas e respectivas observações, teve um impacto marcado na distribuição das frequências do RP, como se pode observar na Fig. 1. Apesar de serem elevados (médias de 0,868, 0,861 e 0,860 para a BD3, BD10 e BD15, respectivamente), verifica-se que na BD3 os touros apresentam, em média, RP mais altos e a tendência é decrescente à medida que aumenta a restrição do número mínimo de observações por nível de efeitos fixos nas 3 BD. Seria de esperar que, quanto maior fosse o número de contemporâneos em cada grupo, menor a variância do erro de predição e, portanto, maior o RP. No entanto, verificou-se que o peso da contribuição colectiva do número total de filhas (que diminui com a eliminação de explorações) foi determinante para a magnitude deste parâmetro, pois, tal como Tosh e Wilton (1994) referem, o número total de observações tem uma elevada correlação com o RP. A eliminação de explorações de dimensão menor poderá portanto ter, consequências negativas no melhoramento genético, que neste caso se traduzem na perda de oportunidade de selecção de alguns reprodutores.

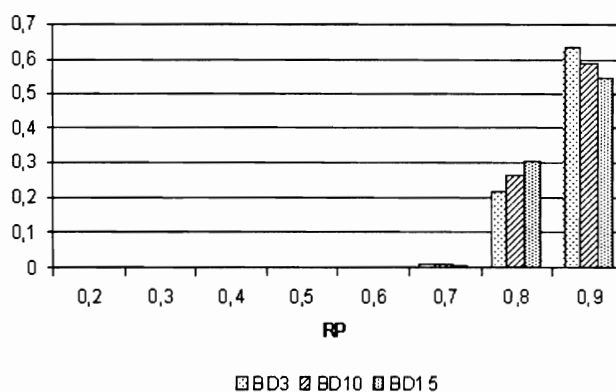


Figura 1. Frequência de valores do rigor de predição (RP) estimados em machos com 15 ou mais descendentes, em pelo menos 5 explorações, nas 3 bases de dados em estudo (BD3, BD10 e BD15 com 3, 10 e 15 ou mais observações por grupo contemporâneo, respectivamente).

Por outro lado, eliminar os rebanhos mais pequenos pode ser também uma forma de tratamento preferencial no que diz respeito à selecção de reprodutores machos. De uma maneira geral, as explorações maiores investem mais na genética (Henderson, 1973; Schaeffer, 1991), sendo de esperar que o progresso genético anual nestes rebanhos seja proporcionalmente maior do que nos mais pequenos. Este aspecto está reflectido nos gráficos apresentados na Fig. 2. Enquanto o progresso genético nas fêmeas foi praticamente igual nas 3 BD, o mesmo não se verificou nos machos, onde o impacto desta forma de tratamento preferencial foi bastante acentuado (Fig. 2). O caso extremo é representado pelo progresso genético anual dos machos com 10 e mais filhas, onde a diferença observada entre a BD3 e a BD15 (1996 a 1998, inclusive), é de 141 kg anuais de leite por lactação (Quadro II). Esta diferença reflecte a eliminação selectiva de filhas provenientes de explorações que em média são geneticamente inferiores e que, na sua maioria, correspondem às explorações mais pequenas. A principal consequência desta forma de tratamento preferencial de touros é um incremento artificial do seu valor genético, que potencialmente pode contribuir para a inflação dos custos da componente reprodução.

QUADRO II- COEFICIENTES DE REGRESSÃO (KG/ANO) DO PROGRESSO GENÉTICO PARA A PRODUÇÃO ACUMULADA DE LEITE AOS 305 DIAS EM MACHOS NASCIDOS ENTRE 1996 E 1998 (INCLUSIVE), NAS 3 BASES DE DADOS EM ESTUDO.

	BD3 ^a	BD10	BD15
M3 ^b	101,14	86,91	92,15
M5	79,65	110,94	171,18
M10	91,19	105,83	231,93

^aBD3 = 3 ou mais observações por grupo contemporâneo; BD10 = 10 ou mais observações por grupo contemporâneo; BD15 = 15 ou mais observações por grupo contemporâneo. ^bM3, M5 e M10 = machos com pelo menos 3, 5 e 10 filhas respectivamente).

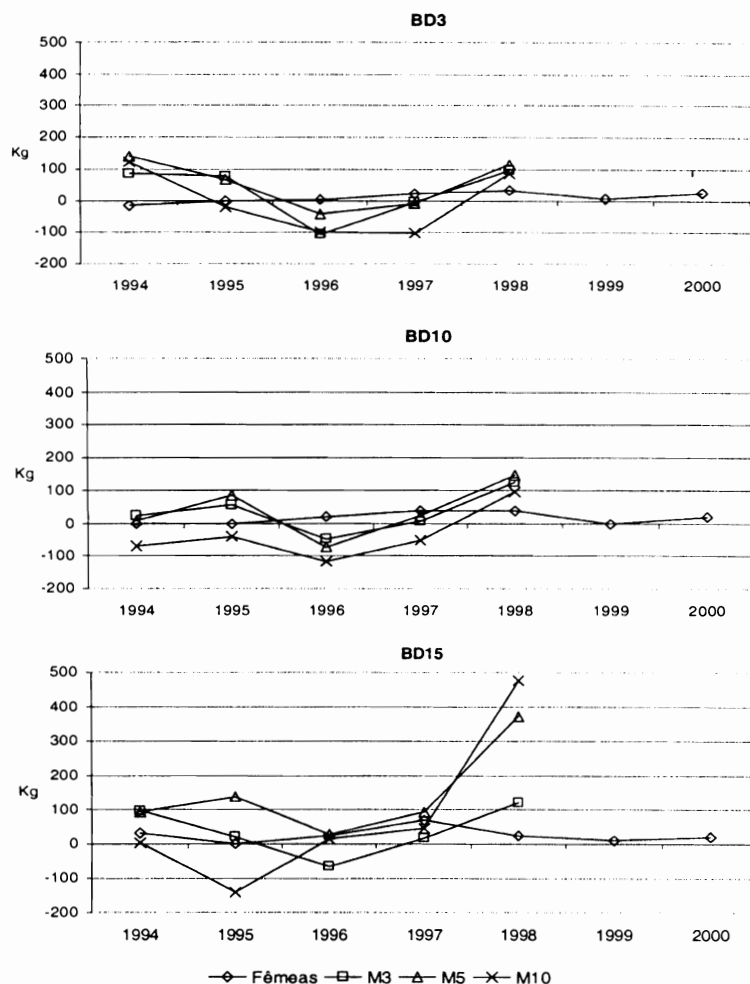


Figura 2. Progresso genético anual do efectivo Nacional Holstein para a produção acumulada de leite aos 305 dias, nas 3 bases de dados em estudo (BD3, BD10 e BD15 com 3, 10 e 15 ou mais observações por grupo contemporâneo, respectivamente; M3, M5 e M10 = machos com pelo menos 3, 5 e 10 filhas respectivamente).

O RP é um indicador do nível de confiança estatística dos VGA estimados e, é com base nestes últimos que, em termos práticos, se identificam os animais geneticamente superiores e se decide sobre a sua eventual selecção como reprodutores. Em características limitadas ao sexo, a avaliação (neste caso, dos machos) é indirecta e depende maioritariamente do seu número de filhas e em

quantas explorações elas são contrastadas (Pollak e Quaas, 1983). À partida, é portanto de esperar que a exclusão de rebanhos, mesmo que de pequenas dimensões, venha a alterar o ranking dos VGA dos animais avaliados, com especial ênfase nos machos. O impacto destas alterações na economia das explorações e no progresso genético da população, pode ser de extrema importância, pois é um dos factores que determinam a competitividade entre touros e consequentemente, influencia de forma dramática a indústria produtora de sêmen seleccionado. Relativamente à BD3, a percentagem dos animais "top 100" em comum com a BD10 foi de 21% em machos e de 35% em fêmeas. Comparando com a BD15, essa percentagem baixou consideravelmente, ficando apenas com 7% de machos e 17% de fêmeas em comum com a BD3. O Quadro III dá uma indicação deste reordenamento, mostrando a "negrito" os animais em comum nas 3 BD, para os primeiros 10 animais de cada sexo. As correlações de ranking entre os VGA obtidos com base na avaliação das 3 BD (Quadro IV) confirmam estas alterações e mostram que elas são mais marcadas nos machos do que nas fêmeas. Com correlações de ranking tão baixas (como 0,21), torna-se claro que a exclusão de rebanhos deve ser limitada ao mínimo e o tamanho dos GC deve ser garantido recorrendo a outros procedimentos (e.g., técnicas de agrupamento). Esta conclusão é partilhada por vários autores (Schmitz *et al.*, 1991; Strabel e Szwaczkowski, 1999; Carvalheira, 2000; Carabaño *et al.*, 2004) ficando, no entanto, por definir a melhor metodologia de modo a simultaneamente, minimizar o enviesamento dos VGA e maximizar o RP.

QUADRO III - LISTA DOS GENETICAMENTE 10 MELHORES ANIMAIS NAS 3 BASES DE DADOS EM ESTUDO.

Ranking	Machos			Fêmeas		
	BD3 ^a	BD10	BD15	BD3	BD10	BD15
1	875531^b	616143	169117	832024	872987	501393
2	26676	709711	309137	782647	783242	596715
3	709711	364910	1022843	526378	834592	834592
4	633788	875531	216819	565022	501393	773560
5	200133	111330	364291	653428	565022	856449
6	713843	530611	155730	804910	653428	832024
7	227807	215927	388763	477853	1067455	597146
8	878655	620028	531645	501393	570613	475229
9	67548	67548	713687	866800	969399	373476
10	1085933	146361	77111	703292	490978	792701

^aBD3 = 3 ou mais observações por grupo contemporâneo; BD10 = 10 ou mais observações por grupo contemporâneo; BD15 = 15 ou mais observações por grupo contemporâneo. ^banimais em "negrito" são comuns entre bases de dados.

QUADRO IV - COEFICIENTES DE CORRELAÇÃO ENTRE O RANKING GENÉTICO PARA MACHOS (ACIMA DA DIAGONAL) E FÊMEAS (POR BAIXO DA DIAGONAL), NAS 3 BASES DE DADOS EM ESTUDO.

	BD3 ^a	BD10	BD15
BD3	--	0,32	0,21
BD10	0,83	--	0,28
BD15	0,82	0,83	--

^a**BD3** = 3 ou mais observações por grupo contemporâneo; **BD10** = 10 ou mais observações por grupo contemporâneo; **BD15** = 15 ou mais observações por grupo contemporâneo.

CONCLUSÕES

Os resultados obtidos neste estudo comprovam que a exclusão de rebanhos de pequena dimensão da avaliação genética pode ter consequências dramáticas na selecção de reprodutores, influenciando negativamente o progresso genético da população. Se, por um lado, é necessário garantir um número mínimo de observações em cada GC, por outro lado há que maximizar o número de filhas e de rebanhos que contribuem para a avaliação dos touros. Agrupar rebanhos pode ser uma solução para este problema, mas torna-se necessário encontrar e desenvolver técnicas que definam quais os critérios de semelhança a serem usados entre eles, de modo a minimizar outras fontes de enviesamento à avaliação genética.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pela FCT, através da bolsa SFRH/BD/11397/2002.

BIBLIOGRAFIA

- CARABAÑO, M.J., MORENO, A., LÓPEZ-ROMERO, P. e DIAZ, C., 2004. Comparing alternative definitions of the contemporary group effect in Avileña Negra Ibérica beef cattle using classical and Bayesian criteria. *J. Anim. Sci.*, 82:3447-2457.
- CARVALHEIRA, J., 1996. Development and application of an autoregressive process for estimating genetic parameters of daily milk yield in a herd of tropical Lucerna cattle and in US Holstein herds. Doctoral Dissertation, Cornell Univ., Ithaca, NY, pp.200.

- CARVALHEIRA, J., 2000. Predição de parâmetros genéticos para a produção diária de leite em gado Holstein, Pardo Suíço e Modicana no sudeste Siciliano. X Congresso de Zootecnia, p.103.
- CARVALHEIRA J., BLAKE, R.W., POLLAK, E.J., QUAAS, R.L. e DURAN-CASTRO, C.V., 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in US Holstein herds. J. Dairy Sci., 81:2738-2751.
- CARVALHEIRA J., POLLAK, E.J., QUAAS, R.L. e BLAKE, R.W., 2002a. An autoregressive repeatability animal model for test-day records in multiple lactations. J. Dairy Sci., 85:2040-2045.
- CARVALHEIRA, J., VASCONCELOS, J., MARTINS, A., PETIM-BATISTA, M.F., SILVESTRE, A., FERREIRA, A. e COLAÇO, J., 2002b. Parâmetros e Tendências Genéticas para Produção Diária de Leite, Gordura e Proteína do Efectivo Nacional Holstein Usando um Modelo de Contrastes Auto-regressivo. XII Congresso de Zootecnia, pp.531- 537.
- DRUET, T., JAFFREZIC, F. e DUCROCO, V., 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. Genet. Sele. Evol., 37(3):257-271.
- HENDERSON, C.R., 1973. Sire evaluation and genetic trends. In: Proc. Anim. Breeding Genet. Symp. In Honor of Dr. J. L. Lush, Am. Soc. Anim. Sci., Am. Dairy Sci. Assoc., Champaign, IL, p.10.
- HENDERSON, C.R., 1975. Comparison of alternative sire evaluation methods. J. Anim. Sci., 41:760-770.
- JAMROZIK, J. e SCHAEFFER, L.R., 1997. Estimates of Genetic Parameters for a Test Day Model with Random Regressions for Yield Traits of First Lactation Holsteins. J. Dairy Sci., 80:762-770.
- JAMROZIK, J., SCHAEFFER, L.R. e DEKKERS, J.C.M., 1997. Genetic Evaluation of Dairy Cattle Using Test Day Yields and Random Regression Model. J. Dairy Sci., 80:1217-1226.
- MELO C.M.R., PACKER, I.U., COSTA, C.N. e MACHADO, P.F., 2005. Genetic parameters for test day and lactation milk yields of first lactation Holstein cows. Rev. Bras. Zoot., 34(3):796-806.
- MEYER, K., 1987. Estimates of variance due to sire x herd interactions and environmental covariances between paternal half-sibs for first lactation dairy production. Livest. Prod. Sci., 17:95-115.
- PTAK, E. e SCHAEFFER, L.R., 1993. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci., 34: 23-34.
- POLLAK, E.J., e QUAAS, R.L., 1983. Definition of group effects in sire evaluation models. J. Dairy Sci. 66(7):1503-1509.
- SCHAEFFER, L., 1987. Estimation of variance components under a selection model. J. Dairy Sci., 70: 661-671.
- SCHAEFFER, L., 1991. C.R. Henderson: Contributions to predicting genetic merit. J. Dairy Sci., 74: 4052-4066.
- SCHAEFFER, L. e SULLIVAN, B., 1994. Genetic evaluation of dairy goats using test day yields. Proc. 5th World Cong. Gen. Appl. Livest. Prod., 18:182-185.
- SCHMITZ, F., EVERETT, R.W. e QUAAS, R.L., 1991. Herd-year-season clustering. J. Dairy Sci., 74: 629-636.

- SMITH, S.P. e GRASER, H.U., 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.*, 69: 1156-1165.
- STRABEL, T. e SZWACZKOWSKI, T., 1999. The use of test day models with small size of contemporary groups. *J. Anim. Breed. Genet.*, 116: 379-386.
- SWALVE, H.H., 1995. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *J. Dairy Sci.*, 78: 929-938.
- TOSH, J.J. e WILTON, J.W., 1994. Effects of data structure on variance of prediction error and accuracy of genetic evaluation. *J. Dairy Sci.*, 72: 2568-2577.
- VAN VLECK, L., 1987. Contemporary groups for genetic evaluations. *J. Dairy Sci.*, 70: 2456-2464.
- VASCONCELOS, J., PETIM-BATISTA, M.F., SILVESTRE, A., FERREIRA, A., MARTINS, A., COLAÇO, J. e CARVALHEIRA, J., 2001. Predição de lactações padronizadas aos 305 dias, de leite, gordura e proteína do efectivo leiteiro Nacional, usando um modelo de contrastes autoregressivo. XI Congresso de Zootecnia, pp.105-106.
- VISSCHER, P.M. e GODDARD, M.E., 1993. Fixed and random contemporary groups. *J. Dairy Sci.*, 76:1444-1454.
- WADE, K.M., 1990. Parameter estimation for mixed linear models with first order autoregressive covariance structure. Doctoral Diss., Cornell Univ. Ithaca, NY, p.117.
- WIGGANS, G.R. e GODDARD, M.E., 1997. A Computationally Feasible Test Day Model for Genetic Evaluation of Yield Traits in the United States. *J. Dairy Sci.*, 80:1795-1800.

Artigo 4*

Effects of Clustering Herds with Small Contemporary Groups in Dairy Cattle

Genetic Evaluations

Efeitos da Aglomeração de Rebanhos com Grupos Contemporâneos

Pequenos na Avaliação Genética

*Vasconcelos, J., F. Santos, A. Bagnato, and J. Carvalheira. 2008. Effects of Clustering Herds in Dairy Cattle Genetic Evaluations. J. Dairy Sci. 91:377-384.

Efeitos da Aglomeração de Rebanhos com Grupos Contemporâneos Pequenos na Avaliação Genética

RESUMO

A maioria dos modelos de contrastes usados em avaliações genéticas define grupos contemporâneos (**GC**) como os animais pertencentes ao mesmo rebanho-ano-mês (**RAM**). Considerar o RAM como um efeito fixo no modelo poderá minimizar o enviesamento das avaliações, mas implica a necessidade de um número mínimo de observações por GC para maximizar o número efectivo de observações, minimizar o erro residual e a variância do erro de predição. A base de dados de contrastes do efectivo leiteiro nacional, com aproximadamente 4 milhões de observações produzidas entre Junho de 1994 e Março de 2006, por 238.271 vacas em 1330 explorações, foi utilizada para avaliar o efeito da aglomeração de GC de pequenas explorações com base na semelhança dos seus ambientes produtivos. A aplicação de uma análise de componentes principais permitiu a redução das 14 variáveis descritivas originais para 5 vectores próprios que explicaram 88% da variância total. Baseadas na matriz de distâncias, foram aplicadas duas metodologias para a formação de “clusters”. Para cada metodologia foram criadas 4 bases de dados nas quais o número mínimo observações por GC foi de 3, 5, 10 e 15 contrastes. Na metodologia “A” todas as explorações, com ou sem o número mínimo de observações foram candidatas para formar “clusters”, enquanto que na “B”, apenas as explorações com pelo menos um GC com um número de observações inferior ao requerido puderam participar no processo de aglomeração e formar “clusters”. As bases de dados foram analisadas usando um modelo animal para contrastes auto-regressivo de múltiplas lactações, considerando o efeito RAM como fixo e os resultados comparados com os obtidos pela metodologia actualmente usada em Portugal nas avaliações genéticas. A metodologia “B”, com um mínimo de 3 observações por GC, foi a que obteve menores diferenças entre valores genéticos aditivos, assim como maiores rigores de predição e menores variâncias dentro de CG, revelando um maior ajustamento aos dados. Esta metodologia foi a que permitiu o maior grau de preservação da estrutura de dados original, maximizando o número de grupos de comparação usados no processo de avaliação genética. As correlações entre os valores genéticos dos animais, de ranking, variância do erro de predição e rigor de predição obtidas (0,97, 0,97, 0,85 e 0,82, respectivamente) foram elevadas, sugerindo uma reduzida influência do processo de aglomeração na ordenação dos animais.

Effects of Clustering Herds with Small-Sized Contemporary Groups in Dairy Cattle Genetic Evaluations

J. Vasconcelos,*† F. Santos,* A. Bagnato,‡ and J. Carnevalheira*†¹

*CIBIO/ICETA [Au: Please spell out acronyms], Research Center in Biodiversity and Genetic Resources, University of Porto, Rua Padre Armando Quintas-Crasto, 4485-661 Vairão, Portugal

†Instituto de Ciências Biomédicas Abel Salazar (ICBAS), University of Porto, Portugal

‡Department of Veterinary Science and Technology for Food Safety, Faculty of Veterinary Medicine, University of Milan, Italy

ABSTRACT

Most test-day models used in genetic evaluations of dairy cattle define contemporary groups (CG) as the herd-test-date effect. Fitting this effect as fixed may minimize prediction bias, but requires a minimum number of observations per CG to simultaneously maximize the effective number of observations and minimize the residual error and prediction error variance. Nearly 4 million test-day records from the Portuguese Holstein database of 238,271 cows calving in 1,330 herds from 1994 through 2006 were used to evaluate the effect of clustering CG from small herds based on the similarity of their production environments. Principal component analysis was used to summarize 14 descriptive variables in 5 eigenvectors that explained 88% of the total variation. Based on the distance matrix, 2 different approaches were applied to group the herds. For each approach, 4 data sets were built having at least 3, 5, 10, or 15 observations per CG, respectively. For the data sets of group A, all herds, with or without the required number of observations per CG, were used in the clustering process. For the data sets of group B, only herds without the minimum number of observations were candidates to form clusters. All data sets were analyzed by an autoregressive test-day animal model fitting a fixed herd test date in a multiple-lactation setting, and results were compared with the current clustering procedure used in the Portuguese genetic evaluations. The data set from group B, with a minimum of 3 records per CG, was the one that provided the highest accuracy of prediction and the smaller within-CG variance, revealing a better fit for the data. This procedure also preserved the original herd structure of the database, better maximizing the number of herd groups. Correlations among EBV, rank, prediction error variance, and accuracies of prediction for this data

set were high (0.97, 0.97, 0.85, and 0.82, respectively), suggesting that no major reranking is to be expected.

Key words: contemporary group, principal component analysis, cluster analysis, autoregressive test-day model

INTRODUCTION

In genetic evaluations, animal performances are compared within contemporary groups (CG) of assumed similar environmental conditions (Schaeffer, 1987; Visscher and Goddard, 1993; Crump et al., 1997; Van Bebber et al., 1997). Most test-day (TD) models used in dairy cattle genetic evaluations define the CG as the herd-test-date (HTD) effect. Fitting this effect as fixed may minimize prediction bias if nonrandom associations with sires occur (Schaeffer, 1987; Van Vleck, 1987; Visscher and Goddard, 1993), as is the case when herd size or level of production is systematically related to the quality of semen used in AI. This may be an example of preferential treatment (Meyer, 1987; Van Vleck, 1987; Tierney and Schaeffer, 1994; Raffrenato et al., 2003) and is a common situation in Portugal, where high-ranking sires are preferentially used in high-production herds. As a consequence, criteria for forming CG that correctly account for the management impact on recorded phenotypes are required.

Fitting CG as fixed, random, or a combination of both is a matter of balance between accuracy and bias (Van Vleck 1987; Visscher and Goddard, 1993; Van Bebber et al., 1997). The random approach may recover some information across CG (Strabel et al., 2005), especially when the number of observations in some CG is small, but it will introduce bias to the breeding value predictions if there are nonrandom associations of animals with effects considered to be fixed (Van Bebber et al., 1997). Partitioning the environmental variables in fixed and random effects (Wade and Quaas, 1993; Van Bebber et al., 1997) may be a viable approach to simultaneously account for prediction bias and loss of information. Following this line of thinking, Carabaño et al.

Received March 16, 2007.

Accepted September 20, 2007.

¹Corresponding author: jgc3@mail.icav.up.pt

(2006) compared alternative models to account for herd effects in TD milk yield data by comparing several combinations of HTD as fixed or random and with or without a random or fixed herd-year effect. They found small differences among the models, but fitting a fixed HTD and a random herd-year effect showed the worst results. Although the models that included a random HTD had the best results, the authors emphasized the need for additional work for more conclusive recommendations about modeling CG.

The need to simultaneously minimize the variance of residuals and the prediction error variance (**PEV**), assuming a fixed HTD, implies a minimum number of observations per CG. Operationally, the HTD effect is defined as the nested effect of month-of-test within year and herd, which, in small herds, may have only 1 or 2 observations. The definition of this minimum number of observations may become a limiting factor if the number of available TD records for the genetic analysis is critical; therefore, all data, including those from small herds without enough records per CG, have to be included (Van Vleck, 1987; Strabel and Szwaczkowski, 1999; Vasconcelos et al., 2005). Schmitz et al. (1991) suggested a minimum of 10 to 15 observations per CG, whereas Carabaño et al. (2004) and Vasconcelos et al. (2006) suggested a minimum of 5 and 3 observations, respectively. With these criteria, there are many small and recent farms that may be excluded from the genetic evaluations unless some methodology is devised to form larger CG and at the same time minimize the impact that the change in the environmental and genetic structure of the data may bring to the genetic evaluations.

In Portugal, for example, defining CG by levels of the HTD effect resulted in 97 and 40% of dairy farms having at least one HTD class with less than 15 and 3 observations per CG, respectively. In terms of the percentage of CG lacking that number of observations, the average was 37 and 7%, respectively. If these farms are not included in the evaluations, the sire EBV will be biased (Vasconcelos et al., 2005) because of the loss of information on their daughters, with negative consequences on the genetic progress (Swalve, 1995). Furthermore, from a breeder's point of view, discarding these farms (animals) may not be acceptable because this information is needed for on-farm selection programs (Carabaño et al., 2004).

Defining the optimum CG continues to be a controversial issue. If a CG spans a long period of time, the accuracy of the evaluations will be greater because of the larger number of contemporaries in the same group. However, assigning records from animals that are not subject to the same management and environmental conditions to the same CG can introduce bias (Crump et al., 1997; Carabaño et al., 2004). A balance between

high accuracy and small bias needs to be reached to optimize the definition of CG (Van Bebber et al., 1997; Carabaño et al., 2004). Some authors suggested clustering CG, either joining HTD classes within the herd (Schmitz et al., 1991; Carabaño et al., 2004), across herds (Vasconcelos et al., 2006), or a combination of both (Strabel and Szwaczkowski, 1999). If management information (e.g., adjusted monthly production means within a year) is also a required output of the evaluation process, the month classes within year and herd should be maintained, and clustering should be performed among herds with similar environments, keeping the test-date structure unchanged (Vasconcelos et al., 2006).

The ongoing genetic evaluations in Portugal have followed this latter approach. However, the optimal minimum number of observations per CG that maximizes the improvement of accuracy has never been tested. Furthermore, the criteria used to define management similarities among herds have been based only on the phenotypic daily mean (milk, fat, and protein) and standard deviation at first lactation. Additionally, the clustering process was only stopped at the tree level, where all formed clusters had, simultaneously, the minimum predefined number of observations per level of fixed effects. This resulted in very large clusters because, although some of the clusters had already fulfilled the clustering criteria at a lower tree level, they all depended on the other clusters that still needed more observations in some CG. The major implication of this approach was that the final number of herds forming a cluster depended only on the tree organization (defined by the distance matrix), making it impossible to maximize the number of comparison herd groups.

The aim of the present work was to study the consequences of clustering herds in dairy cattle genetic evaluations by varying the minimal number of observations per CG, using either all herds in the data set as potential candidates or only herds without the minimum number of observations in any CG. For consistency, HTD effects were assumed fixed in all analytical models. Principal components (**PC**) analysis was also used to refine the definition of which descriptors of the production environments should be used to evaluate similarities among herds. To improve the control of the clustering process, additional parameters were introduced to minimize the number of herds per cluster (increasing the total number of CG).

MATERIALS AND METHODS

Data

The data were provided by the Portuguese Dairy Cattle Breeders Association and consisted of TD milk yields

Table 1. Eigenvalues (λ_i) and proportion of the variance (PV) accounted for by the first 5 principal components (PC), based on 14 descriptive variables to measure herd production environments

Principal component	λ_i	PV	Cumulative PV
PC ₁	8.0147	0.5725	0.5725
PC ₂	1.3863	0.0990	0.6715
PC ₃	1.1798	0.0843	0.7558
PC ₄	0.9236	0.0660	0.8217
PC ₅	0.7548	0.0539	0.8751

of Portuguese Holstein cows recorded from June 1994 through March 2006. Monthly records were from supervised recording plans (morning and afternoon milking according to standard A4 of the International Committee for Animal Recording) from cows in the first 3 parities. Lactations were required to have at least 2 TD records. The TD records were validated when the first reported TD did not exceed 75 DIM, and the time interval between consecutive TD was between 26 and 33 d for the interval from 5 to 305 DIM. Further edits followed International Committee for Animal Recording (1995) rules for dairy cow TD recording. After editing, the data set comprised 3,988,993 TD records of 238,271 cows in 1,330 herds.

Descriptive Variables

Fourteen descriptive variables were considered as potential indicators of production environments for each herd and were used in the PC analysis to measure their importance as sources of variation among herds: herd size (**HS**)—the number of cows calving in the herd during the study period; average age at first calving; peak yield (**PY**)—average of the maximum TD yield of cows in the herd; peak yield variability—standard deviation of PY; production level (**PL**)—average daily milk yield; and production variability—standard deviation of PL in the herd. The last 4 variables (PY, peak yield variability, PL, and production variability) were calculated for all 3 lactations in each herd. Not all herds had production information for all 3 lactations (less than 0.7%). In these cases, missing values (required for PC analysis) were estimated by linear regression within herd.

PC and Cluster Analysis

Given these descriptive variables, 14 PC were computed based on the analysis of the correlation matrix (PROC PRINCOMP, SAS Institute, 1989). The first 5 PC accounted for 88% of the total variance (Table 1) and were used as input variables for the subsequent cluster analyses.

The method used to find environmental similarities between pairs of herds was Ward's minimum variance (PROC CLUSTER, SAS Institute, 1989). Based on the distance matrix, 2 different approaches (A and B) were applied to group the herds. For each approach, 4 data sets were built having at least 3, 5, 10, and 15 observations per CG (A₃, A₅, A₁₀, A₁₅, B₃, B₅, B₁₀, and B₁₅). For the A data sets, all herds, with or without the required number of observations per CG, were used in the clustering process. For the B data sets, only herds without the minimum number of observations were candidates to form clusters. This allowed more herds to keep their original structure. Herds were clustered according to the following criteria:

1. Herds were sorted according to distances provided by PROC CLUSTER (SAS Institute, 1989) so that adjacent herds in the list represented the closest ones in terms of similarities in management.
2. In the A approach, cluster formation had to start with a herd without the minimum number of observations in at least one CG, which was not required for the B approach.
3. Herds were joined following the branches of the tree, reflecting the shortest distance between them.
4. If there was more than one herd at the same distance, priority to join the cluster was given to the one with the smaller HS. This criterion permitted the formation of more clusters.
5. After each agglomeration (herd-herd or herd-cluster), the program recounted all observations at each CG and stopped the process for that specific cluster if all CG had the required minimum number.
6. If not, the clustering process continued, first by checking if there were herds available at lower levels in the tree, and then by going up to the next parent node and starting to look for candidates on the other descending branch.

This part of the clustering process was performed by using specific software developed by the authors.

Model

The analytical autoregressive TD animal model was

$$\mathbf{y}_{ijkLmn} = \mathbf{HTD}_i + \mathbf{Age}(\mathbf{H})_j + \mathbf{DIM}(\mathbf{H})_{k(L)} + \mathbf{a}_n + \mathbf{p}_{m(L)} + \mathbf{t}_{n(mL)} + \mathbf{e}_{ijkLmn},$$

where $\mathbf{y}_{ijkLmn}$ is the vector of TD yield, $\mathbf{HTD}_i$ is the vector of the fixed effect of herd test date, $\mathbf{Age}(\mathbf{H})_j$ is the vector of the fixed effect of age at calving nested

Table 2. (Co)variance components (kg²) and genetic parameters for daily milk yield of Portuguese Holstein cattle

Parameter	Component
Genetic variance	7.428
Error variance	2.966
LTE ¹ variance	<0.001
LTE autocorrelation	<0.001
STE ² variance (Lact 1)	11.004
STE autocorrelation (Lact 1)	0.819
STE variance (Lact 2)	22.231
STE autocorrelation (Lact 2)	0.831
STE variance (Lact 3)	26.044
STE autocorrelation (Lact 3)	0.816
Heritability (Lact 1)	0.347
Heritability (Lact 2)	0.228
Heritability (Lact 3)	0.204

¹LTE = long-term environmental effect, accounting for the autocorrelation generated by the cow across lactations.

²STE = short-term environmental effect, accounting for the autocorrelations attributable to the repeating effect of cow between test days for each lactation.

within herd (**H**), **DIM(H)**_{k(L)} is the vector of the fixed effect of DIM nested within herd and lactation number (**L**), **a_n** is the vector of the random additive genetic effect of the animal, **p_{m(L)}** is the vector of the random long-term environmental effect accounting for the correlations generated by the cow across lactations, **t_{n(mL)}** is the vector of the random short-term environmental effect accounting for the correlations attributable to the cow between TD within the same lactation, and **e** is the vector of the random residual effect. The effects of **p** and **t** are fitted with first-order autocorrelation structures (Carvalho et al., 1998, 2002). The same (co)variance components and autocorrelations were used in all evaluations (Table 2).

Comparison of Clustering Procedures

Results from all data sets were compared with the current clustering procedure used in the Portuguese genetic evaluations (data set C), where only herds with at least one CG with less than 3 observations were candidates to form new clusters. Management similarities among herds were based only on the phenotypic daily mean and standard deviation at first lactation.

Comparisons were based on differences and correlations among EBV, ranking, PEV, and accuracies of prediction (**ACC**), with the latter 2 computed only for sires with 10 or more daughters (1,962 sires). Prediction error variance corresponds to exact estimators extracted from the diagonal of the inverse of the coefficient matrix, and the **ACC** was obtained as

$$ACC_i = \sqrt{1 - \frac{PEV_i}{\hat{\sigma}_a^2 (1 + F_i)}}$$

where **PEV_i** and **F_i** [Au: should these be roman and boldface, or italic and lightface (as in the equation)?] are, respectively, the prediction error variance and the inbreeding coefficient for sire **i**, [Au: see note above] and $\hat{\sigma}_a^2$ is the estimated additive genetic variance (Table 2). Between- and within-CG variances were also used to compare alternative definitions of herd groups. These variance components were obtained by using the model

$$y_{ij}^* = CG_i + e_{ij},$$

where y_{ij}^* is the vector of observation corrected by the corresponding BLUE [Au: Please define] and BLUP solutions for other than the CG effects (Carabaño et al., 2004). Corresponding coefficients of determination were also calculated and used as measures of goodness of fit.

RESULTS AND DISCUSSION

The importance of the production variables and their variance in describing herd environmental variability was clearly represented in the first 3 PC, whereas the fourth and fifth PC showed greater loads of the management factors, such as HS and age at first calving. These findings are in agreement with several studies (Zwald et al., 2003; Windig et al., 2005) in which milk yield descriptors were, in general, the most important environmental variables. As expected, the final number of comparison herd groups decreased with the increase in the minimum number of observations required per CG. The criteria used to form data sets for group B also contributed to increasing the relative final number of herd groups and the total number of CG for the analysis (Table 3). Although data set C had a high total number of CG, the number of clusters was only 14 (with a mean number of 37.6 ± 19.7 herds per cluster) compared with 74 clusters in B₃ (7.1 ± 8.4 herds per cluster). Considering that one of the objectives of this study was to retain the HTD structure to the extent possible (year-months within each herd), data sets A₃ and B₃ were the ones that best preserved the original database structure, especially data set B₃, in which herd groups were maximized, resulting in the data set with the largest number of CG (Table 3).

Table 4 has the correlations for EBV, rank, **ACC**, and **PEV** among data set C and all other data sets. Estimated breeding value correlations were generally high, ranging from 0.85 to 0.97. The rank correlation followed the same pattern, with data set B₃ having the stronger relationship with data set C. The same types of relationships were found within sex (0.98 for males

Table 3. Percentage of herds clustered, corresponding to the total number of herd groups (including clusters) and the contemporary group (CG) structure for the data sets

Data set ¹	Herds clustered, ² %	Total herd groups, n	CG, n	Average (min; max) observations in CG			
				Full data set		Only in clusters	
A ₃	59.7	761	93,248	42.8	(3; 863)	43.3	(3; 364)
A ₅	80.2	536	67,739	59.9	(5; 863)	64.9	(5; 782)
A ₁₀	97.4	191	25,987	153.5	(10; 1134)	162.0	(10; 1134)
A ₁₅	99.9	80	11,269	354.0	(17; 2180)	557.8	(17; 2180)
B ₃	39.6	877	100,699	39.6	(3; 863)	64.0	(3; 668)
B ₅	57.4	665	80,929	49.3	(5; 963)	93.8	(5; 963)
B ₁₀	87.1	273	35,590	112.1	(10; 1924)	183.1	(10; 1924)
B ₁₅	96.8	110	14,563	273.9	(15; 897)	374.1	(17; 897)
C	39.6	813	93,815	42.5	(3; 963)	289.8	(4; 963)

¹In data sets A and B, management similarities among herds were based on the distance matrix using 5 principal components obtained from 14 descriptive variables, and CG had to have at least 3, 5, 10, and 15 observations. In data set A, all herds were candidates to form clusters, whereas in data set B, only those without the minimum number of observations in any CG were used in the clustering process. In data set C, only herds with less than 3 observations per CG were candidates to form new clusters, and management similarities among herds were based only on the phenotypic daily mean and SD at first lactation.

²Percentages from a total of 1,330 herds.

and 0.97 for females for data set B₃). High EBV and rank correlations were also reported in other studies evaluating the effect of size of CG (Reents et al., 1995; Carabaño et al., 2004). Varying the size of the CG from 3 to 15 observations by using the definitions of management similarity in the present study had only a minor effect on EBV and ranking of the animals. This fact may facilitate the implementation of new clustering approaches, especially those that provide a better partitioning of the CG for herd management purposes. With the exception of data sets A₃ and A₅, the ACC and PEV correlations for the selected sires were also high and of similar magnitude. The high correlation coefficients

found in this study indicate that only a slight change in the rank of sires and cows may be expected (especially between data sets C and B₃).

The high percentage of elite animals (top 100) in common between C and the other data sets (Table 5) was related to the high correlations obtained in this study. Again, data set B₃ was the one showing more similarities with data set C, indicating that this clustering methodology will be the one with less impact in reranking sires and cows in future evaluations.

Response to selection during the last decade was also compared in this study (Figure 1 [Au: No Figure 2 was included in the manuscript, and the callout was deleted.]). Except for males in A₁₅, trends were

Table 4. Correlation coefficients for EBV, rank, accuracy of prediction (ACC), and prediction error variance (PEV) of data set C¹ with the other 8 data sets studied

Data set ²	EBV	Rank	ACC ³	PEV ³
A ₃	0.97	0.96	0.56	0.50
A ₅	0.95	0.94	0.65	0.58
A ₁₀	0.89	0.88	0.86	0.84
A ₁₅	0.85	0.84	0.87	0.84
B ₃	0.97	0.97	0.85	0.82
B ₅	0.96	0.95	0.85	0.83
B ₁₀	0.91	0.90	0.77	0.73
B ₁₅	0.88	0.86	0.87	0.84

¹In data set C, only herds with less than 3 observations per contemporary group (CG) were candidates to form new clusters, and management similarities among herds were based only on the phenotypic daily mean and SD at first lactation.

²In data sets A and B, management similarities among herds were based on the distance matrix using 5 principal components obtained from 14 descriptive variables, and CG had to have at least 3, 5, 10, and 15 observations. In data set A, all herds were candidates to form clusters, whereas in data set B, only those without the minimum number of observations in any CG were used in the clustering process.

³ACC and PEV refer only to males with 10 or more daughters.

Table 5. Percentage of animals in common of the top 100 elite animals among data set C¹ and all the other 8 data sets studied (sires with 10 or more daughters)

Data set ²	Sires	Cows
A ₃	84	82
A ₅	77	57
A ₁₀	56	39
A ₁₅	54	35
B ₃	89	86
B ₅	84	54
B ₁₀	66	42
B ₁₅	58	38

¹In data set C, only herds with less than 3 observations per contemporary group (CG) were candidates to form new clusters, and management similarities among herds were based only on the phenotypic daily mean and SD at first lactation.

²In data sets A and B, management similarities among herds were based on the distance matrix using 5 principal components obtained from 14 descriptive variables, and CG had to have at least 3, 5, 10, and 15 observations. In data set A, all herds were candidates to form clusters, whereas in data set B, only those without the minimum number of observations in any CG were used in the clustering process.

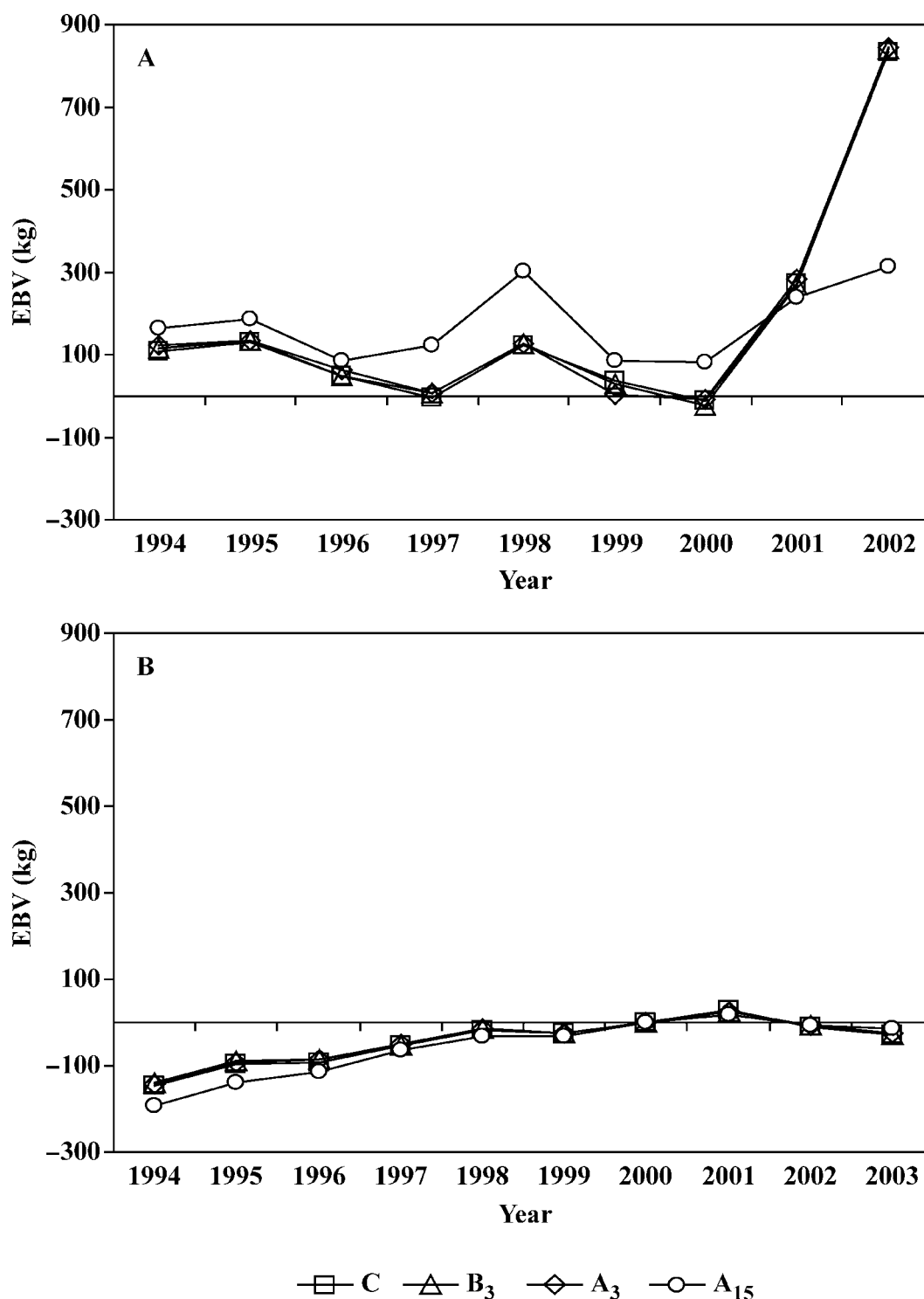


Figure 1. Milk genetic progress for sires (A) with 10 or more daughters and dams (B) by year of birth in data sets C, B₃, A₃, and A₁₅. In data set C, only herds with less than 3 observations per contemporary group (CG) were candidates to form new clusters, and management similarities among herds were based only on the phenotypic daily mean and SD at first lactation. In data sets A and B, management similarities among herds were based on the distance matrix using 5 principal components obtained from 14 descriptive variables. The CG had to have at least 3 (B₃ and A₃) or 15 (A₁₅) observations. In A₃ and A₁₅, all herds were candidates to form clusters, whereas in B₃, only those without at least 3 observations in any CG were used in the clustering process.

Table 6. Between (σ_b^2) and within (σ_w^2) contemporary group (CG) variances (kg^2) for the adjusted test days and the coefficients of determination (R^2) for the 9 data sets

Data set ¹	σ_b^2	σ_w^2	R^2
A ₃	26.546	1.381	0.9517
A ₅	27.128	1.415	0.9513
A ₁₀	23.482	1.508	0.9400
A ₁₅	22.476	1.549	0.9357
B ₃	25.441	1.047	0.9615
B ₅	25.091	1.397	0.9483
B ₁₀	26.307	1.475	0.9474
B ₁₅	20.188	1.531	0.9298
C	30.499	1.391	0.9574

¹In data set C, only herds with less than 3 observations per CG were candidates to form new clusters, and management similarities among herds were based only on the phenotypic daily mean and SD at first lactation. In data sets A and B, management similarities among herds were based on the distance matrix using 5 principal components obtained from 14 descriptive variables, and CG had to have at least 3, 5, 10, and 15 observations. In data set A, all herds were candidates to form clusters, whereas in B, only those without the minimum number of observations in any CG were used in the clustering process.

very similar for both sexes, confirming that only minor changes are to be expected if the new clustering methods are to be adopted.

The average PEV and ACC of data set C was significantly different ($P < 0.05$, data not shown) from the PEV and ACC of the other data sets. The difference was minimal, and its significance may be related to the large number of degrees of freedom used in the comparisons. Except for A₃ and A₅, the PEV (ACC) of C was larger (smaller) than for the other data sets. The variance components used in the analytical model were fixed for all data sets, implying that the differences in ACC were caused by the differences in PEV, influenced by the size of the CG and the data structure (Carabaño et al., 2004). Schmitz et al. (1991) concluded that ACC of a sire evaluation is more dependent on the number of observations per CG than on the procedure for grouping them. Ugarte et al. (1992) found that increasing the size of CG decreased the PEV in all models studied. In the present study, clustering herds to increase the size of CG also led to smaller PEV (higher ACC) by increasing the representation of a particular sire in a particular CG.

In the present study, the within-CG variance for the adjusted TD from data set C was larger than the one obtained for data set B₃ (Table 6). A better fit of the CG should minimize the within-CG variance and maximize the variance among CG levels. Even though the difference is small, the clustering methodology used in data set B₃ provided a larger R^2 , which may also be interpreted as a measure of a better fit for the data (Table 6).

In future research, we will study the impact that herds changing clusters between consecutive evalua-

tions might have in selection programs. The consequences of considering the systematic environmental effects as fixed, random, or both, with or without a covariance structure, will also be investigated.

CONCLUSIONS

Using PC analysis to reduce the number of descriptive environmental variables to define management similarities was a viable approach to cluster herds with insufficient contemporaries per level of fixed effects in the analytical model. Management similarities among herds were based on the distance (correlation) matrix, using 5 PC obtained from 14 descriptive variables. A larger number of clusters permits more comparison groups and was one of the objectives of this study, aiming for the best possible contrast among daughters, which would result in faster genetic progress per year.

Among all clustering strategies studied, data set B₃ (CG had to have at least 3 records, and clustering was only among herds without the minimum number of observations in any CG) was the one giving the better fit for the data with smaller changes in the original herd structure of the database. The smaller within-CG variance and the larger R^2 associated with consistently higher correlations among the parameters studied suggest that this clustering procedure may be the choice for future dairy cattle genetic analysis in Portugal.

ACKNOWLEDGMENTS

The authors thank A. Martins and A. Ferreira from the Portuguese Dairy Cattle Breeders Association [Au: Please add location] for providing the data set. J. Vasconcelos received financial support from Fundação para a Ciência e a Tecnologia (SFRH/BD/11397/2002) [Au: Please add location].

REFERENCES

- Carabaño, M. J., M. T. Fernández, C. Díaz, and M. Serrano. 2006. Comparing alternative models to account for herd effects in test day milk yield data. CD-ROM communication 25-09 in Proc. 8th World Congr. Gen. Appl. Livest. Prod., Belo Horizonte, Brazil. [Au: Please provide name of editor and name and location of publisher] C
- Carabaño, M. J., A. Moreno, P. López-Romero, and C. Díaz. 2004. Comparing alternative definitions of the contemporary group effect in Avileña Negra Ibérica beef cattle using classical and Bayesian criteria. *J. Anim. Sci.* 82:3447–3457.
- Carvalho, J., R. W. Blake, E. J. Pollak, R. L. Quaas, and C. V. Duran-Castro. 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in US Holstein herds. *J. Dairy Sci.* 81:2738–2751.
- Carvalho, J., E. J. Pollak, R. L. Quaas, and R. W. Blake. 2002. An autoregressive repeatability animal model for test-day records in multiple lactations. *J. Dairy Sci.* 85:2040–2045.
- Crump, R. E., N. R. Wray, R. Thompson, and G. Simm. 1997. Assigning pedigree beef performance records to contemporary

- groups taking account of within-herd calving patterns. *Anim. Sci.* 65:193–198.
- International Committee for Animal Recording. 1995. Recording guidelines. Appendices to the information agreement of recording practices (milk recording). Int. Comm. for Anim. Rec., Rome, Italy.
- Meyer, K. 1987. Estimates of variances due to sire $\times$ herd interactions and environmental covariances between paternal half-sibs for first lactation dairy production. *Livest. Prod. Sci.* 17:95–115.
- Raffrenato, E., R. W. Blake, P. A. Oltenacu, J. Carvalheira, and G. Licitra. 2003. Genotype by environment interaction for yield and somatic cell score with alternative environmental definitions. *J. Dairy Sci.* 86:2470–2479.
- Reents, R., J. C. M. Dekkers, and L. R. Schaeffer. 1995. Genetic evaluation for somatic cell score with a test day model for multiple lactations. *J. Dairy Sci.* 78:2858–2870.
- SAS Institute. 1989. SAS/STAT User's Guide, Version 6. 4th ed. Vol. 1. SAS Inst. Inc., Cary, NC.
- Schaeffer, L. 1987. Estimation of variance components under a selection model. *J. Dairy Sci.* 70:661–671.
- Schmitz, F., R. W. Everett, and R. L. Quaas. 1991. Herd-year-season clustering. *J. Dairy Sci.* 74:629–636.
- Strabel, T., and T. Szwachkowski. 1999. The use of test day models with small size of contemporary groups. *J. Anim. Breed. Genet.* 116:379–386.
- Strabel, T., J. Szyda, E. Ptak, and J. Jamrozik. 2005. Comparison of random regression test-day models for Polish Black and White cattle. *J. Dairy Sci.* 88:3688–3699.
- Swalve, H. H. 1995. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *J. Dairy Sci.* 78:929–938.
- Tierney, J. S., and L. R. Schaeffer. 1994. Inclusion of semen price of the sire in an animal model to account for preferential treatment. *J. Dairy Sci.* 77:576–582.
- Ugarte, E., R. Alenda, and M. J. Carabaño. 1992. Fixed or random contemporary groups in genetic evaluations. *J. Dairy Sci.* 75:269–278.
- Van Bebber, J., N. Reinsch, W. Junge, and E. Kalm. 1997. Accounting for herd, year and season effects in genetic evaluations of dairy cattle: A review. *Livest. Prod. Sci.* 51:191–203.
- Van Vleck, L. 1987. Contemporary groups for genetic evaluations. *J. Dairy Sci.* 70:2456–2464.
- Vasconcelos, J., A. Martins, A. Ferreira, and J. Carvalheira. 2005. Effects of the exclusion of small herds from genetic evaluations of dairy cattle in Portugal. *Revista Portuguesa de Zootecnia, Ano XII* 2:105–117.
- Vasconcelos, J., F. Santos, R. Barroso, A. Martins, A. Ferreira, and J. Carvalheira. 2006. Effects of clustering dairy herds for genetic evaluations using different descriptors to define similarities between production environments. CD-ROM communication 24-30 in Proc. 8th World Cong. Gen. Appl. Livest. Prod., Belo Horizonte, Brazil. **[Au: Please provide name of editor and name and location of publisher]**
- Visscher, P. M., and M. E. Goddard. 1993. Fixed and random contemporary groups. *J. Dairy Sci.* 76:1444–1454.
- Wade, K. M., and R. L. Quaas. 1993. Solutions to a system of equations involving a first-order autoregressive process. *J. Dairy Sci.* 76:3026–3031.
- Windig, J. J., M. P. L. Calus, and R. F. Veerkamp. 2005. Influence of herd environment on health and fertility and their relationship with milk production. *J. Dairy Sci.* 88:335–347.
- Zwald, N. R., K. A. Weigel, W. F. Fikse, and R. Rekaya. 2003. Application of a multiple-trait herd cluster model for genetic evaluation of dairy sires from seventeen countries. *J. Dairy Sci.* 86:376–382.

Artigo 5*

Evaluation of the Impact of Clustering Herds with Small Size Contemporary Groups on
Consecutive Genetic Evaluations of Dairy Cattle

*Avaliação do Impacto da Aglomeração de Rebanhos com Grupos Contemporâneos
Pequenos em Avaliações Genéticas Consecutivas*

*Vasconcelos, J., A. Bagnato, A. Pereira, and J. Carvalheira. Evaluation of the Impact of Clustering with Small Size Contemporary Groups on Consecutive Genetic Evaluations of Dairy Cattle. Unpublished paper.

Avaliação do Impacto da Aglomeração de Rebanhos com Grupos Contemporâneos Pequenos em Avaliações Genéticas Consecutivas

RESUMO

Em avaliações genéticas de bovinos leiteiros os grupos contemporâneos (**GC**) de animais são habitualmente definidos pelas classes do efeito rebanho-ano-mês (**RAM**). Assumir este efeito como fixo no modelo implica a existência de um número mínimo de observações em cada classe. A formação de “clusters”, seguindo a metodologia que utiliza apenas explorações com pelo menos um GC com menos de 3 observações no processo de aglomeração, soluciona esta limitação. No entanto, tornou-se necessário avaliar o impacto que a possibilidade das explorações mudarem de “clusters” em avaliações sucessivas teria no valor genético aditivo e ranking dos animais. Foram usados aproximadamente 4 milhões de contrastes produzidos por 250.390 vacas Holstein, em 1344 explorações para construir 4 bases de dados representando 4 avaliações consecutivas, uma por quadrimestre. Após o processo de aglomeração, foi aplicado um modelo para contrastes auto-regressivo que considerou o RAM como efeito fixo no modelo e os seus resultados comparados. As elevadas correlações obtidas entre os valores genéticos aditivos estimados em avaliações consecutivas, quer para o conjunto de todos os animais, quer separadamente entre sexos ($> 0,97$), indicou a inexistência de grandes alterações. Mais de 83% dos touros da lista do “top 100” mantiveram-se de uma avaliação para a outra. Apesar das correlações entre médias genéticas das explorações e respectivo ranking não serem tão elevadas ($> 0,64$), os seus valores não foram extremos, podendo ser justificados pelas práticas de manejo usualmente aplicadas nas explorações em Portugal. Estes resultados sugerem que este método de formação de “clusters” de explorações com base na semelhança dos seus níveis de manejo poderá ser aplicado na avaliação genética de bovinos leiteiros em Portugal.

Evaluation of the Impact of Clustering Herds with Small Size Contemporary Groups on Consecutive Genetic Evaluations of Dairy Cattle

J. Vasconcelos^{*,†}, A. Bagnato[‡], A. Pereira^{*}, and J. Carvalheira^{*,†}

^{*}CIBIO-ICETA, Research Center in Biodiversity and Genetic Resources - University of Porto, Rua Padre Armando Quintas-Crasto, 4485-661 Vairão, Portugal

[†]ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Portugal

[‡]Department of Veterinary Science and Technology for Food Safety, Faculty of Veterinary Medicine, University of Milan, Italy

Julio Carvalheira

Campus Agrário de Vairão,

Rua Padre Armando Quintas – Crasto, 4485-661 Vairão

PORTUGAL

Phone: 351-252 660400, FAX: 351-252 661780

E-mail: jgc3@mail.icav.up.pt

ABSTRACT

Contemporary groups (**CG**) in genetic evaluations of dairy cattle are usually defined as classes of the herd-test-date (**HTD**) effect. Assuming this effect as fixed in the analytical model, requires the definition of a minimum number of observations at each class. A methodology to cluster herds with less than 3 records per CG was used to fulfill this requirement. The impact of herds changing clusters between consecutive evaluations on the animals EBV and respective ranking was evaluated. Nearly 4 million test-day records produced by 250,390 Portuguese Holstein cows in 1,344 herds were used to build 4 data sets representing 4 consecutive genetic evaluations taken 4 months apart. After the clustering process, all data sets were analyzed by an autoregressive test-day animal model fitting a fixed HTD in a multiple lactation setting. EBV and rank correlations for all animals and within sex were high (> 0.97) indicating that the clustering procedure did not induce major changes between evaluations. More than 83% of the top 100 sires were in common from one evaluation to the next. Correlations among herd genetic means and rank were lower than those among animals (> 0.64). These figures were considered within the expectations and may be explained by the usual farm management dynamics occurring in the Portuguese dairy cattle population. These results suggest that the methodology used to cluster herds based on management similarities may be applied in future dairy cattle genetic evaluations.

(Key Words: contemporary group, EBV re-ranking, cluster analysis, autoregressive test-day model)

Defining as contemporary groups (**CG**) the herd-test-date (**HTD**) effect in test-day (**TD**) models used in dairy cattle genetic evaluations has many advantages (Ptak and Schaeffer, 1993). Fitting this effect as random or fixed or a combination of both is a matter of balance between accuracy and bias (Van Vleck 1987; Visscher and Goddard, 1993; Van Bebber et al., 1997). In Portugal, the risk of nonrandom associations between animals and effects considered fixed is high, especially because top sires are preferentially used in high production herds. In these cases, assuming the CG as a fixed effect in the analytical model may minimize prediction bias (Meyer 1987; Schaeffer, 1987; Van Vleck 1987; Visscher and Goddard 1993; Tierney and Schaeffer, 1994; Van Bebber et al., 1997). However, fitting HTD as a fixed effect requires a minimum number of observations per CG in order to simultaneously maximize the effective number of observations and minimize the residual error and prediction error variance.

Using principal components (**PC**) analysis to reduce the number of descriptive environmental variables to define management similarities, Vasconcelos et al. (2008) developed a methodology to cluster herds with insufficient contemporaries per level of fixed effects in the analytical model. The assumption was that dairy farms with similar descriptive environmental variables should have similar management levels, minimizing in this manner the impact that any change in the environmental and genetic structure of the data may bring to the genetic evaluations. In that study the authors evaluated the consequences of clustering those herds by varying the minimal number of observations per CG, using either all herds in the data set as potential candidates, or just herds without the minimum number of observations in any CG. Comparisons with the current clustering procedure used in the Portuguese genetic evaluations were based on differences and correlations between EBV, ranking, PEV and accuracies of prediction (**ACC**). The clustering strategy that defined a minimum of 3 records per CG, with the exclusive participation of herds with CG lacking that number, was the one that gave the better fit for the data with minimum changes in the original herd structure of the database (Vasconcelos et al., J. Dairy Sci., accepted).

The major implication of adopting this herd clustering procedure is related with the possibility of observing unexpected changes in the predicted parameters between consecutive genetic evaluations. These changes may be due to herds that enter or leave a cluster between evaluations, either because they register important changes in management level (HTD means) and/or changes in their herd structure (more or less animals in different ages and production groups). Because in genetic evaluations, animal performances are compared within CG, changes in their composition may result in changes in animal and herd ranking with consequences for the industry. Significant variations could result in oscillations of the economic market value of the animals,

decreasing the confidence level of animal breeders on the genetic evaluations (Strabel and Szwaczkowski, 1999; Lin and Togashi, 2002). Therefore, the aim of the present work was to study the impact of herds changing clusters between consecutive evaluations on the animals EBV and respective ranking and determine the consequences for the herd management.

The data was provided by the Portuguese Dairy Cattle Breeders Association and consisted of TD yields of milk from Holstein cows recorded from June 1994 to December 2006. Monthly records were from a supervised recording plan (morning and afternoon milking according to the standard plan A4 of ICAR) from cows in the first 3 parities. Lactations were required to have at least 2 TD records. The TD records were validated when the first reported TD did not exceed 75 DIM, and time interval between consecutive TD was between 26 and 33 days, for the interval between 5 and 305 DIM. Further edits followed ICAR (1995) rules for dairy cows TD recording. After editing, the data comprised 4,244,449 TD records, produced by 250,390 cows in 1,344 herds.

Four data sets were built to represent 4 consecutive national genetic evaluations, taken 4 months apart, from December 2005 to December 2006 (evaluations **A**, **B**, **C** and **D** respectively). The data sets are described in Table 1 showing the trends in number of records, herds and clusters during the period of study. At each evaluation, fourteen herd production and environmental variables obtained from the respective data set were used in a PC analysis to determine the most influencing PC vectors (explaining more than 85% of the total variance) and then used to build a distance (correlation) matrix based on the method of Ward's minimum-variance (PROC CLUSTER; SAS, 1989). Herds were then sorted according to those distances so that adjacent herds in the list represented the closest ones in terms of similarities in management. The clustering process started by joining the closest herds, one at the time. After each agglomeration (herd-herd or herd-cluster), the program recounted all observations at each CG and stopped the process for that specific cluster if all CG had the required minimum of 3 records (see details in Vasconcelos et al., J. Dairy Sci., accepted).

After the clustering process, the data sets were analyzed using the autoregressive test day animal model which fits multiple lactation data and defines the systematic environmental effects using 3 fixed effects: herd-test-dates, age at calving within herd and DIM within herd and lactation (Carvalho et al., 2002):

$$y_{ijkLmn} = \text{HTD}_i + \text{Age}(\mathbf{H})_j + \text{DIM}(\mathbf{H})_{k(L)} + \mathbf{a}_n + \mathbf{p}_{m(L)} + \mathbf{t}_{n(mL)} + \mathbf{e}_{ijkLmn},$$

where, $\mathbf{y}$ is the vector of TD yield, HTD is the vector of fixed effect of herd-test-date, $\text{Age}(\mathbf{H})$ is the vector of fixed effect of age at calving nested within herd ($\mathbf{H}$), $\text{DIM}(\mathbf{H})$ is the

vector of fixed effect of DIM nested within herd and lactation number (**L**), **a** is the vector of random additive genetic effect of the animal, **p** is the vector of random long term environmental effect accounting for the correlations generated by the cow across lactations, **t** is the vector of random short term environmental effect accounting for the correlations due to the cow between TD within the same lactation, and **e** is the vector of random residual effect. The effects of **p** and **t** are fitted with first-order autocorrelation structures (Carvalho et al., 1998, 2002). The same (co)variance components and autocorrelations were used in all evaluations.

Results from the 4 consecutive evaluations were compared based on correlations between EBV and animal ranks (Table 2). Although some changes are expected (some cows are new, others dry off, the number of records increase, herds may move from one cluster to another, etc.), they should not be large enough to cause extreme modifications to the rank structure of the population. EBV and rank correlations for all animals between evaluations were high (> 0.97) indicating that the clustering procedure did not induce major changes. Correlations within sex were also high (> 0.97), confirming this expectation. If the cow's rank is taken within herd, the correlation coefficients approach unity (Within farm female rank, Table 2). Changes also occurred in the elite animals. Comparing the top 100 sires (with 10 or more daughters) between evaluations B and A, C and B and D and C, there were 86%, 86% and 83% of sires in common, respectively. These figures are not extremes and may be explained by the normal farm management dynamics occurring in our dairy cattle population.

Genetic means from all herds were also calculated based on the EBV of their animals (with production records). Herd genetic means and rank correlations between evaluations were lower than those among animals (Table 2). Several reasons may be found to explain this higher susceptibility of herd genetic means to vary in time, besides changing clusters. For example, it is a common practice among dairy cattle breeders in Portugal to import pregnant heifers from around Europe. In general, these animals come from highly selected pedigrees, with a non-negligible positive influence on the genetic mean of the importing herd. This implies periodic changes in the ranking relationship among herds.

The direct influence of the clustering effect in the parameters studied is difficult to quantify because of the many variables involved. Nevertheless, in the author's opinion, the strict criteria used to define management similarities to cluster herds should minimize this effect and the methodology here presented may be used in future dairy cattle genetic evaluations. However, some further developments can be made. The criteria to cluster herds are very rigorous especially when only one CG in the herd has less than 3 observations. Another reason is that not all herds have observations throughout the

whole period of data collection and that farmers change management practices over time. Although increasing the complexity of the analytical models, future research should study the consequences of considering the systematic environmental effects as fixed and random. An approach that combines fixed herd-years and random HTD effects with a simple correlation structure may minimize the prediction bias and eliminates the need for clustering herds with small size CG.

ACKNOWLEDGEMENTS

The authors thank A. Martins and A. Ferreira from the Portuguese Dairy Cattle Breeders Association for providing the data set. J. Vasconcelos receives financial support from Fundação para a Ciência e a Tecnologia (FCT).

REFERENCES

- Carvalho, J., R.W. Blake, E.J. Pollak, R.L. Quaas, and C.V. Duran-Castro. 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in US Holstein herds. *J. Dairy Sci.* 81:2738-2751.
- Carvalho, J., E. J. Pollak, R. L. Quaas and R. W. Blake. 2002. An autoregressive repeatability animal model for test-day records in multiple lactations. *J. Dairy Sci.* 85:2040-2045.
- ICAR, International Committee for Animal Recording. 1995. Recording Guidelines. Appendices to the information agreement of recording practices (milk recording). Rome, Italy.
- Lin, C. Y., and K. Togashi. 2002. Genetic improvement in the presence of genotype by environment interaction. *Anim. Sci. J.* 73:3-11.
- Meyer, K. 1987. Estimates of variances due to sire $\times$ herd interactions and environmental covariances between paternal half-sibs for first lactation dairy production. *Livestock Prod. Sci.* 17:95-115.
- SAS/STAT® User's Guide, Version 6, Volume 1. 1989. SAS Institute Inc.
- Schaeffer, L. 1987. Estimation of variance components under a selection model. *J. Dairy Sci.* 70:661-671.
- Strabel, T., and T. Szwaczkowski. 1999. The use of test day models with small size of contemporary groups. *Journal of Animal Breeding and Genetics* 116(5):379-386.
- Tierney, T. S., and L. R. Schaeffer. 1994. Inclusion of semen price of the sire in an animal model to account for preferential treatment. *J. Dairy Sci.* 77:576-582.
- Ugarte, E., R. Alenda, M. J. Carabaño. 1992. Fixed or random contemporary groups in genetic evaluations. *J. Dairy Sci.* 75:269-278.

- Van Bebber, J., N. Reinsch, W. Junge, E. Kalm. 1997. Accounting for herd, year and season effects in genetic evaluation of dairy cattle: a review. *Livestock Prod. Sci.* 51:191-203.
- Van Vleck, L. 1987. Contemporary groups for genetic evaluations. *J. Dairy Sci.* 70:2456-2464.
- Vasconcelos, J., F. Santos, R. Barroso, A. Martins, A. Ferreira and J. Carvalheira. 2006. Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil 24-30.
- Vasconcelos, J., A. Bagnato, A. Pereira, and J. Carvalheira. 2008. Effects of clustering herds in dairy cattle genetic evaluations. 91:377-384.
- Visscher, P. M. and M. E. Goddard. 1993. Fixed and random contemporary groups. *J. Dairy Sci.* 76:1444-1454.

Table 1. Descriptive statistics and structure of the data used in 4 consecutive genetic evaluations from December 2005 to December 2006. Herds with less than 3 records in any contemporary group (CG) were clustered¹

	Evaluations ²			
	A	B	C	D
TD records, no.	3,898,867	4,019,138	4,135,680	4,244,449
Cows, no.	233,302	239,676	244,941	250,390
Sires, no.	1997	2079	2133	2094
Herds, no.	1327	1335	1342	1344
Clusters, no.	72	71	55	67
Herds in Clusters, no.	527	526	534	540
1 st Lactation				
Average records/cow, no.	8.6	8.6	8.7	8.7
Mean daily yield, kg	24.68	24.80	24.89	24.94
SD, kg	6.45	6.51	6.53	6.53
2 nd Lactation				
Average records/cow, no.	8.5	8.5	8.5	8.5
Mean daily yield, kg	27.25	27.39	27.50	27.56
SD, kg	8.67	8.72	7.74	8.74
3 rd Lactation				
Average records/cow, no.	8.4	8.4	8.4	8.4
Mean daily yield, kg	27.79	27.94	28.05	28.12
SD, kg	9.03	9.09	9.11	9.12

¹Herds with less than 3 records in any CG were clustered after sorting according to the distance matrix calculated using the Ward's minimum-variance methodology. Adjacent herds in the list represented the closest ones in terms of similarities in management. The clustering process started by joining the closest herds, one at the time. After each agglomeration (herd-herd or herd-cluster), the program recounted all observations at each CG and stopped the process for that specific cluster if all CG had the required minimum of 3 records.

²A = December 2005; B = April 2006; C = August 2006; D = December 2006.

Table 2. Milk EBV and rank correlations obtained from 4 consecutive genetic evaluations using data corresponding to evaluations from December 2005 to December 2006. Herds with less than 3 records in any contemporary group (CG) were clustered¹

	Evaluations ²		
	A - B	B - C	C - D
All animals (EBV)	0.982	0.978	0.976
All animals rank	0.978	0.975	0.973
Females (EBV)	0.982	0.979	0.976
Females rank	0.980	0.976	0.974
Males (EBV)	0.976	0.975	0.974
Males rank	0.974	0.973	0.974
Within farm female rank	0.997	0.997	0.997
Herd genetic means	0.677	0.659	0.641
Herd rank	0.734	0.676	0.698

¹Herds with less than 3 records in any CG were clustered after sorting according to the distance matrix calculated using the Ward's minimum-variance methodology. Adjacent herds in the list represented the closest ones in terms of similarities in management. The clustering process started by joining the closest herds, one at the time. After each agglomeration (herd-herd or herd-cluster), the program recounted all observations at each CG and stopped the process for that specific cluster if all CG had the required minimum of 3 records.

²A = December 2005; B = April 2006; C = August 2006; D = December 2006.

4. CONCLUSÕES GERAIS

Neste capítulo serão apresentadas as principais conclusões do trabalho desenvolvido e dos artigos publicados, assim como de outros trabalhos desenvolvidos paralelamente no âmbito desta tese.

A primeira fase deste projecto de investigação consistiu num estudo sobre a extensão e projecção de lactações incompletas pelo MCAR (Carvalho et al., 2002b; Petim-Batista et al., 2002; Vasconcelos et al., 2002a; Vasconcelos et al., 2002b; Vasconcelos et al., 2004). Apesar das vantagens da utilização do valor diário da produção no dia do contraste em detrimento da produção padronizada aos 305 dias de lactação na estimativa dos valores genéticos já ter sido amplamente demonstrada, existem, contudo, situações em que o seu cálculo se torna necessário. Estas são utilizadas como ferramenta de apoio a decisões de manejo e em casos de limitação na capacidade de armazenamento e processamento de grandes quantidades de dados (Wiggans et al., 2002; Bormann et al., 2002; Vasconcelos e Carvalho, 2003). Em Portugal, a última estimativa dos valores genéticos aditivos (**VGA**) do efectivo leiteiro publicada a nível nacional foi baseada na produção aos 305 dias calculada pelo método clássico do intervalo entre testes (**MIT**), utilizando apenas lactações terminadas e com mais de 210 dias. A adopção de modelos de contrastes, ao definir o contraste como carácter de interesse, eliminou estas restrições e tornou possível, não só o incremento do número de observações (pelo menos dois contrastes por lactação), como também deu oportunidade à maioria dos animais de contribuírem para a caracterização genética da população. A possibilidade de incluir vacas múltiparas, sem registo de valores de produção da primeira lactação no processo de avaliação genética, adicionado à possibilidade da utilização de lactações incompletas ou em curso, representou um acréscimo de 23 a 30% de observações (Vasconcelos et al., 2004). A avaliação de animais com lactações ainda em curso possibilitou análises genéticas com dados muito recentes, reduzindo substancialmente o intervalo entre gerações. A adicionar a estes aspectos, há também a considerar o facto de os modelos de contrastes, ao incorporar estruturas de covariância auto-regressivas, optimizam os ajustes necessários à correcção do enviesamento que os efeitos temporais provocam na predição dos parâmetros genéticos.

A aplicação do MCAR aos dados nacionais demonstrou maior precisão e rigor na estimativa dos valores de contrastes em falta quando comparada com a avaliação obtida por modelos clássicos de lactações acumuladas aos 305 dias (Vasconcelos et al., 2004).

As diferenças entre lactações padronizadas de 305 dias reais (estimadas a partir de lactações completas pelo MIT) e estimadas pelo MCAR a partir de lactações incompletas foram insignificantes e menores do que as estimadas pelo MIT. Os elevados valores de correlação para lactações acumuladas aos 305 dias estimadas pelos dois métodos revelam uma elevada capacidade de predição do MCAR, cuja aplicação contribuiu significativamente para a diminuição do intervalo entre gerações (Vasconcelos et al., 2002a; Vasconcelos et al., 2002b; Vasconcelos e Carvalheira, 2003; Vasconcelos et al., 2004). As diferenças entre contrastes predictos pelo MCAR e contrastes reais foi reduzida e menor do que a obtida aplicando o MIT. As correlações entre valores estimados de produção acumulada aos 305 dias pelo MCAR e valores reais foram superiores a 0,5, mesmo nos casos de extensão da lactação a partir de apenas dois contrastes (Vasconcelos et al., 2002b; Vasconcelos et al., 2004). O incremento na precisão de predição dos valores genéticos dos reprodutores (Gama et al, 2000; Jamrozik et al, 2000; VanRaden, 1997) e a diminuição do intervalo entre gerações, são dois factores fundamentais para a maximização do progresso genético dos animais. Os modelos de contrastes contribuíram significativamente, positivamente para este fim. O acesso antecipado à informação da capacidade produtiva do efectivo poderá assim contribuir de forma decisiva para o aumento da rentabilidade económica das explorações. A fácil implementação do MCAR, sem necessidade de grandes recursos computacionais, possibilita aos produtores o cálculo das lactações acumuladas das suas fêmeas nas próprias explorações, o que representa uma valiosa ferramenta informativa para as decisões de manejo e gestão do efectivo (Vasconcelos e Carvalheira, 2003).

Tendo sido o objectivo principal desta tese o estudo das consequências da aplicação de um MCAR, tornou-se necessária uma análise prévia da base de dados do rebanho nacional de bovinos leiteiros. Esta consistiu na avaliação dos níveis de endogamia e depressão de endogamia dos animais constantes no pedigree nacional, um problema crescente nos programas de melhoramento de bovinos leiteiros (Vasconcelos et al., 2005b).

A análise do pedigree nacional de bovinos da raça Holstein revelou um coeficiente de endogamia (**CE**) médio e uma proporção de animais endogâmicos relativamente pequena quando comparada com outros países (Miglior et al., 1992; Wiggans e VanRaden, 1995; Young e Seykora, 1996; Boichard et al., 1997; Swalve et al., 2003). A proporção de animais endogâmicos apresentou uma tendência para o crescimento desde 1981, sendo maior nos últimos 10 anos, o que se pode considerar preocupante

pois aumenta a probabilidade da diminuição na performance de caracteres com importância económica, como a produção de leite, a sobrevivência e o desempenho reprodutivo, assim como o aumento do número de células somáticas no leite (Smith et al., 1998). Contudo, este aumento do número de animais endogâmicos não foi acompanhado pela subida do CE médio, como seria de esperar, o que pode estar relacionado com a utilização dos programas de emparelhamento (Vasconcelos et al., 2003), ou devido ao elevado número de animais com progenitores desconhecidos, o que implicou uma subestimativa dos CE (Vasconcelos et al., 2005b). Verificou-se uma tendência para a redução no número efectivo de animais fundadores nas últimas gerações, o que indica uma importante perda na variabilidade genética na população. Uma das consequências da utilização intensiva de touros na inseminação artificial é fazer com que um número relativamente pequeno de touros seja responsável pela manutenção de quase todo o efectivo (Miglior et al., 1992). Em Portugal, os 10 touros geneticamente mais influentes contribuíram com 21% para o “pool” de genes da população actual, o que poderá ter contribuído para um aumento do CE, ou seja, a probabilidade de um indivíduo herdar cópias do mesmo gene devido à presença de ancestrais comuns (Hartl e Clark, 1997). Apesar do efeito da depressão de endogamia obtido neste estudo não parecer causar reduções importantes na produção, o aumento intencional da endogamia deve ser evitado, pois os efeitos negativos no animal são globais e cumulativos, acarretando perdas económicas significativas principalmente devido à diminuição da expressão dos caracteres ligados ao “fitness” (reprodução e sobrevivência). A importância da qualidade do ficheiro de pedigree foi assim demonstrada, sendo urgente incluir a identificação dos touros disponíveis nos programas informáticos de emparelhamento. Tal irá prever os níveis de endogamia e o valor genético da descendência, instrumento fundamental de apoio aos programas de selecção e melhoramento, para assim poder maximizar o progresso genético e manter a diversidade genética.

Uma das possíveis consequências resultantes da aplicação de técnicas de formação de “clusters” é a perda de conexão genética entre explorações ao associar de forma sistemática touros e explorações. A utilização exclusiva (ou maioritária) de touros de menor qualidade genética em explorações de baixa média fenotípica e de touros de elevado potencial genético em explorações de média fenotípica superior poderá diminuir esta conexão genética entre os animais das várias explorações. O rigor de predição das avaliações genéticas depende, entre outros factores, do grau de conexão genética entre os grupos de manejo. Tornou-se assim necessário verificar se existia em Portugal uma utilização sistemática dos touros por região, induzindo à formação de subpopulações

isoladas, sem ligações genéticas entre si. Apesar de em Portugal serem realizadas avaliações genéticas periódicas, não existia uma quantificação da conexão genética dos dados, bem como da sua evolução ao longo do tempo.

Vasconcelos e Carvalheira (2004; 2005) concluíram que a não inclusão das explorações pequenas na avaliação genética provoca uma diminuição da conexão genética, logo do rigor de predição da mesma. A inclusão das explorações pequenas aglomeradas em grupos de manejo maiores não teve efeitos negativos na conexão genética da base de dados nacional, tornando-se assim numa solução viável para este problema. A utilização das técnicas de aglomeração permite desta forma a utilização de um maior número de animais na avaliação, aumentando o valor médio do relacionamento genético na população, e consequentemente o incremento do rigor de predição na avaliação do mérito genético aditivo dos animais.

Antes de avaliar em detalhe as consequências da aplicação de técnicas de formação de “clusters”, foi imperativo estudar o impacto que a alternativa à sua aplicação, ou seja, a eliminação de rebanhos pequenos pode ter nas avaliações genéticas. O objectivo dos estudos desenvolvidos por Vasconcelos et al. (2005a; 2005c) centrou-se na análise da variação de alguns parâmetros genéticos da população Holstein nacional, de acordo com a aplicação de diferentes critérios para a determinação do tamanho mínimo dos GC. Neste estudo foram simuladas 3 bases de dados (**BD**), a partir dos dados nacionais, onde as explorações pequenas (mínimo de 3, 10 ou 15 observações por nível de efeitos fixos no modelo) foram eliminadas progressivamente. O objectivo era evidenciar variações e possíveis tendências nos componentes de variância, rigor de predição dos valores genéticos e alterações no ranking dos animais que permitissem concluir sobre a viabilidade da utilização destas BD truncadas nas avaliações genéticas.

Verificou-se a ausência de significância estatística para as diferenças entre os componentes de variância obtidos nestas BD e os previamente estimados em Portugal usando o mesmo modelo animal (Carvalheira et al., 2002a). Apesar de esperado que quanto maior o número de contemporâneos em cada grupo, menor a variância do erro de predição e, portanto, maior o rigor de predição, verificou-se uma tendência decrescente à medida que aumentou a restrição do número mínimo de animais contemporâneos em cada GC de 3 para 15 observações. O peso da contribuição colectiva do número total de filhas (que diminui com a eliminação de explorações) foi determinante para a magnitude deste parâmetro. A eliminação de explorações pequenas pode ter, portanto, consequências negativas no melhoramento genético, diminuindo a

oportunidade de selecção de alguns reprodutores e a intensidade de selecção (Vasconcelos et al., 2005a; Vasconcelos et al., 2005c). A eliminação dos rebanhos mais pequenos pode ser também uma forma de tratamento preferencial no que diz respeito à selecção de reprodutores machos, pois de uma maneira geral, as explorações maiores investem mais na genética (Henderson, 1973; Schaeffer, 1991), sendo de esperar que o progresso genético anual nestes rebanhos seja proporcionalmente maior do que nos mais pequenos. A principal consequência é um incremento artificial do valor genético destes touros. Como esperado, a exclusão de rebanhos, mesmo que de pequenas dimensões, alterou o ranking dos VGA dos animais avaliados, com especial ênfase nos machos. É, portanto, necessário garantir um número mínimo de observações em cada GC, assegurando simultaneamente a maximização do número de filhas e de rebanhos com contribuição para a avaliação genética (Vasconcelos et al., 2005a; Vasconcelos et al., 2005c). A exclusão de rebanhos deve ser evitada e o tamanho dos GC deve ser garantido recorrendo a outros procedimentos. Agrupar rebanhos apresentou-se como uma solução para este problema, conclusão partilhada por outros autores (e.g. Schmitz et al., 1991; Strabel e Szwaczkowski, 1999; Carvalheira, 2000; Carabaño et al., 2004).

Depois de comprovada a necessidade de aglomeração de explorações com número insuficiente de observações por GC e de efectuado o estudo das consequências da sua aplicação, a etapa seguinte do projecto consistiu na estimativa dos parâmetros para a formação de “clusters”, nomeadamente o desenvolvimento de um método para a definição de semelhanças entre explorações e a influência de diferentes números mínimos de observações por GC (Vasconcelos et al., 2006; Vasconcelos et al., 2008). A definição de semelhanças entre explorações foi baseada numa análise de componentes principais aplicada às médias por exploração de variáveis consideradas como potenciais indicadores do nível produtivo e de manejo. Para um maior controlo do processo de aglomeração foram introduzidos novos parâmetros com o objectivo de minimizar o número de explorações por cluster, maximizando o número de GC final, assegurando, no entanto um mínimo de observações por nível de efeitos fixos. Após o processo de aglomeração, as BD foram analisadas usando MCAR e foram comparados os resultados dos VGA obtidos (ranking e progresso genético) bem como o seu rigor de predição. Comparando com a metodologia actualmente utilizada nas avaliações genéticas em Portugal em que apenas as explorações com pelo menos um GC com menos de 3 observações puderam participar no processo de aglomeração e formar “clusters”, foi a que obteve menores diferenças de VGA, assim como de maiores rigores de predição e menores variâncias dentro de CG, revelando um melhor ajustamento dos dados. Esta metodologia foi a que melhor preservou a estrutura original dos dados, maximizando o

número de grupos de comparação usados no processo de avaliação genética. Os elevados valores de correlação entre ranking de VGA obtidos sugeriram uma reduzida influência do processo de aglomeração na ordenação dos animais, tendo sido confirmado pela elevada percentagem de animais em comum na lista do “top100” dos VGA. A formação de mais “clusters” permitiu também a existência de um maior número de grupos de comparação e, portanto, um maior contraste entre filhas, revelando um maior progresso genético anual dos machos.

Após a definição da melhor metodologia a usar na aglomeração de explorações, antes da sua aplicação regular no programa de selecção nacional, surgiu a necessidade de estudar o impacto das mudanças de explorações de um cluster para o outro em avaliações consecutivas, quer a nível de alterações no ranking genético dos animais como do respectivo rigor de predição (Vasconcelos et al., dados não publicados). A metodologia aplicada na formação dos “clusters” foi a que obteve os resultados mais satisfatórios no trabalho anterior (Vasconcelos et al., 2006; Vasconcelos et al., J. Dairy Sci., 2008), em que apenas as explorações com pelo menos um GC com menos de três observações foram usadas no processo de aglomeração. Foram usados os dados nacionais para construir 4 BD que representassem 4 avaliações nacionais sucessivas quadrimestrais, de Dezembro de 2005 a Dezembro de 2006. Após o processo de aglomeração das explorações pequenas, foram efectuadas avaliações genéticas usando um MCAR, que considerou o RAM como efeito fixo no modelo e os seus resultados comparados. Os elevados valores de correlação obtidos para VGA estimados em avaliações consequentes, quer para o conjunto de todos os animais, quer separadamente entre sexos, indicou a não existência de grandes alterações no ranking dos animais. O ranking genético das fêmeas por exploração e o “top 100” genético também não evidenciaram grandes reordenações dos animais ao nível das explorações. Apesar dos valores para as correlações entre médias genéticas das explorações e respectivo ranking não serem tão elevadas ($> 0,64$), os seus valores não foram extremos, podendo ser justificados pela dinâmica e práticas de manejo usualmente aplicadas nas explorações em Portugal (Vasconcelos et al., dados não publicados). Por exemplo, a maioria dos produtores nacionais não faz a recria do seu efectivo, sendo prática corrente a introdução de grandes grupos de novilhas nas explorações, o que induz oscilações nas dimensões da exploração, média de idades do efectivo e coincidências na fase da lactação. Desta forma, a inexistência de alterações significativas no ranking genético dos animais em avaliações sucessivas provocadas pela mudança da exploração de “cluster” torna viável a aplicação deste método de formação de “clusters” na avaliação genética de bovinos leiteiros em Portugal.

Apesar de os resultados serem conclusivos relativamente às vantagens da aplicação de técnicas de aglomeração de explorações nas avaliações nacionais, ficou clara a necessidade de investigar mais com o objectivo de eliminar algumas limitações da aplicação desta técnica, como por exemplo, assumir alguns dos efeitos ambientais sistemáticos como variáveis aleatórias. A restrição relativa ao número mínimo de observações por nível de efeitos fixos é bastante rigorosa, tendo em conta que a presença de apenas um GC pequeno implica a aglomeração de toda a exploração. O período de vida produtiva das explorações pode não ser coincidente, existindo sempre explorações recentes e explorações extintas. Além destes factores, é também necessário considerar o facto de as práticas de manejo de uma exploração poderem variar ao longo do tempo.

A técnica de aglomeração de explorações com número insuficiente de observações por GC, desenvolvida e apresentada nos estudos que compõem esta tese, levou à adopção do MCAR como metodologia de avaliação oficial por parte das estruturas nacionais responsáveis pela avaliação genética das características de produção de bovinos leiteiros.

5. BIBLIOGRAFIA

- Animal Improvement Programs Laboratory, USDA. Acedido em Setembro 2007 em: <http://www.ars.usda.gov>.
- Associação Nacional para o Melhoramento dos Bovinos Leiteiros. 2006. Publicação de Resultados.
- Boichard, D., L. Maignel, e É. Verrier. 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.* 29:5-23.
- Bormann, J., G. R. Wiggans, T. Druet, and N. Genglar. 2002. Estimating effects of permanent environment, lactation stage, age, and pregnancy on test-day yield. *J. Dairy Sci.* 85(Jan.). Online. Available: <http://www.adsa.org/jds/>.
- Carabaño, M. J., A. Moreno, P. López-Romero, and C. Díaz. 2004. Comparing alternative definitions of the contemporary group effect in Avileña Negra Ibérica beef cattle using classical and Bayesian criteria. *J. Anim. Sci.* 82:3447-3457.
- Carvalho, J., 1996. Development and application of an autoregressive process for estimating genetic parameters of daily milk yield in a herd of tropical Lucerna cattle and in US Holstein herds. Doctoral Dissertation, Cornell Univ., Ithaca, NY.
- Carvalho J. 2000. Predição de parâmetros genéticos para a produção diária de leite em gado Holstein, Pardo Suíço e Modicana no sudeste Siciliano. X Congresso de Zootecnia, Estação Zootécnica Nacional, Vale de Santarém: 103.
- Carvalho, J., R.W. Blake, E.J. Pollak, R.L. Quaas, and C.V. Duran-Castro. 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in US Holstein herds. *J. Dairy Sci.* 81:2738-2751.
- Carvalho, J., E. J. Pollak, R. L. Quaas and R. W. Blake. 2002a. An autoregressive repeatability animal model for test-day records in multiple lactations. *J. Dairy Sci.* 85:2040-2045.
- Carvalho, J., J. Vasconcelos, A. Martins, M.F. Petim-Batista, A. Silvestre, A. Ferreira e J. Colaço. 2002b. Parâmetros e Tendências Genéticas para Produção Diária de Leite, Gordura e Proteína do Efectivo Nacional Holstein Usando um Modelo de Contrastes Auto-regressivo. XII Congresso de Zootecnia. Vila Real, 21 a 23 de Novembro de 2002, p.531-537.
- Crump, R. E., N. R. Wray, R. Thompson, and G. Simm. 1997. Assigning pedigree beef performance records to contemporary groups taking account of within-herd calving patterns. *Animal Science* 65:193-198.
- Direcção Geral de Veterinária. 2003. Avaliação Genética da Raça Frísia.
- Direcção Geral de Veterinária. 2005. Boletim Estatístico nº11.

- Gabinete de Planeamento e Política Agro-Alimentar. 2005. Anuário Pecuário 2005.
- Gabinete de Planeamento e Política Agro-Alimentar. 2006. Agricultura Portuguesa – Principais Indicadores 2005.
- Gama, L.T., I. Carolino, e N. Carolino. 2000. Correlações genéticas e fenotípicas entre a produção de leite e gordura em bovinos da raça Frísia. Proc. Congresso de Zootecnia, Estação Zootécnica Nacional, Vale de Santarém: 104.
- Gengler N., A. Tijani, G. R. Wiggans and I. Misztal. 1999. Estimation of (Co)Variance Function Coefficients for Test Day Yield with a Expectation-Maximization Restricted Maximum Likelihood Algorithm. J. Dairy Sci. 82:1849.
- Gengler N., A. Tijani, G. R. Wiggans and J. C. Philpot. 2001. Estimation of (Co)Variance Functions for Test-Day Yields during First and Second Lactation in the United States. J. Dairy Sci. 84:542.
- Hartl, D. L., and A. G. Clark. Principles of Population Genetics. 3. ed. Sunderland: Sinauer Associates Inc. 1997. 682p.
- Harris, B., and D. Johnson. 1998. Approximate reliability of genetic evaluations under an animal model. J. Dairy Sci. 81:2723-2728.
- Henderson, C. R., 1973. Sire evaluation and genetic trends. In Proc. Anim. Breeding Genet. Symp. In Honor of Dr. J. L. Lush, Am. Soc. Anim. Sci., Am. Dairy Sci. Assoc., Champaign, IL.
- Henderson, C. R., 1975. Comparison of alternative sire evaluation methods. J. Anim. Sci. 41:760-770.
- Jamrozik, J., and L. R. Schaeffer, 1997. Estimates of Genetic Parameters for a Test Day Model with Random Regressions for Yield Traits of First Lactation Holsteins. J. Dairy Sci. 80:762-770.
- Jamrozik, J., L. R. Schaeffer and J. C. M. Dekkers, 1997. Genetic Evaluation of Dairy Cattle Using Test Day Yields and Random Regression Model. J. Dairy Sci. 80:1217-1226.
- Macciotta, N. P. P., A. Capio-Borlino, and G. Pulina. 1999. Analysis of environmental effects on test day milk yields of Sarda dairy ewes. J. Dairy Sci. 82:2212-2217.
- Meyer, K. 1987. Estimates of variances due to sire $\times$ herd interactions and environmental covariances between paternal half-sibs for first lactation dairy production. Livestock Prod. Sci. 17:95-115.
- Miglior, F., B. Szkotnicki, and E. B. Burnside. 1992. Analysis of levels of inbreeding and inbreeding depression in Jersey cattle. J Dairy Sci. 75, 1112-1118.
- Misztal, I., T. Stabel, J. Jamrozik, E. A. Mäntysaari, and T.H.E. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. J. Dairy Sci. 83:1125-1134.

- Norman, H. D., P. M. VanRaden, J. R. Wright and J. S. Clay. 1999. Comparison of test interval and best prediction methods for estimation of lactation yield from monthly, a.m.-p.m., and trimonthly testing. *J. Dairy Sci.* 82:438-444.
- Petim-Batista, M.F., A. Silvestre, J. Colaço, J. Vasconcelos, A. Martins, A. Ferreira e J. Carvalheira. 2002. Estimativa de Parâmetros Genéticos para Produção aos 305 Dias e Análise da Evolução Genética em Bovinos Leiteiros Nacionais. XII Congresso de Zootecnia. Vila Real, 21 a 23 de Novembro de 2002, p.566-570.
- Ptak, E. and L. R. Schaeffer, 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci.* 34: 23-34.
- Ptak, E., and A. Zarnecki. 2000. Comparison of different test day models used for genetic evaluation of Polish Black-and-White cattle. *J. of Anim. And Feed Sci.* 9(2):261-269.
- Rekaya, R., M. J. Carabaño, and M. A. Touro. 1999. Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle. *Livestock Prod. Sci.* 57(3):203-217.v
- SAS Institute Inc., SAS/STAT® User's Guide, Version 6, Volume 1. 1989.
- Schaeffer, L. 1987. Estimation of variance components under a selection model. *J. Dairy Sci.* 70:661-671.
- Schaeffer, L. R.. 1991. C. R. Henderson: Contributions to Predicting Genetic Merit. *J. Dairy Sci* 1991 74: 4052-4066.
- Schaeffer, L., and B. Sullivan, 1994. Genetic evaluation of dairy goats using test day yields. *Proc. 5th World Cong. Gen. Appl. Livest. Prod.* 18:182-185.
- Schaeffer, L. R., J. Jamrozik, G. J. Kistemaker, and B. J. Van-Doormaal. 2000. Experience with a test-day model. *J. Dairy Sci.* 83: 1135:1144.
- Schmitz, F., R. W. Everett, and R. L Quaas. 1991. Herd-year-season clustering. *J. Dairy Sci.* 74:629-636.
- Smith, L. A., B. G. Cassell, and R. E. Pearson. 1998. The effects of inbreeding on the lifetime performance of dairy cattle. *J Dairy Sci*, 81, 2729-2737.
- Strabel, T., and T. Szwaczkowski. 1999. The use of test day models with small size of contemporary groups. *Journal of Animal Breeding and Genetics* 116(5):379-386.
- Swalve, H. H., F. Rosner, and W. Wemheuer. 2003. Inbreeding in the German Holstein cow population. *Proc. EAAP – 54th Annual Meeting, Rome, Italy, 2003*, 17.
- Swalve, H. H. 1995. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *J. Dairy Sci.* 78:929-938.
- Tierney, T. S., and L. R. Schaeffer. 1994. Inclusion of semen price of the sire in an animal model to account for preferential treatment. *J. Dairy Sci.* 77:576-582.
- Van Bebber, J., N. Reinsch, W. Junge, E. Kalm. 1997. Accounting for herd, year and season effects in genetic evaluation of dairy cattle: a review. *Livestock Prod. Sci.*

- 51:191-203.
- Van Vleck, L. 1987. Contemporary groups for genetic evaluations. *J. Dairy Sci.* 70:2456-2464.
- VanRaden, P. M. 1997. Lactation yields and accuracies computed from test day yields and (co)variances by best prediction. *J. Dairy Sci.* 80:3015-3022.
- VanRaden, P. M., and L. A. Smith. 1999. Selection and mating considering expected inbreeding of future progeny. *J Dairy Sci.* 82:2771-2778.
- Vasconcelos, J. e J. Carvalheira. 2003. Monitoring herd performance in dairy farms using Test-Day models. EAAP – 54th Annual Meeting, Rome – Italy. 31 August – 3 September 2003, Book of Abstracts No. 9 (2003), 170.
- Vasconcelos, J., e J. Carvalheira. 2004. Análise Preliminar da Conexão Genética entre Grupos de Maneio na Raça Holstein-Frísia em Portugal. XIV Congresso de Zootecnia. Angra do Heroísmo, 7 a 9 de Outubro de 2004, p.98-99.
- Vasconcelos, J. e J. Carvalheira. 2005. Effect of Small-Herd Clustering on the Genetic Connectedness of the Portuguese Holstein Cattle Population. EAAP – 56th Annual Meeting, Uppsala – Sweden. 5 a 8 de Junho 2005, Abstract no. 424.
- Vasconcelos, J., and J. Carvalheira. 2007. Assessment of heterogeneity of residual variances in an autoregressive test-day multiple lactations model. EAAP – 58th Annual Meeting, Dublin – Ireland. 26 - 29 de August 2007, Session 10, Poster 39.
- Vasconcelos, J., A. Martins, M. F. Petim-Batista, A. Silvestre, A. Ferreira, J. Colaço, J. Carvalheira. 2002a. Estimation of lactation yields from test day yields of milk, fat and protein for Portuguese Holstein cattle using an autoregressive test day model. 7th World Congress on Genetics Applied to Livestock Production. 2002. Proc. 7th World Cong. Gen. Appl. Livest. Prod., CD-ROM communication n° 01-57.
- Vasconcelos, J., A. Martins, M.F. Petim-Batista, A. Silvestre, J. Colaço, A. Ferreira e J. Carvalheira. 2002b. Validação do Modelo de Contrastes Auto-regressivo para a Predição de Lactações Padronizadas aos 305 dias, de Leite, Gordura e Proteína do Efectivo Leiteiro Nacional. XII Congresso de Zootecnia. Vila Real, 21 a 23 de Novembro de 2002, p.519- 524.
- Vasconcelos, J., A. Martins, A. Ferreira, S. Pinto e J. Carvalheira. 2003. Análise Preliminar dos Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal. XIII Congresso de Zootecnia. Évora, 1 a 4 de Outubro de 2003, p.52.
- Vasconcelos, J., A. Martins, M. F. Petim-Batista, J. Colaço, R. W. Blake, and J. Carvalheira. 2004. Prediction of Daily and Lactation Yields of Milk, Fat, and

- Protein Using an Autoregressive Repeatability Test Day Model. *J. Dairy Sci.* 87:2591-2598
- Vasconcelos, J., A. Martins, A. Ferreira, and J. Carvalheira. 2005a. Consequências da eliminação de rebanhos pequenos da avaliação genética de bovinos leiteiros em Portugal. *Revista Portuguesa de Zootecnia*, Ano XII, 2:105-117.
- Vasconcelos, J., A. Martins, A. Ferreira, S. Pinto, e J. Carvalheira. 2005b. Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal. *Revista Portuguesa de Ciências Veterinárias* 100(553-554): 33-38.
- Vasconcelos, J., A. Martins, A. Ferreira e J. Carvalheira. 2005c. Consequências da Eliminação de Pequenos Rebanhos da Avaliação Genética de Bovinos Leiteiros em Portugal. XV Congresso de Zootecnia. Vila Real, 3 a 5 de Novembro de 2005, p.160-163.
- Vasconcelos, J., F. Santos, R. Barroso, A. Martins, A. Ferreira, J. Carvalheira. 2006. Effects of Clustering Dairy Herds for Genetic Evaluations using Different Descriptors to Define Similarities between Production Environments. 8th World Congress on Genetics Applied to Livestock Production. 2006. Proc. 8th World Cong. Gen. Appl. Livest. Prod., CD-ROM communication n° 24-30.
- Vasconcelos, J., A. Bagnato, A. Pereira, and J. Carvalheira. 2008. Effects of clustering herds with small size contemporary groups in dairy cattle genetic evaluations. *J. Dairy Sci.* 91:377-384.
- Visscher, P. M. and M. E. Goddard. 1993. Fixed and random contemporary groups. *J. Dairy Sci.* 76:1444-1454.
- Wade, K. M., 1990. Parameter estimation for mixed linear models with first order autoregressive covariance structure. Doctoral Diss., Cornell Univ. Ithaca, NY, pp 117.
- Wiggans, G. R. and M. E. Goddard, 1997. A Computationally Feasible Test Day Model for Genetic Evaluation of Yield Traits in the United States. *J. Dairy Sci.* 80:1795-1800.
- Wiggans, G. R., e P. M. VanRaden. 1995. Calculation and use of inbreeding coefficients for genetic evaluation of United States dairy cattle. *J. Dairy Sci.* 78:1584-1590.
- Wiggans, G. R., P. M. VanRaden, J. Bormann, J. C. Philpot, T. Druet, and N. Gengler. 2002. Deriving lactation yields from test-day yields adjusted for lactation stage, age, pregnancy, and herd test date. *J. Dairy Sci.* 85(Jan.). Online. Available: <http://www.adsa.org/jds/>.
- Young, C. W., e A. J. Seykora. 1996. Estimates of inbreeding and relationship among registered Holstein females in the United States. *J. Dairy Sci.* 79:502-505.

6. ANEXOS

Anexo I*

Estimation of lactation yields from test day yields of milk, fat and protein for Portuguese Holstein cattle using an autoregressive test day model

*Vasconcelos, J., A. Martins, M. F. Petim-Batista, A. Silvestre, A. Ferreira, J. Colaço, and J. Carneiro. 2002. Estimation of lactation yields from test day yields of milk, fat and protein for Portuguese Holstein cattle using an autoregressive test day model. 7th World Congress on Genetics Applied to Livestock Production. 2002. Proc. 7th World Cong. Gen. Appl. Livest. Prod., CD-ROM communication n° 01-57.

ESTIMATION OF LACTATION YIELDS FROM TEST DAY YIELDS OF MILK, FAT AND PROTEIN FOR PORTUGUESE HOLSTEIN CATTLE USING AN AUTOREGRESSIVE TEST DAY MODEL. J. Vasconcelos, A. Martins, M. Petim-Batista, A. Silvestre, A. Ferreira, J. Colaço and J. Carvalheira. CECA and ICBAS, Universidade do Porto, Rua do Monte-Crasto, 4485-661 Vairão, Portugal.

Genetic evaluations in Portugal are from terminated lactations (L) with 210d or more, extended by test interval methodology (TIM). All L with less than 210d or still in progress, are excluded. Objectives were to develop an autoregressive test day (ARTD) model to predict 305d L. Data was provided by ANABLE consisting of 6,694,962 TD records. ARTD procedures allowed a substantial increase of 305d yields in the first 3 L for milk, fat and protein. Correlations between ARTD L and TIM L ranged from 0.97 to 0.99. The high correlations obtained in this study suggest that the ARTD model may be a viable alternative to estimate 305d L yields. More studies are necessary to determine differences in accuracy between the two methods.

ESTIMATION OF LACTATION YIELDS FROM TEST DAY YIELDS OF MILK, FAT AND PROTEIN FOR PORTUGUESE HOLSTEIN CATTLE USING AN AUTOREGRESSIVE TEST DAY MODEL

J. Vasconcelos¹, A. Martins², M. Petim-Batista³, A. Silvestre³, A. Ferreira², J. Colaço³ and J. Carvalheira¹

¹Instituto de Ciências Biomédicas Abel Salazar and Centro de Estudos de Ciência Animal, Universidade do Porto, Rua do Monte-Crasto, 4485-661 Vairão, Portugal

²Associação para o apoio à Bovinicultura Leiteira do Norte

³Universidade de Trás-os-Montes e Alto Douro

INTRODUCTION

The Portuguese dairy cattle associations estimate cumulative lactation yields for milk and milk components by the Test Interval Method (TIM, Sargent, 1968). The Portuguese National milk data set is relatively small with an average of about 90,000 new lactations (L) records added per year. However, National genetic evaluations use only yields from terminated L with 210 or more days-in-milk (DIM), extended to 305-d if necessary, using pre-calculated factors (Gama *et al.*, 2000). All L with less than 210-d are excluded from the genetic analysis, representing each year 11% to 14% of all terminated L (ANABLE, 1999). On top of this number, is also necessary to add all L still in progress at the moment of data retrieval for analysis. At the end of the year 2000, for example, there were in the first three L, 61,000 cows in milk (30% of all terminated L in that year) that were also excluded from any analysis. In a relatively small data set as this, any information that may be added to the evaluation would be an important contribution to improve the reliability of the predicted breeding values.

Test day (TD) models are becoming the standard for genetic evaluations of production traits in dairy cattle. This methodology may also be applied to estimate 305-d cumulative yields of milk, fat and protein using the predicted daily yields from the model. This method is being applied in Canada (Schaffer and Jamrozik, 1996) using the Wood equation (Wood, 1967) in a multiple trait approach incorporating the covariances between the three traits. In the US, VanRaden (1997) obtained more accurate 305-d yields using a best prediction TD model to estimate these traits. More recently, Carvalheira *et al.* (1998, 2001) indicated that autoregressive TD repeatability models (ARTD) could estimate accurate L curves (on a herd basis) and predict all TD residual yields from all cows present in the analysis. Given the acknowledged potential of this modeling approach to reduce the size of residual variance components, ARTD models may be a viable alternative to derive 305-d L yields from observed TD yields. In this manner, the estimated 305-d L records may reflect the general advantages associated with TD models by better accounting for environmental effects.

The objective of this work was to develop and test an ARTD model to predict missing TD for every cow with records. The cow effect would be partitioned in two independent variables: the first measuring correlations between L, and the second accounting for correlations between TD within L. The covariance matrices for these variables, will assume first-order autoregressive processes. Cumulative 305-d L may then be estimated, summing the appropriate means to the predicted TD residuals.

MATERIAL AND METHODS

Data. The Portuguese National Association of Dairy Cattle Breeders (ANABLE) provided the data for this study, with a total of 6,694,962 TD records, produced between June 1994 and December 2000. The editing process conformed the data for the analysis following the assumptions of the ARTD model. Only records from a.m.-p.m. supervised test plans belonging to the first three L were used. Records missing relevant information, outliers, or with less than 5 or more than 305 DIM were deleted. Criteria on minimum and maximum age limits within L were adopted from ANABLE. The sequence between TD within L for each cow was checked and those with intervals smaller than 15-d or greater than 75-d were not used. Cows with only one record per L were also deleted. Table 1 gives the structure of the data set for milk, fat and protein submitted to the analysis.

Table 1. Number of test day milk, fat and protein yields per lactation (L) after editings (data is from June 1994 to December 2000).

Data	Milk	Fat	Protein
Original	6,694,962	6,694,962	6,694,962
Final	2,101,098	2,093,524	2,103,506
1 st L	915,338	913,001	916,500
2 nd L	695,628	692,260	696,289
3 rd L	490,132	488,263	490,717

The final data consisted of records produced by 141,681, 141,777 and 141,730 cows for milk, fat and protein, respectively.

The model. The ARTD model takes into account environmental effects that are specific to each test date, age at calving and lactation curve within each herd. This partition of the fixed effects permits the estimation of specific means for each herd, increasing the accuracy of the predicted TD. The model used in the present study was a “cow model” assuming no genetic relationships between animals. The ARTD model for milk, fat and protein yields, was as follows:

$$y_{ijkLmn} = \text{HTD}_i + \text{Age(H)}_j + \text{DIM(H)}_{k(L)} + p_{m(L)} + t_{n(mL)} + e_{ijkLmn},$$

Where, y is the TD yield, **HTD** is the fixed effect of herd-test-date, **Age(H)** is the fixed effect of age at calving nested within herd, **DIM(H)** is the fixed effect of DIM nested with herd and L, p is the random effect of the long term environmental effects accounting for the correlations generated by the cow across L, t is the random effect of the short term environmental effects accounting for the correlations due to the cow between TD and within each L, and e is the random residual effect. Both, p and t are fitted with first-order autocorrelation structures.

A detailed description of the model expectations and (co)variance structure for this type of models, may be found in *Carvalho et al.*, (2001). To help ensure accuracy, at least 3 observations in each contemporary group were required. Farms (herds) failing this requirement were grouped based in cluster analysis using the method of Ward (SAS, 1989), where phenotypic mean and standard deviation in first L defined similarity between herds. It was assumed that farms with similar phenotypic mean and standard deviation had similar management. Analysis were for single trait and (co)variance components were estimated by DFREML-SIMPLEX method (Nelder *et al.*, 1965). Cumulative 305-d L yields were then estimated summing the appropriate within herd means to the predicted TD residuals for each cow.

RESULTS AND DISCUSSION

Presently, only terminated 305-d L yields with more than 210 DIM (approximately the number of records with > 210-d in Table 2) are used for the National genetic evaluations. ARTD procedures allowed a substantial increase in the total number of 305-d L records available for those analysis (records with < 210-d in Table 2). These records (either L in progress or terminated) were projected to 305-d L yields corresponding to a recuperation of 20.8%, 20.5% and 20.2% of the data for milk, fat and protein, respectively, for the first 3 L. This data set now contains 245,466, 254,474 and 254,610 of ARTD 305-d L yields of milk, fat and protein, respectively, and is available for further analysis. In total, the National data set for the first 3 L, was incremented by 51,106, 52,274 and 51,369 records for milk, fat and protein, respectively, which, for a relatively small database, may be a significant improvement.

Table 2. Number of estimated 305-d lactation yields derived from test day yields of milk, fat and protein from the Portuguese Holstein database (from June1994 to December 2000) using an autoregressive test day model.

Lactation		Milk	Fat	Protein
1 st	< 210-d	19,688	19,960	19,577
	> 210-d	89,022	88,817	89,194
	Total	108,710	108,777	108,771
2 nd	< 210-d	18,113	18,348	18,022
	> 210-d	66,948	66,652	67,070
	Total	85,061	85,000	85,092
3 rd	< 210-d	13,305	13,966	13,770
	> 210-d	46,890	46,736	46,977
	Total	60,695	60,702	60,747

Milk, fat and protein phenotypic means and standard deviations for the first 3 ARTD L are in Table 3. These means were slightly greater but not significantly different than those estimated by TIM. Trends of ARTD 305-d yields were positive for all traits and L for the period studied (109,4 kg/yr, 115,3 kg/yr e 105,9 kg/yr for milk, 4,1 kg/yr, 4,3 kg/yr e 4,0 kg/yr for fat and 4,3 kg/yr, 4,5 kg/yr, 4,2 kg/yr for protein, for 1st, 2nd and 3rd L, respectively).

Table 3. Phenotypic means (M) and standard deviations (SD) for yields of milk, fat and protein (kg) for the first 3 lactations (L) estimated at 305-d by the autoregressive test day model.

L	Milk		Fat		Protein	
	M	SD	M	SD	M	SD
1 st	6,946.5	1,472.6	251.6	57.0	220.3	46.9
2 nd	7,563.5	1,749.2	272.9	66.2	241.3	54.3
3 rd	7,667.8	1,741.4	277.0	66.5	242.0	53.9

Correlations between ARTD 305-d L yields and corresponding TIM records (only for those L having 210 or more DIM) are in Table 4. The high correlations obtained in this study suggest that the ARTD model may be a viable alternative to estimate 305-d L yields, although more studies are necessary to determine differences in accuracy between the two methods.

Table 4. Correlations between the 305-d lactation (L) yields estimated by the autoregressive test day model and by the test interval method (only for those lactations of length 210-d or more) for the first 3 lactations of Portuguese Holstein cattle.

Trait	1 st L	2 nd L	3 rd L
Milk	0.99	0.98	0.98
Fat	0.98	0.98	0.97
Protein	0.99	0.98	0.98

REFERENCES

- ANABLE (1999) « Publicação de Resultados », Editors Verde Milho, Aveiro, Portugal.
- Carvalho J., R. W. Blake, E. J. Pollak, R. L. Quaas, and C. V. Duran-Castro (1998) *J. Dairy Sci.* **81**(10):2738-51.
- Carvalho J., Pollak, E. J., Quaas, R. L. Blake, R. W. (2001) *Proc. 52nd EAAP*:6.
- Gama, L.T., Carolino, I. and Carolino, N. (2000) *Proc. Congresso de Zootecnia*:104.
- Nelder, J. A., and R. Mead. (1965) *Computer J.* **7**:308-313.
- Sargent, F. D., V. H. Lytton, e O. J. Wall, Jr. (1968) *J. Dairy Sci.* **51**:170.
- SAS Institute Inc., SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC: SAS Institute, Inc., (1989). 943 pp.
- Schaeffer, L. R., J. Jamrozik (1996) *J. Dairy Sci.* **79**:2044-2055.
- Vanraden, P. M. (1997) *J. Dairy Sci.* **80**:3015-3022.
- Wood, P. D. P. (1967) *Nature* **216**:164-165.

ESTIMATION OF STANDARDIZED LACTATION YIELDS FROM TEST DAY YIELDS OF MILK, FAT AND PROTEIN FOR THE PORTUGUESE HOLSTEIN CATTLE USING AN AUTOREGRESSIVE REPEATABILITY TEST DAY MODEL

J. Vasconcelos¹, A. Martins², M. Petim-Batista³, A. Silvestre³, A. Ferreira², J. Colaço³ and J. Carnevalheira¹

¹Instituto de Ciências Biomédicas Abel Salazar and Centro de Estudos de Ciência Animal, Universidade do Porto, Rua do Monte-Crasto, 4485-661 Vairão, Portugal

²Universidade de Trás-os-Montes e Alto Douro

³Associação para o apoio à Bovinicultura Leiteira do Norte

INTRODUCTION

The Portuguese dairy cattle associations estimate cumulative lactation yields for milk and milk components by the Test Interval Method (TIM; Sargent, 1968), using only yields from terminated lactations (L) with 210 or more days-in-milk (DIM), extended to 305-d if necessary, using pre-calculated factors (Gama *et al.*, 2000). An average of about 90,000 L records are added per year. This is a relatively small data base but, is still further reduced because:

- 11% to 14% of all terminated L are excluded from the genetic analysis for having less than 210-d (ANABLE, 1999).
- At the moment of the data retrieval for this analysis, about 30% of all L records would also be excluded from a genetic analysis because they were in progress at that time.

Test day (TD) models are becoming the standard for genetic evaluations of production traits in dairy cattle. This methodology may also be applied to estimate 305-d cumulative yields of milk, fat and protein using predicted daily yields from the model. Carnevalheira *et al.* (1998, 2001) indicated that autoregressive TD repeatability models (ARTD) can:

- Estimate accurate L curves and other environmental means (on a herd basis) and predict all TD residual yields from all cows present in the analysis.
- Reduce the size of residual variance components.
- Take better accounting for environmental effects that are specific to each test date, age at calving and lactation curve within each herd.

The objective of this work was to develop and test an ARTD model to predict missing TD for every cow with records.

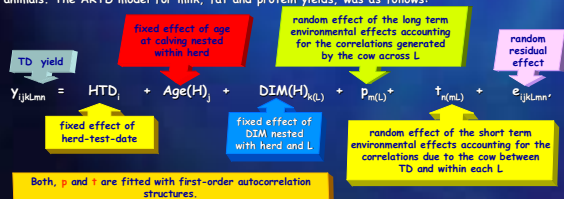
MATERIAL AND METHODS

- Data was provided by the Portuguese National Association of Dairy Cattle Breeders (ANABLE). Table 1 gives the structure of the data set for milk, fat and protein submitted to the analysis.
- 6,694,962 TD records, produced between June 1994 and December 2000.
- The editing process confirmed the data for the analysis following the assumptions of the ARTD model:
 - Only records from a.m.-p.m. supervised test plans belonging to the first three L were used.
 - Records missing relevant information, outliers, or with less than 5 or more than 305 DIM were deleted.
 - Criteria on minimum and maximum age limits within L were adopted from ANABLE.
 - The sequence between TD within L for each cow was checked and those with intervals smaller than 15-d or greater than 75-d were not used.
 - Cows with only one record per L were also deleted.

Table 1. Number of test day milk, fat and protein yields per lactation (L) after editing (data is from June 1994 to December 2000)

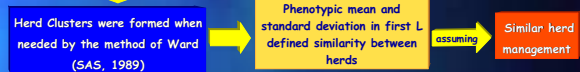
Data	Milk	Fat	Protein
Original	6,694,962	6,694,962	6,694,962
Final	2,101,098	2,093,524	2,103,506
1 st L	915,338	913,001	916,500
2 nd L	695,628	692,260	696,289
3 rd L	490,132	488,263	490,717

The model used in the present study was a "cow model" assuming no genetic relationships between animals. The ARTD model for milk, fat and protein yields, was as follows:



A detailed description of the model expectations and (co)variance structure may be found in Carnevalheira *et al.*, (2001).

To help ensure accuracy, at least 3 observations in each contemporary group were required.



Analysis were for single trait and (co)variance components were estimated by DFREML (Nelder *et al.*, 1965). Cumulative 305-d L yields were then estimated summing the appropriate within herd means to the predicted TD residuals for each cow.

RESULTS AND DISCUSSION

Presently, only terminated 305-d L yields with more than 210 DIM (approximately the number of records with > 210-d in Table 2) are used for the National genetic evaluations. ARTD procedures allowed a substantial increase in the total number of 305-d L records available for those analysis (records with < 210-d in Table 2). These records (either L in progress or terminated) were projected to 305-d L yields corresponding to a recuperation of 20.8%, 20.5% and 20.2% of the data for milk, fat and protein, respectively, for the first 3 L. This data set now contains 254,466, 254,479 and 254,610 of 305-d L yields for milk, fat and protein, respectively, and is available for further analysis. In total, the National data set for the first 3 L, was incremented by 51,106, 52,274 and 51,369 L records for milk, fat and protein, respectively, which, for a relatively small database, may be a significant improvement.

Table 2. Number of estimated 305-d lactation yields derived from test day yields of milk, fat and protein from the Portuguese Holstein database (from June 1994 to December 2000) using an autoregressive test day model.

Lactation	Milk	Fat	Protein
1 st	< 210-d: 19,688 > 210-d: 89,022 Total: 108,710	19,960 88,817 108,777	19,577 89,194 108,771
2 nd	< 210-d: 18,113 > 210-d: 66,948 Total: 85,061	18,348 66,652 85,000	18,022 67,070 85,092
3 rd	< 210-d: 13,305 > 210-d: 46,890 Total: 60,195	13,966 46,736 60,702	13,770 46,977 60,747

Milk, fat and protein phenotypic means and standard deviations for the estimated L by ARTD are in Table 3. These means were slightly greater but not significantly different than those estimated by TIM. Trends for these 305-d L yields were positive for all traits and L for the period studied (109.4 kg/yr, 115.3 kg/yr and 105.9 kg/yr for milk, 4.1 kg/yr, 4.3 kg/yr and 4.0 kg/yr for fat and 4.3 kg/yr, 4.5 kg/yr, 4.2 kg/yr for protein, for 1st, 2nd and 3rd L, respectively).

Table 3. Phenotypic means ($\bar{x}$) and standard deviations (SD) for yields of milk, fat and protein (kg) for the first 3 lactations (L), estimated at 305-d by the autoregressive test day model.

L	Milk		Fat		Protein	
	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1 st	6,946.5	1,472.6	251.6	57.0	220.3	46.9
2 nd	7,563.5	1,749.2	272.9	66.2	241.3	54.3
3 rd	7,667.8	1,741.4	277.0	66.5	242.0	53.9

Correlations between 305-d L yields estimated by ARTD and corresponding TIM records (only for those L having 210 or more DIM) are in Table 4. The high correlations obtained in this study suggest that the ARTD model may be a viable alternative to estimate 305-d L yields, although more studies are necessary to determine differences in accuracy between the two methods.

Table 4. Correlations between the 305-d lactation (L) yields estimated by the autoregressive test day model and by the test interval method (only for those lactations of length 210-d or more) for the first 3 lactations of Portuguese Holstein cattle.

Trait	1 st L	2 nd L	3 rd L
Milk	0.99	0.98	0.98
Fat	0.98	0.98	0.97
Protein	0.99	0.98	0.98

REFERENCES

- ANABLE (1999) « Publicação de Resultados », Editores Verde Milho, Aveiro, Portugal.
 Carnevalheira, J., R. W. Blake, E. J. Pollak, R. L. Quaas, and C. V. Duran-Gastro (1998) J. Dairy Sci. 81(10):2738-51.
 Carnevalheira, J., Pollak, E. J., Quaas, R. L., Blake, R. W. (2001) Proc. 52nd EAAP.
 Gama, L. T., Carolina, I. and Carolina, N. (2000) Proc. Congresso de Zootecnia 104.
 Nelder, J. A., and R. Mead. (1965) Computer J. 7:308-313.
 Sargent, E. D., V. H. Lyttle, and O. J. Wall, Jr. (1968) J. Dairy Sci. 51:170.
 SAS Institute Inc., SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC: SAS Institute, Inc., (1989), 943 pp.
 Schaeffer, L. R., J. Jannink (1998) J. Dairy Sci. 79:2044-2055.
 VanRaden, P. M. (1997) J. Dairy Sci. 80:3015-3022.
 Wood, P. D. P. (1967) Nature 216:164-165.

Anexo II*

Validação do Modelo de Contrastes Auto-regressivo para a Predição de Lactações Padronizadas aos 305 dias, de Leite, Gordura e Proteína do Efectivo Leiteiro Nacional

*Vasconcelos, J., A. Martins, M.F. Petim-Batista, A. Silvestre, J. Colaço, A. Ferreira, e J. Carneiro. 2002. Validação do Modelo de Contrastes Auto-regressivo para a Predição de Lactações Padronizadas aos 305 dias, de Leite, Gordura e Proteína do Efectivo Leiteiro Nacional. XII Congresso de Zootecnia. Vila Real, 21 a 23 de Novembro de 2002, p.519- 524.

Validação do Modelo de Contrastes Auto-regressivo para a Predição de Lactações Padronizadas aos 305 dias, de Leite, Gordura e Proteína do Efectivo Leiteiro Nacional

J. Vasconcelos¹, A. Martins³, M.F.Petim-Batista², A. Silvestre², J. Colaço², A. Ferreira³ e J. Carvalheira^{1,4}

¹CIBIO-ICETA - Universidade do Porto, ²Universidade de Trás-os-Montes e Alto Douro (UTAD), ³A.N.A.B.L.E., ⁴ICBAS - Universidade do Porto, PORTUGAL

Resumo

Actualmente, a avaliação genética do efectivo leiteiro Nacional é feita utilizando modelos de análise da produção acumulada aos 305 dias, apenas de lactações (LA) terminadas com mais de 210 dias. As LA com duração superior a 210 dias são projectadas para 305, utilizando factores pré-calculados de ajustamento. Estas restrições implicam a não utilização de grande parte da informação disponível (LA incompletas e em curso). A aplicação do modelo de contrastes auto-regressivo (MCAR) permite a recuperação destas LA, necessitando que tenham apenas dois contrastes consecutivos para predizer os contrastes em falta, que depois podem ser usados na extensão e ajustamento das LA aos 305 dias. O objectivo deste trabalho consistiu na validação desta metodologia, comparando-a com o método clássico do intervalo entre testes (MIT). Para este efeito, foi seleccionada uma amostra na base de dados do efectivo leiteiro Nacional, contendo apenas LA completas, i.e. todas as LA com 10 contrastes e pertencentes às três primeiras LA. Estes contrastes foram utilizados para calcular as LA padronizadas de 305 dias (L305) usando ambas as metodologias: MIT e MCAR. Para completar o estudo, foi ainda necessário criar uma base de dados reduzida, retirada da amostra original, onde, de forma aleatória, foram eliminados contrastes de cada uma das três LA, com o objectivo de simular as frequências encontradas nas LA incompletas da base de dados Nacional. O MCAR foi aplicado a estes dados para predizer os contrastes em falta e obter em simultâneo as L305. Foram assim criadas três novas variáveis: L305 calculada pelo MIT (L305_M) na amostra completa, L305 calculada pelo MCAR na amostra completa (L305_C) e L305 calculada pelo MCAR na amostra reduzida (L305_R). A validação do MCAR para predizer L305 a partir de LA completas ou incompletas foi feita através de uma análise de correlação (entre as produções e entre o “ranking”) para os valores obtidos pelas duas metodologias. As correlações entre L305_M e L305_C foram bastante elevadas e com valores superiores a 0,99 para as três LA e para cada uma das características, o que dá uma indicação da fiabilidade do novo método. As correlações entre L305_C e L305_R foram também muito altas (de 0,98 a 0,99), indicando uma elevada capacidade preditiva do MCAR. As correlações entre L305_M e L305_R variaram para as três LA nas três características entre 0,97 e 0,98, valores que permitiram concluir que as projecções de LA incompletas ou em curso estimadas pelo MCAR não alteram significativamente as L305, o que viabiliza a utilização deste modelo na obtenção de LA de 305 dias para futuras avaliações genéticas, com a vantagem de acrescentar cerca de 20% de registos à base de dados Nacional.

Palavras-chave: Modelo de contrastes, padronização de lactações, validação do modelo.

Introdução

Em Portugal, a avaliação genética do efectivo leiteiro Nacional é feita utilizando métodos de análise do registo da produção acumulada aos 305 dias, em apenas lactações (LA) terminadas. As LA com menos de 210 dias são rejeitadas e aquelas que possuem mais de 210, são projectadas até 305 dias, utilizando para tal factores pré-calculados de ajustamento (DMAP, 1999). Esta restrição (somente LA com mais de 210 dias) implica a não utilização de grande parte da informação disponível [11%, 13,6%, 13,7% e 13% das LA terminadas nos anos de 1996, 1997, 1998 e 1999 respectivamente (ANABLE, 1999)]. A acrescentar a este número é ainda necessário considerar as lactações em curso, que também não entram na avaliação Nacional. Estudos anteriores demonstraram que a utilização de LA mesmo que incompletas ou ainda em curso, não só não altera significativamente o rigor de predição do mérito genético de touros (Famula, 1981), como também permite a obtenção de maiores valores de heritabilidade e repetibilidade (Norman et al., 1985).

As produções acumuladas aos 305 dias são simples combinações lineares dos contrastes. O método do intervalo entre testes (MIT), surgiu em 1880 (Sargent, 1968), consistindo basicamente no uso da interpolação linear entre os valores dos contrastes, cuja soma resulta na produção total aos 305 dias. Entretanto, no Canadá, Schaeffer e Jamrozik (1996) e nos E.U.A, VanRaden (1997) desenvolveram métodos baseados em modelos de contrastes e que se revelaram mais eficazes que o MIT. Recentemente (Carvalho et al., 1998; Vasconcelos et al., 2002) foi também demonstrado que a utilização de modelos de contraste auto-regressivos (MCAR) permite estimar com maior precisão (menores erros residuais), as curvas de lactação padronizadas por exploração aos 305 dias, as quais poderão ser usadas directamente na predição dos contrastes para todos os animais em produção. O MCAR permite a recuperação das LA incompletas ou em curso, predizendo os contrastes em falta, necessitando para tal de apenas dois contrastes consecutivos (Vasconcelos et al., 2002). O objectivo deste trabalho consistiu na validação do MCAR para a predição de L305, comparando-as nomeadamente com as LA estimadas pelo MIT.

Material e Métodos

A base de dados dos contrastes de leite, gordura e proteína do efectivo Nacional foi cedida pela Associação Nacional para o Melhoramento dos Bovinos Leiteiros (ANABLE) e compreende todos os contrastes efectuados entre Junho de 1994 e Dezembro de 2000, num total de 6.694.962 observações (Obs). Durante o processo de edição dos dados conservaram-se apenas as Obs pertencentes às primeiras três LA de cada animal e os contrastes do tipo A4 (manhã e tarde). Foram eliminadas todas as Obs com menos de 5 ou mais de 310 dias em LA, assim como as Obs em desacordo com os pressupostos assumidos pelo MCAR (Carvalho et al., 1998, 2002). Durante este processo foram eliminadas cerca de 69% das Obs para o leite, gordura e proteína.

Da base de dados editada foram seleccionadas três amostras, uma para cada carácter, contendo apenas LA completas (104873, 102402 e 104815 LA respectivamente para o leite, gordura e proteína) i. e., todas com 10 contrastes. Estes contrastes foram usados para predizer as LA padronizadas de 305 dias (L305), aplicando as duas metodologias em estudo:

MIT. Este método foi aplicado às três amostras (uma por cada característica) para calcular as produções acumuladas aos 305 dias.

MCAR. A metodologia usada para este modelo encontra-se descrita em detalhe noutros trabalhos (e.g., Vasconcelos et al., 2002), podendo ser descrito da seguinte forma:

$$y_{ijkLmn} = \text{RAM}_i + \text{Idade}_j + \text{DEL(R)}_k(L) + \text{pm}(L) + t_{n(mL)} + r_{ijkLmn},$$

onde,

y_{ijkLmn} representa o valor do contraste; RAM_i mede o efeito fixo do rebanho-ano-mês em que se observou o contraste; Idade_j mede o efeito fixo correspondente à jésima idade ao parto; $\text{DEL(R)}_k(L)$ mede o efeito fixo devido às vacas que são testadas na mesma classe de DEL para cada rebanho e lactação (curvas de lactação); $\text{pm}(L)$ representa o efeito aleatório do ambiente a longo prazo, medindo a correlação criada pela vaca ao produzir em mais do que uma lactação, e para o qual se assume um processo auto-regressivo de 1ª ordem; $t_{n(mL)}$ representa o efeito aleatório do ambiente a curto prazo, que neste caso mede a correlação entre os contrastes da mesma vaca para cada lactação. Assume-se que os contrastes são independentes entre lactações e que também seguem um processo auto-regressivo de 1ª ordem; r_{ijkLmn} é o termo residual aleatório.

Os componentes de variância para as três características, foram estimados pelo método DFREML-SIMPLEX (Nelder *et al.*, 1965). Tal como o MIT, o MCAR foi aplicado às amostras seleccionadas (compostas por LA com 10 contrastes), para calcular as L305: **L305_M** e **L305_C** respectivamente.

Com o objectivo de testar a capacidade preditiva do modelo, foram eliminadas aleatoriamente Obs de cada LA, de modo a obter uma amostra reduzida que simulasse a frequência de contrastes existente nas LA incompletas da base de dados Nacional, para cada uma das características. Estas frequências representam a estrutura da base de dados reduzida e encontram-se descritas na Tabela 1, sendo os valores da gordura e proteína semelhantes aos do leite.

Tabela 1. Frequência dos contrastes (%) nas três primeiras lactações (LA) da base de dados Nacional (produção de leite).

Nº de Contrastes	2	3	4	5	6	7	8	9
1ª LA	4,0	4,0	4,0	3,5	3,6	4,1	6,9	25,9
% 2ª LA	4,2	4,6	4,6	4,3	4,4	5,4	8,6	25,0
3ª LA	4,6	4,8	4,7	4,8	4,5	5,6	8,4	24,9

Das amostras reduzidas foram obtidas as L305 preditas pelo MCAR (**L305_R**).

A validação do MCAR para predição das L305 foi baseada numa análise de correlação (entre as produções e entre o “ranking”) para os valores obtidos pelas duas metodologias.

Resultados e Discussão

As médias e desvios padrão estimados pelas duas metodologias para as L305, por característica encontram-se na Tabela 2.

Tabela 2. Número de lactações (LA) padronizadas aos 305 dias, média (M) e desvio padrão (DP) calculadas pelo Método do Intervalo entre Testes (MIT) e pelo Modelo de Contrastes Auto-Regressivo (MCAR) na amostra completa e na amostra reduzida.

Método	Amostra	Leite			Gordura			Proteína		
		nº LA	M	DP	nº LA	M	DP	nº LA	M	DP
MIT	Completa	104873	7513,0	1609,0	102402	267,2	61,4	104815	237,2	50,7
MCAR	Completa	104873	7496,0	1582,0	102402	267,2	60,0	104815	237,1	50,0
MCAR	Reduzida	104873	7497,0	1565,0	102402	267,3	59,4	104815	237,1	49,3

Os coeficientes de correlação para a produção (Tabela 3), assim como da correlação entre o “ranking” (Tabela 4) entre L305M e L305C foram bastante elevados e com valores superiores a 0,99, para cada característica, quer considerando todas as LA em conjunto, quer individualmente, o que dá uma indicação da fiabilidade do novo método.

Tabela 3. Coeficientes de correlação entre lactações padronizadas aos 305 dias (total e para as três lactações individualmente), calculadas pelo Método do Intervalo entre Testes (**L305_M**) e pelo Modelo de Contrastes Auto-Regressivo na amostra completa (**L305_C**) e na amostra reduzida (**L305_R**).

	Leite				Gordura				Proteína			
	total	1ªLA	2ªLA	3ªLA	total	1ªLA	2ªLA	3ªLA	total	1ªLA	2ªLA	3ªLA
L305 _M e L305 _C	0,997	0,998	0,997	0,995	0,997	0,998	0,997	0,995	0,997	0,998	0,997	0,995
L305 _C e L305 _R	0,984	0,985	0,983	0,979	0,980	0,982	0,978	0,975	0,983	0,984	0,981	0,976
L305 _M e L305 _R	0,984	0,985	0,982	0,980	0,979	0,981	0,977	0,975	0,982	0,984	0,981	0,977

As correlações entre L305C e L305R foram também muito altas (de 0,98 a 0,99), indicando uma elevada capacidade preditiva do MCAR. As correlações entre L305M e L305R variaram para as três LA nas três características entre 0,97 e 0,98, valores que permitiram concluir que as projecções de LA incompletas ou em curso estimadas pelo MCAR não alteram significativamente as L305.

Tabela 4. Coeficientes de correlação do “ranking” entre as lactações padronizadas aos 305 dias (total e para as três lactações individualmente), calculadas pelo Método do Intervalo entre Testes (**L305_M**) e pelo Modelo de Contrastes Auto-Regressivo na amostra completa (**L305_C**) e na amostra reduzida (**L305_R**).

	Leite				Gordura				Proteína			
	total	1ªLA	2ªLA	3ªLA	total	1ªLA	2ªLA	3ªLA	total	1ªLA	2ªLA	3ªLA
L305 _M e L305 _C	0,997	0,998	0,997	0,994	0,997	0,998	0,997	0,996	0,998	0,998	0,998	0,995
L305 _C e L305 _R	0,983	0,984	0,982	0,977	0,980	0,981	0,978	0,974	0,982	0,984	0,981	0,975
L305 _M e L305 _R	0,983	0,984	0,982	0,979	0,979	0,980	0,977	0,974	0,981	0,983	0,980	0,976

Adicionalmente foi feita a correlação entre LA completas (L305M e L305C) e LA incompletas (L305R) da mesma vaca, estimadas a partir da informação de menos de dez contrastes (entre 2 e 9 contrastes). Os coeficientes de correlação variaram entre 0,86, no caso de LA estimadas a partir de 2 contrastes e 0,99 para as LA calculadas com base na informação de 9 contrastes (Tabela 5). Os elevados coeficientes de correlação obtidos, quer entre L305 no geral, quer em cada LA individualmente e ainda em LA incompletas, consoante o número de contrastes em falta, viabilizam a utilização deste modelo na obtenção de LA a

305 dias para futuras avaliações genéticas, com a vantagem de acrescentar cerca de 20% de registos à base de dados Nacional.

Tabela 5. Coeficientes de correlação entre lactações padronizadas aos 305 dias calculadas pelo Método do Intervalo entre Testes (**L305_M**) e pelo Modelo de Contrastes Auto-Regressivo na amostra completa (**L305_C**) e na amostra reduzida a partir de lactações incompletas com 2 a 9 contrastes (**L305_R**).

Nº de contrastes		2	3	4	5	6	7	8	9
L305M L305R	Leite	0,877	0,917	0,945	0,961	0,976	0,986	0,993	0,997
	Gordur								
	a	0,859	0,897	0,928	0,949	0,967	0,980	0,990	0,996
	Proteín								
	a	0,864	0,910	0,934	0,955	0,974	0,984	0,992	0,997
L305C L305R	Leite	0,877	0,917	0,944	0,976	0,986	0,993	0,993	0,997
	Gordur								
	a	0,863	0,898	0,929	0,951	0,968	0,981	0,990	0,996
	Proteín								
	a	0,864	0,910	0,935	0,955	0,974	0,984	0,992	0,997

Agradecimentos

Este estudo insere-se num projecto aprovado pela FCT e pelo POCTI, participado pelo fundo comunitário europeu FEDER.

Bibliografia

- ANABLE, 1999. Melhoramento Nacional de Bovinos Leiteiros. Publicação de Resultados, Verde Milho, Aveiro.
- Carvalho J., R. W. Blake, E. J. Pollak, R. L. Quaas, and C. V. Duran-Castro. 1998. J. Dairy Sci. 81:2738-2751.
- Carvalho J., E. J. Pollak, R. L. Quaas and R. W. Blake. 2002. J. Dairy Sci. 85:2040-2045.
- DMAP - Divisão de Melhoramento Animal e de apoio à Produção. 1999. Avaliação Genética da Raça Frísia. Direcção de Serviços de Produção e Melhoramento Animal, Direcção Geral de Veterinária.
- Famula, T. R. and L. D. Van Vleck. 1981. J. Dairy Sci. 64:484-490.
- Gama, L.T., I. Carolino, e N. Carolino. 2000. Congresso de Zootecnia, Estação Zootécnica Nacional, Vale de Santarém: 104.
- Nelder, J. A., and R. Mead. 1965. Computer J. 7: 308-313.
- Norman, H. D., F. N. Dickinson and J. R. Wright. 1985. J. Dairy Sci. 68:2646-2654.
- Schaeffer, L. R., J. Jamrozik. 1996. J. Dairy Sci. 79:2044-2055.
- Vanraden, P. M. 1997. Lactation yields and accuracies computed from test day yields and (co)variances by best prediction. J. Dairy Sci. 80:3015-3022.
- Vasconcelos, J., A. Martins, M. F. Petim-Batista, A. Silvestre, A. Ferreira, J. Colaço, J. Carvalho. 2002. 7th World Congress on Genetics Applied to Livestock Production. CD-ROM communication n° 01-57.



Validação do Modelo de Contrastes Auto-regressivo para a Predição de Lactações Padronizadas aos 305 dias, de Leite, Gordura e Proteína do Efectivo Leiteiro Nacional



J. Vasconcelos¹, A. Martins³, M.F. Petim-Batista², A. Silvestre², J. Colaço², A. Ferreira³ e J. Carnevalheira^{1,4}

¹CIBIO-ICETA - Universidade do Porto, ²Universidade de Trás-os-Montes e Alto Douro (UTAD), ³A.N.A.B.L.E., ⁴ICBAS - Universidade do Porto, PORTUGAL

INTRODUÇÃO

Em Portugal, a avaliação genética do efectivo leiteiro Nacional é feita utilizando métodos de análise do registo da produção acumulada aos 305 dias, em apenas lactações (LA) terminadas. As LA com menos de 210 dias são rejeitadas e aquelas que possuem mais de 210, são projectadas até 305 dias, utilizando factores pré-calculados de ajustamento (DMAP, 1999). Esta restrição (somente LA com mais de 210 dias) implica a não utilização de grande parte da informação disponível:

- 11% a 14% das LA terminadas entre 1996 e 1999 foram excluídas da avaliação genética por possuírem menos de 210 dias (ANABLE, 1999).

A acrescentar a este número é ainda necessário considerar as lactações em curso, que também não entram na avaliação Nacional:

- 20.000 (30%) das três primeiras LA iniciadas entre 1994 e 2000, estavam em progresso no dia 30/12/2000.
- Num estudo anterior foi desenvolvido um modelo de Contrastes Auto-regressivo (MCAR) para vacas (Vasconcelos *et al.*, 2002), onde a maioria destas LA incompletas foram recuperadas e padronizadas a 305 dias.

Importa agora validar este processo comparando a metodologia MCAR com a metodologia standard: Método do Intervalo entre Testes (MIT):

Método do Intervalo entre Testes

MIT

- Surgiu em 1880 (Sargent, 1968)
- Consiste basicamente no uso da interpolação linear entre os valores dos contrastes, cuja soma resulta na produção total aos 305 dias

VS

Modelo de Contrastes Auto-regressivo

MCAR

- Modelos de contrastes mais eficazes que o MIT (Schaeffer *et al.*, 1996; VanRaden, 1997)
- Permite estimar com maior precisão as curvas de lactação padronizadas por exploração aos 305 dias (Carvalho *et al.*, 1998; Vasconcelos *et al.*, 2002) usadas directamente na predição dos contrastes para todos os animais em produção
- Permite a recuperação das LA incompletas ou em curso, predizendo os contrastes em falta, necessitando para tal de apenas dois contrastes consecutivos (Vasconcelos *et al.*, 2002).

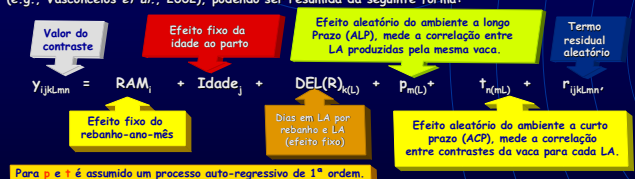
MATERIAL E MÉTODOS

- A base de dados dos contrastes de leite, gordura e proteína do efectivo Nacional foi cedida pela Associação Nacional para o Melhoramento dos Bovinos Leiteiros (ANABLE)
- Contrastes efectuados entre Junho de 1994 e Dezembro de 2000, num total de 6.694.962 observações (Obs).
- Durante o processo de edição dos dados, conservaram-se apenas as Obs de acordo com os pressupostos assumidos pelo MCAR (Carvalho *et al.*, 1998, 2002). Durante este processo foram eliminadas cerca de 69% das Obs para o leite, gordura e proteína.

Da base de dados editada foram seleccionadas três amostras, uma para cada característica, contendo apenas LA completas (104873, 102402 e 104815 LA respectivamente para o leite, gordura e proteína) i. e., todas com 10 contrastes. Estes contrastes foram usados para predizer as LA padronizadas de 305 dias (L305), aplicando as duas metodologias em estudo:

MIT ■ Aplicado às três amostras (uma por cada característica) para calcular as produções acumuladas aos 305 dias: L305_M.

MCAR ■ A metodologia usada para este modelo encontra-se descrita em detalhe noutros trabalhos (e.g., Vasconcelos *et al.*, 2002), podendo ser resumida da seguinte forma:



Os componentes de variância para as três características, foram estimados pelo método DFREML-SIMPLEX (Nelder *et al.*, 1965). Tal como o MIT, o MCAR foi aplicado às amostras seleccionadas (compostas por LA com 10 contrastes), para calcular as L305: L305_C.

Com o objectivo de testar a capacidade preditiva do modelo, foram eliminadas aleatoriamente Obs de cada LA, de modo a obter uma amostra reduzida que simulasse a frequência de contrastes existente nas LA incompletas da base de dados Nacional, para cada uma das características. Estas frequências representam a estrutura da base de dados reduzida (Tabela 1), sendo os valores da gordura e proteína semelhantes aos do leite.

Tabela 1. Frequência dos contrastes (%) nas três primeiras lactações (LA) da base de dados Nacional (produção de leite).

Nº de Contrastes	2	3	4	5	6	7	8	9
1ª LA	4,0	4,0	4,0	4,0	4,0	4,0	4,0	4,0
%	2ª LA	4,2	4,2	4,2	4,2	4,2	4,2	4,2
	3ª LA	4,6	4,6	4,6	4,6	4,6	4,6	4,6

Das amostras reduzidas foram obtidas as L305 preditas pelo MCAR (L305_C).

A validação do MCAR para predição das L305 foi baseada numa análise de correlação (entre as produções e entre o "ranking") para os valores obtidos pelas duas metodologias.

RESULTADOS E DISCUSSÃO

As médias e desvios padrão estimados pelas duas metodologias para as L305, por característica encontram-se na Tabela 2.

Tabela 2. Número de lactações (LA) padronizadas aos 305 dias (média (M) e desvio padrão (DP) calculados pelo Método do Intervalo entre Testes (MIT) e pelo Modelo de Contrastes Auto-regressivo (MCAR) na amostra completa e na amostra reduzida.

Método	Amostra	Leite			Gordura			Proteína		
		nº LA	M	DP	nº LA	M	DP	nº LA	M	DP
MIT	Completa	104873	7513,0	1609,0	102402	267,2	61,4	104815	237,2	50,7
MCAR	Completa	104873	7496,0	1582,0	102402	267,2	60,0	104815	237,1	50,0
MCAR	Reduzida	104873	7497,0	1565,0	102402	267,3	59,4	104815	237,1	49,3

Os coeficientes de correlação para a produção (Tabela 3), assim como da correlação entre o "ranking" dos animais (Tabela 4) entre L305_M e L305_C, foram bastante elevados e com valores superiores a 0,99, para cada característica, quer considerando todas as LA em conjunto, quer individualmente, o que dá uma indicação da fiabilidade do novo método.

Tabela 3. Coeficientes de correlação entre lactações padronizadas aos 305 dias (total e para as três lactações individualmente), calculados pelo Método do Intervalo entre Testes (L305_M) e pelo Modelo de Contrastes Auto-regressivo na amostra completa (L305_C) e na amostra reduzida (L305_R).

		Leite				Gordura				Proteína			
		total	1ª LA	2ª LA	3ª LA	total	1ª LA	2ª LA	3ª LA	total	1ª LA	2ª LA	3ª LA
L305 _M e L305 _C	0,997	0,998	0,997	0,995	0,997	0,998	0,997	0,995	0,997	0,998	0,997	0,995	0,995
L305 _C e L305 _R	0,984	0,985	0,983	0,979	0,980	0,982	0,978	0,975	0,983	0,984	0,981	0,976	0,976
L305 _M e L305 _R	0,984	0,985	0,982	0,980	0,979	0,981	0,977	0,975	0,982	0,984	0,981	0,977	0,977

As correlações entre L305_M e L305_C foram também muito altas (de 0,98 a 0,99), indicando uma elevada capacidade preditiva do MCAR. As correlações entre L305_M e L305_R variaram para as três LA nas três características entre 0,97 e 0,98, valores que permitem concluir que as projecções de LA incompletas ou em curso estimadas pelo MCAR não alteram significativamente as L305.

Tabela 4. Coeficientes de correlação do "ranking" dos animais entre as lactações padronizadas aos 305 dias (total e para as três lactações individualmente), calculados pelo Método do Intervalo entre Testes (L305_M) e pelo Modelo de Contrastes Auto-regressivo na amostra completa (L305_C) e na amostra reduzida (L305_R).

		Leite				Gordura				Proteína			
		total	1ª LA	2ª LA	3ª LA	total	1ª LA	2ª LA	3ª LA	total	1ª LA	2ª LA	3ª LA
L305 _M e L305 _C	0,997	0,998	0,997	0,994	0,997	0,998	0,997	0,996	0,998	0,998	0,998	0,998	0,998
L305 _C e L305 _R	0,983	0,984	0,982	0,977	0,980	0,981	0,978	0,974	0,982	0,984	0,981	0,975	0,975
L305 _M e L305 _R	0,983	0,984	0,982	0,979	0,979	0,980	0,977	0,974	0,981	0,983	0,980	0,976	0,976

Para testar o efeito da redução do número de contrastes na predição da LA completa, foi feita a correlação entre LA incompletas (L305_R), estimadas a partir da informação de menos de dez contrastes (entre 2 e 9 contrastes) e as respectivas LA completas (L305_M e L305_C) da mesma vaca.

Os coeficientes de correlação variaram entre 0,86, no caso de LA estimadas a partir de 2 contrastes e 0,99 para as LA calculadas com base na informação de 9 contrastes (Tabela 5).

Os elevados coeficientes de correlação obtidos, quer entre L305 no geral, quer em cada LA individualmente e ainda em LA incompletas, consoante o número de contrastes em falta, viabilizam a utilização deste modelo na obtenção de LA a 305 dias para futuras avaliações genéticas, com a vantagem de acrescentar cerca de 20% de registos à base de dados Nacional.

Tabela 5. Coeficientes de correlação entre lactações padronizadas aos 305 dias calculadas pelo Método do Intervalo entre Testes (L305_M) e pelo Modelo de Contrastes Auto-regressivo na amostra completa (L305_C) e na amostra reduzida a partir de lactações incompletas com 2 a 9 contrastes (L305_R).

Nº de contrastes		2	3	4	5	6	7	8	9
L305M L305R	Leite	0,877	0,917	0,945	0,961	0,976	0,986	0,993	0,997
	Gordura	0,859	0,897	0,928	0,949	0,967	0,980	0,990	0,996
	Proteína	0,864	0,910	0,934	0,955	0,974	0,984	0,992	0,997
L305C L305R	Leite	0,877	0,917	0,944	0,976	0,986	0,993	0,993	0,997
	Gordura	0,863	0,898	0,929	0,951	0,968	0,981	0,990	0,996
	Proteína	0,864	0,910	0,935	0,955	0,974	0,984	0,992	0,997

AGRADECIMENTOS

Este estudo insere-se num projecto apoiado pela FCT e pelo POCTE, financiado pelo fundo comunitário europeu FEDER.

BIBLIOGRAFIA

- ANABLE, 1999. Melhoramento Nacional de Bovinos Leiteiros. Publicação de Resultados, Verde Milho, Aveiro.
- Carvalho, J. R. W. Blake, E. J. Pollak, R. L. Quares, and C. V. Duran-Castro. 1998. J. Dairy Sci. 81:2738-2751.
- Carvalho, J. R. W. Blake, R. L. Quares and R. W. Blake. 2002. J. Dairy Sci. 85:2040-2045.
- DMAP - Divisão de Melhoramento Animal e de Apoio à Produção. 1999. Avaliação Genética da Raça Frísia. Direcção de Serviços de Produção e Melhoramento Animal, Direcção Geral de Veterinária.
- Familio, T. R. and L. D. Van Vleck. 1981. J. Dairy Sci. 64:484-490.
- Gama, L. T., I. Carvalho, e N. Carvalho. 2000. Congresso de Zootecnia, Estação Zootécnica Nacional, Vale de Santarém: 104.
- Nelder, J. A., and R. Mead. 1965. Computer J. 7: 308-313.
- Norman, H. D., F. N. Dickinson and J. R. Wright. 1985. J. Dairy Sci. 68:2646-2654.
- Schaeffer, L. R., J. Jamrozik. 1996. J. Dairy Sci. 79:2044-2055.
- Vanraden, P. M. 1997. J. Dairy Sci. 80:3015-3022.
- Vasconcelos, J., A. Martins, M. F. Petim-Batista, A. Silvestre, A. Ferreira, J. Colaço, J. Carnevalheira. 2002. 7th WCGALP, CD-ROM communication n.º 01-57.

Anexo III*

Monitoring herd performance in dairy farms using Test-Day models

*Vasconcelos, J. e J. Carvalheira. 2003. Monitoring herd performance in dairy farms using Test-Day models. EAAP – 54th Annual Meeting, Rome – Italy. 31 August – 3 September 2003, Book of Abstracts No. 9 (2003), 170.

Monitoring Herd Performance in Dairy Farms Using Test-Day Models

J. Vasconcelos^{*,1} and J. Carvalheira^{1,2}, ¹Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-ICETA) - Universidade do Porto, Rua Padre Armando Quintas-Crasto, 4485-661 Vairão, PORTUGAL, ²Instituto de Ciências Biomédicas de Abel Salazar (ICBAS) - Universidade do Porto

The objectives of this study were to analyse the ability of an autoregressive test-day (TD) animal model to monitor management factors affecting herd performance in a monthly basis. TD data were used to estimate systematic environmental effects to evaluate schemes of herd management and individual cow productivity on a farm basis. Management curves, lactation curves and predictions of daily and cumulative yields were updated every month, showing the dynamics of each herd and were translated into a series of tables and graphics for further analysis. Management curves indicate the within-year variation in average daily yield and seasonal effects on quality, availability and cost of feed. These estimates are adjusted to a common age, stage of lactation, genetic differences and random environmental residuals. They represent herd production as if all cows were at the same age, stage of lactation, etc., on every TD. As an indicator of the farmer's feeding and management program, management curves shows what the cows would have produced if they were all the same. An example with the performance of a farm for a period of one year is discussed. This information may provide farmers and consultants with a better understanding of the dairy farm enterprise and contribute for a greater efficiency.

MONITORING HERD PERFORMANCE IN DAIRY FARMS USING TEST-DAY MODELS

J. Vasconcelos^{1,1} and J. Carvalho^{1,2},

¹Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-ICETA) - Universidade do Porto, Rua Padre Armando Quintas-Crasto, 4485-661 Vairão, PORTUGAL

²Instituto de Ciências Biomédicas de Abel Salazar (ICBAS) - Universidade do Porto

INTRODUCTION

Test-day (TD) data can be a useful source of information to support dairy farm management in the process of decision-making (Pietersma et al., 2001). TD data can be used to estimate systematic environmental effects to evaluate the management performance of the herd and the individual cow productivity (Carvalho et al., 2002b).

The objectives of this study were to evaluate the ability of an Autoregressive Repeatability TD (ARTD) animal model to extract and process management information in order to monitor factors affecting herd performance in a monthly basis.

MATERIAL AND METHODS

The Data: Records produced by the Portuguese Holstein cattle population from June 1994 up to the month in consideration (from January to December, 2002).

- Only monthly records (twice daily milking)
- From supervised plans
- Editings conformed the TD data to ICAR (1995) rules for TD milk recording and to the assumptions of the ARTD model (Carvalho et al., 1998, 2002a).
- The data for the present study was from farms tested together in the first week of every month and then simultaneously fitted to the model.

The model for milk, fat and protein was as follows:

$$Y_{ijklmn} = \text{HTD}_i + \text{Age}(H)_j + \text{DIM}(H)_{klm} + a_m + p_{m(L)} + t_{n(L)} + e_{ijklmn}$$

Where Y_{ijklmn} = vector of TD yields; HTD_i = fixed effect of herd TD; $\text{Age}(H)_j$ = fixed effect of age at calving nested within herd; $\text{DIM}(H)_{klm}$ = fixed effect of DIM nested within herd and L , a_m is the animal random effect; $p_{m(L)}$ = long term environmental random effect assumed to follow a first-order autoregressive process between L ; $t_{n(L)}$ = short term environmental random effect assumed to follow a first-order autoregressive process between TD within L ; e_{ijklmn} = random residual effect. Expectations and other statistical assumptions are in Carvalho et al. (2002a).

Model solutions from the systematic environmental effects were used to build graphics and tables reflecting the effects of management (herd and cow performance) for the previous 12 months for every month of 2002.

RESULTS AND DISCUSSION



Milk Management Curves (Figure 1)

Indicate:

- Monthly variation in average daily yield adjusted to a common age, stage of L, genetic differences and random environmental residuals.
- Farmer's feeding and management program, showing what the cows would have produced if they were all at the same age, stage of L, etc., on each TD.
- Observed variation will mainly correspond to the quality, availability and cost of feed effects.

Each graph:

- Gives the adjusted daily average for the previous 12 months;
- Represent residuals from the adjusted year average;
- Show gains and losses from the farm performance potential.

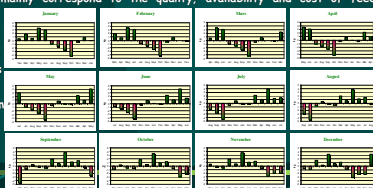


Figure 1. Milk management curves for each month of 2002 from the selected farm.

- The negative trend during the summer months (Figure 1) may be associated with important changes in the feeding programme (end of high quality silage and beginning of the corn harvest season). Farms that compensate for this low feed quality period show smaller variations with a substantial increase in the year yield average.
- Fat and protein management curves (not showed) had very similar patterns compared to the milk curves.

Herd Standard Lactation Curves (Figure 2)

The analysis of herd standard L curves can be a valuable tool providing the perception of L shape, peak and persistency.

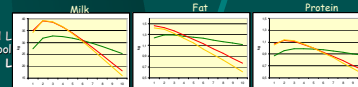


Figure 2. First (—), second (---), third and later (---) standard lactation curves for milk, fat and protein from the selected farm (December, 2002).

1st L curves were more persistent than the 2nd, 3rd and later L

Anomalous 3rd and later L (below the 2nd calvers)

May indicate that older cows are being neglected in terms of feed management

Peakless fat L curves (especially 2nd, 3rd and later L)

May indicate an inadequate diet and insufficient reserves of body tissues to support early L, emphasising the possibility of a neglecting feed management practice, especially in older cows (Raffrenato et al., 2003).

Individual Production Curves (Figure 3)

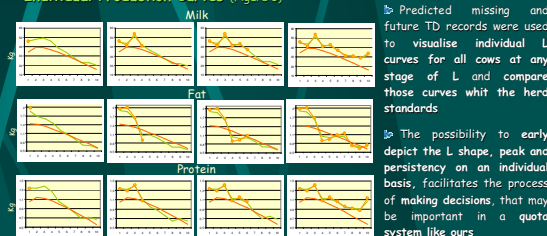


Figure 3. Example for predictions of a cow lactation curves for milk, fat and protein yields (—) using 1, 4, 6 and 10 actual TD yields (●). The herd 3rd and plus standard lactation curve (---) is showed for comparisons.

Cow List (Table 1)

Table 1 gives an example of a list with one cow currently in production (December, 2002) for milk, fat and protein, respectively.

Trait	Test Day (kg)										305-d of L ¹	305-d of L ²	305-d of L ³
	1	2	3	4	5	6	7	8	9	10			
Milk	52.1	44.3	36.7	34.0	28.5	30.8	26.8	23.3	23.1	18.9	10118.7	91	226
Fat	1.80	1.48	1.18	1.01	1.01	0.94	0.97	0.83	0.90	0.77	359.9	13	274
Prot	1.35	1.14	0.98	0.93	0.74	0.83	0.83	0.74	0.76	0.64	308.8	104	276

Ranking the cows from the most to the least productive animal associated with their genetic rank may be a basis for selection and culling decisions.

CONCLUSION

The application of the ARTD animal model to TD data provided useful information to evaluate the management performance of the herd and individual cow productivity for the farms under study. This procedure should be runned every month including the most recent TD information in order to keep farmers and their advisers constantly up-to-date and ready to intervene as soon as possible on the management of their farms, helping them to stay competitive and profitable.

REFERENCES

- Carvalho J., R. W. Blake, E. J. Pollak, R. L. Quass, and C. V. Bureau-Castro. 1998. J. Dairy Sci. 81:2738-2750.
 Carvalho J., E. J. Pollak, R. L. Quass, and R. W. Blake. 2002a. J. Dairy Sci. 85:2040-2045.
 Carvalho J., R. L. Quass, R. W. Blake, S. Vennart, and R. W. Blake. 2002b. 7th World Congress on Genetics Applied to Livestock Production, CD-ROM communication n° 01-11.
 International Committee for Animal Recording. 1999. Recording Guidelines. Appendixes to the information agreement of recording practices (milk recording). Rome, Italy.
 Glens V. E., and P. C. B. Gledhill. 1999. Proc. 99th International Meet. August 24-27, 1999, Zurich, Switzerland.
 Pietersma, D., B. Lacroix, D. Lefebvre, E. Black, and K. M. Wade. 2001. J. Dairy Sci. 84:730-739.
 Raffrenato, E., R. W. Blake, P. A. Offenberg, J. Carvalho, and R. L. Quass. 2003. J. Dairy Sci. 86:2470-2479.

ACKNOWLEDGMENT

This work contributed to the project POCT/33162/07/2005, which was partially funded by the Fundação para a Ciência e a Tecnologia (FCT) and FEDER (EU).

Anexo IV*

Análise Preliminar dos Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal

*Vasconcelos, J., A. Martins, A. Ferreira, S. Pinto e J. Carvalheira. 2003. Análise Preliminar dos Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal. XIII Congresso de Zootecnia. Évora, 1 a 4 de Outubro de 2003, p.52.

Análise Preliminar dos Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal

J. Vasconcelos¹, A. Martins², A. Ferreira², S. Pinto³ e J. Carvalheira^{1,4}

¹CIBIO-ICETA, Universidade do Porto.

²ABLN, Vila do Conde.

³APCRF, Samora Correia.

⁴ICBAS, Universidade do Porto.

O ficheiro do Pedigree Nacional da raça Holstein, com 238.982 machos e 825.393 fêmeas (1.064.375) registados (até 31 de Dezembro de 2002), foi analisado para a ocorrência de endogamia. A proporção de animais endogâmicos foi de 9% (94.941 animais) dos quais 31.240 eram machos e 63.701 fêmeas. Esta proporção é relativamente pequena, mas a tendência verificada foi de crescimento contínuo desde 1981 (1,12% por ano). Se considerarmos apenas os últimos 10 anos, este crescimento aumenta para 2,0% ao ano, valor preocupante em certas condições. A média dos coeficiente de endogamia (CE) de todos os animais registados e nascidos em 2001 foi baixa (0,4%), com uma pequena variação entre sexos (0,4% em fêmeas e 0,3% em machos). No que diz respeito aos animais endogâmicos, apesar de em número estarem a aumentar rapidamente, a média dos seus CE baixou muito entre 1985 e 1991 (-3,1% por ano), tendo praticamente estabilizado após esse período (-0,16% ao ano). Esta aparente contradição entre as tendências verificadas nos últimos 10 anos (2% de aumento anual de animais endogâmicos e 0,16% de diminuição anual na média dos CE), pode estar relacionada com a utilização mais ou menos generalizada dos programas de emparelhamento que, não podendo evitar totalmente o acasalamento entre animais relacionados sem afectar o valor genético da descendência, conseguem mesmo assim, minimizar o incremento do CE médio. O valor médio do CE nos animais endogâmicos nascidos em 2001 foi de 1,5%. Para determinar o efeito depressivo da endogamia na produção de leite, foram utilizadas as lactações ajustadas aos 305 dias de 141.053 primíparas, paridas entre 1994 e 2002. Estes dados foram analisados por um modelo fixo tendo como efeitos principais o rebanho/ano/estação do parto, a idade ao parto categorizada em 8 classes e o efeito linear dos CE como covariável. Todos os efeitos foram muito significativos ($P < 0,0001$), explicando 75,8% da variação total. O efeito de endogamia provocou uma depressão de $-6,6 \pm 1,5$ kg no leite produzido por cada 1% de aumento no CE. Comparativamente a estudos similares realizados na mesma raça em outros países, este resultado não parece ser suficiente para causar uma grande apreensão em termos produtivos. No entanto, a alta frequência de animais com ascendência desconhecida (40,1% - 27,5% com os dois e 12,6% com um dos pais desconhecidos), bem como a presença de erros de identificação no pedigree Nacional, poderão ser factores importantes para a justificação deste valor relativamente baixo. Por outro lado, e de forma semelhante a outros estudos, também se obteve um aumento maior que o esperado do efeito de depressão de endogamia para CE superiores a 15%. Apesar do efeito depressivo obtido neste estudo não parecer causar reduções importantes na produção de leite em vacas com CE médio, o aumento

intencional da endogamia deve ser evitado, pois os efeitos negativos no animal são globais e cumulativos. As perdas económicas que podem resultar da diminuição da produção serão incrementadas principalmente devido a factores relacionados com a diminuição da expressão dos caracteres ligados ao “fitness” (reprodução e sobrevivência).



Análise Preliminar dos Níveis de Endogamia e Depressão de Endogamia no Gado Bovino Leiteiro em Portugal

J. Vasconcelos¹, A. Martins², A. Ferreira², S. Pinto³ e J. Carvalheira^{1,4}

¹CIBIO-ICETA - Universidade do Porto, ²ABLN - Vila do Conde, ³APCRF - Samora Correia, ⁴ICBAS - Universidade do Porto, PORTUGAL



INTRODUÇÃO

O acasalamento de indivíduos relacionados produz descendência endogâmica, correspondendo o seu coeficiente de endogamia (CE) a $\frac{1}{2}$ do relacionamento genético aditivo entre os seus pais.

Uma das consequências da endogamia é a depressão de endogamia, manifestando-se numa redução do valor fenotípico médio, particularmente em caracteres produtivos e especialmente reprodutivos ("fitness").

A utilização quase exclusiva de IA na reprodução de gado leiteiro aliada a uma elevadíssima intensidade de selecção implicam o uso de um reduzido número de touros reprodutores, sendo de esperar que a endogamia venha a aumentar rapidamente. Este aumento da endogamia numa população traduz-se numa diminuição da variabilidade genética.

MATERIAL E MÉTODOS

Ficheiro do Pedigree Nacional da raça Holstein, com 1 064 375 animais registados (até 31 de Dezembro de 2002):



238,982

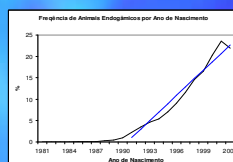
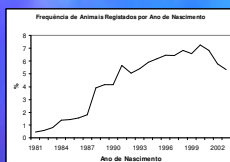


825,393

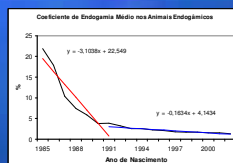
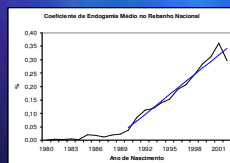
Análise de Proporção de Animais Endogâmicos e cálculo dos CE

Estudo do Efeito Depressivo da Endogamia na prod. de Leite

RESULTADOS E DISCUSSÃO



- ▢ A proporção de animais endogâmicos foi de 9% (94.941 animais) dos quais 31.240 eram machos e 63.701 fêmeas.
- ▢ Esta proporção é relativamente pequena, mas a tendência verificada foi de crescimento contínuo desde 1981 (1,12% por ano).
- ▢ Se considerarmos apenas os últimos 10 anos, este crescimento aumenta para 2,0% ao ano, valor preocupante em certas condições.



- ▢ A média dos coeficientes de endogamia (CE) de todos os animais registados e nascidos em 2001 foi baixa (0,4%), com uma pequena variação entre sexos (0,4% em fêmeas e 0,3% em machos).
- ▢ Os animais endogâmicos, apesar de em número estarem a aumentar rapidamente, a média dos seus CE baixou muito entre 1985 e 1991 (-3,1% por ano), tendo praticamente estabilizado após esse período (0,16% ao ano).

Esta aparente contradição entre as tendências verificadas nos últimos 10 anos (2% de aumento anual de animais endogâmicos e 0,16% de diminuição anual na média dos CE), pode estar relacionada com a utilização mais ou menos generalizada dos programas de emparelhamento que, não podendo evitar totalmente o acasalamento entre animais relacionados sem afectar o valor genético da descendência, conseguem mesmo assim, minimizar o incremento do CE médio.

O valor médio do CE nos animais endogâmicos nascidos em 2001 foi de 1,5%.

Para determinar o efeito depressivo da endogamia na produção de leite, foram utilizadas as lactações ajustadas aos 305 dias de 141.068 primíparas, paridas entre 1994 e 2002.

Estes dados foram analisados pelo seguinte modelo fixo:

$$Y = \text{RAE} + \text{Idade} + \text{CE} + r$$

onde:

- Y: 1305 Leite
- RAE: Efeito fixo do rebanho-ano-estação
- Idade: Efeito fixo da idade ao parto
- CE: Efeito linear do Coeficiente de Endogamia
- r: Termo residual aleatório

RESULTADOS E DISCUSSÃO

- ▢ Todos os efeitos foram muito significativos ($P < 0,0001$), explicando 75,8% da variação total.
- ▢ O efeito de endogamia provocou uma depressão de $-6,6 \pm 1,5$ kg no leite produzido por cada 1% de aumento no CE.

Comparativamente a estudos similares realizados na mesma raça em outros países, este resultado não parece ser suficiente para causar uma grande apreensão das ramagens produtivas. No entanto, a alta frequência de animais com ascendência desconhecida (40,1% - 27,5%) com os quais a 12,5% com um dos pais desconhecidos), bem como a presença de erros de identificação no pedigree Nacional, poderão ser factores importantes para a justificação deste valor relativamente baixo.

Por outro lado, a 1% forma semelhante a outros estudos, também se obtiveram um aumento maior que o esperado do efeito de depressão de endogamia para CE superiores a 15%.

Apesar do efeito depressivo obtido neste estudo não parecer causar reduções importantes na produção de leite em vacas com CE médio, o aumento intencional da endogamia deve ser evitado, pois os efeitos negativos no animal são globais e cumulativos. As perdas económicas que podem resultar da diminuição da produção serão incrementadas principalmente devido a factores relacionados com a diminuição da expressão dos caracteres ligados ao "fitness" (reprodução e sobrevivência).



XIII Congresso de Zootecnia,
1 a 4 de Outubro 2003 - Évora

BIBLIOGRAFIA

Miguel, F., B. Szotnicki, and S. B. Burnside (1992) J. Dairy Sci. 75:1112-1118.
Smith, L. A., B. G. Cassell, and R. E. Pearson (1998) J. Dairy Sci. 81:2729-2737.
Thompson, J. R., R. W. Everett, and N. L. Hammerschmidt (2000) J. Dairy Sci. 83:1856-1864.
Weigel, K. A., and S. W. Lin (2000) J. Dairy Sci. 83:822-828.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pela FCT, através da bolsa do Doutoramento SFRH/BD 11397/2002.

Anexo V*

Análise Preliminar da Conexão Genética entre Grupos de Maneio na Raça Holstein-Frísia em Portugal

*Vasconcelos, J., e J. Carvalheira. 2004. Análise Preliminar da Conexão Genética entre Grupos de Maneio na Raça Holstein-Frísia em Portugal. XIV Congresso de Zootecnia. Angra do Heroísmo, 7 a 9 de Outubro de 2004, p.98-99.

Análise Preliminar da Conexão Genética entre Grupos de Maneio na Raça Holstein-Frísia em Portugal

J. Vasconcelos^{1,2}, e J. Carvalheira^{1,2}

¹CIBIO-ICETA, Universidade do Porto. ²ICBAS, Universidade do Porto.

O rigor de predição das avaliações genéticas depende, entre outros factores, do grau de conexão genética (CG) entre os grupos de maneio (GM), definidos neste estudo como exploração/ano. Apesar de em Portugal serem realizadas avaliações periódicas, não existe uma quantificação da CG dos dados, bem como da sua evolução ao longo do tempo. O uso intensivo da IA nos últimos anos sugere, *à priori*, uma elevada CG. Torna-se, no entanto, necessário verificar se existe utilização sistemática dos touros por região, induzindo à formação de sub-populações isoladas, sem ligações genéticas entre si.

Por outro lado, existe o problema dos GM com reduzido número de efectivo, nos quais a predição dos valores genéticos dos animais se encontra comprometida. A formação de agrupamentos (clusters) de pequenos GM, baseada em critérios de similaridade de variâncias, pode ser uma solução. No entanto, esta pode estar a influenciar directamente a CG, induzindo a um aumento ou diminuição artificial das ligações genéticas existentes entre os novos GM estabelecidos pelos clusters, afectando o rigor de predição da comparação entre valores genéticos.

Para calcular o grau de CG da raça Holstein-Frísia em Portugal, foram utilizadas três bases de dados: (BD1) contem toda a informação relativa a explorações contrastadas entre Junho de 1994 e Dezembro de 2003 (196.019 animais), (BD2) onde explorações com reduzido número de animais contrastados foram eliminadas, (BD3) contem o mesmo número de animais que a BD1, mas as pequenas explorações foram aglomeradas, segundo critérios já definidos, formando GM maiores.

O método mais correcto para avaliar a CG utiliza a variância de diferenças entre os valores genéticos aditivos dos animais, definido por variância das diferenças do erro de predição (PEVd). No entanto, em grandes bases de dados, o calculo do PEVd torna-se computacionalmente difícil, se não impossível. Fries (1998) e mais tarde Roso *et al.* (2004), sugerem a utilização do número de ligações genéticas directas (LGD) entre GM, através dos pais (machos e fêmeas), como método alternativo para o cálculo da CG entre GM. Este método apresenta elevada correlação com o PEVd.

Os resultados são apresentados na Figura 1. Para a BD1, a média do grau de CG de cada GM com os restantes, aumentou desde 1994 até 1999, verificando-se de seguida um plateau até ao ano de 2002. Desde então, foi observado um decréscimo no valor médio da CG. Na BD2, apesar de mostrar o mesmo padrão de curva da BD1, foi verificado uma diminuição nos valores médios do grau de CG. A não inclusão das pequenas explorações na avaliação genética provocou uma diminuição da CG, logo do rigor de predição da mesma. Ao serem incluídas as pequenas explorações aglomeradas (BD3), observou-se um aumento do número de LGD por ano, apresentando, no entanto, o padrão da curva semelhante aos anteriores, apesar de apresentar um ligeiro decréscimo entre 1999 e 2002. Sendo o número de animais idêntico ao da BD1, o aumento

da CG observado é causa directa da aglomeração. Se por um lado a aglomeração de explorações parece induzir um aumento artificial do rigor de predição, por outro permite a utilização de um maior número de animais na avaliação, aumentando o valor médio do relacionamento genético na população. O passo seguinte desta investigação será a realização de um estudo mais exaustivo sobre as consequências da utilização desta técnica. A utilização intensiva da IA permitiu um elevado progresso genético anual, no entanto provocou uma diminuição real do número efectivo de reprodutores, o que pode explicar, pelo menos parcialmente, a diminuição dos LGD nos últimos anos. As implicações desta estratégia reprodutiva devem ser reavaliadas de modo a não inviabilizar o contínuo progresso genético necessário à indústria, ao mesmo tempo que se mantêm os mesmos níveis do rigor de predição.

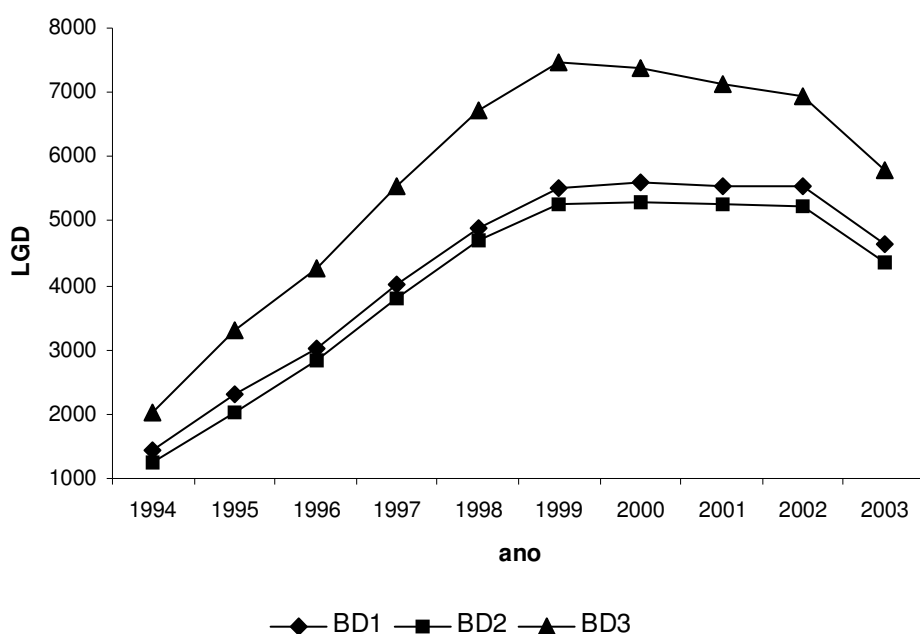


Figura 1. Média anual do grau de conexão genética (número de ligações genéticas directas - LGD) entre os grupos de manejo (exploração/ano).



Análise Preliminar da Conexão Genética entre Grupos de Maneio na Raça Holstein-Frisia em Portugal

J. Vasconcelos^{1,2}, e J. Carvalheira^{1,2}

¹CIBIO-ICETA - Universidade do Porto, ²ICBAS - Universidade do Porto, PORTUGAL



INTRODUÇÃO



- Um dos fatores que influencia o rigor de predição das avaliações genéticas, é o grau de conexão genética entre os grupos de maneio.
- Grupos de Maneio = exploração/ano.
- Em Portugal não existe nenhuma avaliação, quer quantitativa, quer qualitativa da conexão genética. $\Rightarrow$ O uso intensivo da IA nos últimos anos sugere, *a priori*, uma elevada conexão genética.
- Nos grupos de maneio com um número reduzido de dados, a predição dos valores genéticos dos animais encontra-se comprometida. Solução:
 - Formação de agrupamentos (clusters) de pequenos grupos de maneio. No entanto, podem influenciar a conexão genética se houver alguma sistematização do uso dos recursos genéticos.
- É necessário verificar se existe uma utilização sistemática dos touros por região, incluindo a formação de sub-populações isoladas, assim como averiguar se a formação de novos grupos de maneio influencia o grau de conexão genética de dados.

MATERIAL E MÉTODOS

3 Bases de Dados

Retiradas da base de dados Nacional da raça Holstein-Frisia:

- BD1 - contém toda a informação relativa a explorações contrastadas entre Junho de 1994 e Dezembro de 2003 (196.019 animais).
- BD2 - onde explorações com reduzido número de animais contrastados foram eliminados (menos de 3 obs. por grupo contemporâneo de efeitos fixos).
- BD3 - contém o mesmo número de animais que a BD1, mas as pequenas explorações foram aglomeradas em grupos de maneio maiores, usando como critério a similaridade entre as médias a variâncias fenotípicas.

LGD

Utilização do número de Ligações Genéticas Diretas como método alternativo ao PEVd (variância das diferenças de tiro da predição) para o cálculo da conexão genética entre os grupos de maneio (Fries (2003), Rosa *et al.* (2001)).

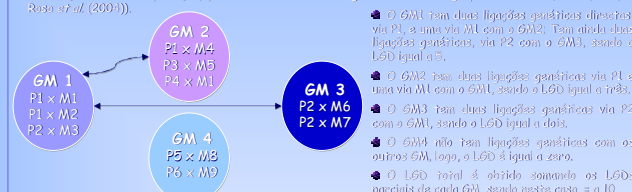
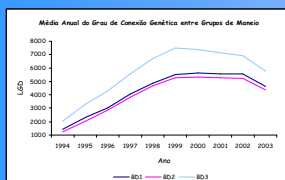


Fig. 1 - Exemplo ilustrativo do cálculo da LGD (onde GM = grupo de maneio, P = pai e M = mãe).

RESULTADOS E DISCUSSÃO



- BD1: A média do grau de CG de cada GM com os restantes, aumentou desde 1994 até 1999, verificando-se de seguida um plateau até ao ano de 2002. Desde então, foi observado um decréscimo no valor médio da Conexão Genética.
- BD2: Apesar de mostrar o mesmo padrão de curva da BD1, foi verificado uma diminuição nos valores médios do grau de Conexão Genética. A não inclusão das pequenas explorações nesta Bd provocou uma diminuição da CG, espelhando-se um impacto negativo no rigor de predição dos valores genéticos.
- BD3: Ao serem incluídas as pequenas explorações aglomeradas, observou-se um aumento do número de LGD por ano. Sendo o número de animais idêntico ao da BD1, o aumento da Conexão Genética pode ser explicado pelo facto de ao dividir o número total de ligações genéticas por um número menor de grupos de maneio com animais a contribuir com ligações, a média de ligações aumenta.

Se por um lado a aglomeração de explorações parece incluir um aumento artificial das conexões genéticas, por outro permite a utilização de um maior número de animais na avaliação, o que aumenta o valor médio do relacionamento genético na população.

A utilização intensiva da IA permitiu um elevado progresso genético anual, no entanto provocou uma diminuição real do número efectivo de reprodutores, o que pode explicar, pelo menos parcialmente, a diminuição dos LGD nos últimos anos.

As implicações desta estratégia reprodutiva devem ser reavaliadas de modo a não inviabilizar o contínuo progresso genético necessário à indústria, ao mesmo tempo que se mantêm os mesmos níveis do rigor de predição.

BIBLIOGRAFIA

- Carvalheira J. (2000) X Congresso de Zootecnia, Estação Zootécnica Nacional, Vale de Santarém: 103.
 Fries, L.A. (1998) Proc. 6th World Cong. Genet. Appl. Livest. Prod., Armidale, Australia. 27:449-500.
 Hanoq, E., e D. Boichard (1999) Genet. Select. Eval. 31(2): 163-176.
 Hanoq, E., L. Tiphine, e D. Boichard (1999) Producteurs animaux 12(2):101-111.
 Kennedy, B.W., D. Trus (1993) J. Anim. Sci. 71:2341-2352.
 Rosa, V.M., F.S. Schenkel, e S.R. Miller (2004) Canadian J. Anim. Sci. 84(1):37-47.



AGRADECIMENTOS

Este trabalho foi parcialmente financiado pela FCT, através do bolsa de doutoramento SFRH/BD 11397/2002.

XVI Congresso de Zootecnia, 7 a 10 de Outubro 2004
 Angra do Heroísmo

Anexo VI*

Effect of Small-Herd Clustering on the Genetic Connectedness of the Portuguese Holstein Cattle Population

*Vasconcelos, J. e J. Carvalheira. 2005. Effect of Small-Herd Clustering on the Genetic Connectedness of the Portuguese Holstein Cattle Population. EAAP – 56th Annual Meeting, Uppsala – Sweden. 5 a 8 de Junho 2005, Abstract no. 424.

Abstract no. 424; Session G4.44
joanav@mail.icav.up.pt

EFFECT OF SMALL-HERD CLUSTERING ON THE GENETIC CONNECTEDNESS OF THE PORTUGUESE HOLSTEIN CATTLE POPULATION

J. Vasconcelos^{*1,2} and J. Carvalho^{1,2}

¹*Centro Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Universidade do Porto,
Rua Padre Armando Quintas, 4485-661 Vairão, PORTUGAL,*

²*ICBAS-Universidade do Porto*

INTRODUCTION

Accuracy of genetic evaluations is directly related with connectedness (Kennedy and Trus, 1993; Hanocq and Boichard, 1999) between management groups (**MG**). Contemporary groups in test-day (**TD**) models may have an unacceptable number of observations (e.g., less than 4) depending on the herd size. Clustering these herds based on, for example, the simultaneous similarity between phenotypic mean and STD may prevent the loss of this information. However, these techniques may induce a false increase in genetic connectedness (**GC**), inflating the accuracy of EBV and promote preferential treatment for some sires. The aim of this study was to evaluate possible effects of clustering herds on the GC of the Portuguese Holstein cattle.

MATERIAL AND METHODS

The Portuguese National Association of Dairy Cattle Breeders (**ANABLE**) provided TD data, produced by 194,980 animals from 1997 to 2003. Ideally, average prediction error variance of differences (**PEVd**) in EBV between animals on different MG should be the basis of measurement of GC (Foulley *et al.*, 1992; Kennedy and Trus, 1993). In a dataset of this dimension the estimation of PEVd is computationally infeasible. Two alternative methods were applied to compute the degree of GC:

GLT - Total number of direct genetic links between MG (Roso *et al.*, 2004), defined in this work as herd/year:

GLT was calculated by summing the total number of common sires and dams in pairs of MG, forming genetic links (Fries, 1998). For each MG, the overall number of genetic links with all other MG was calculated.

GDV - Genetic drift variance (Kennedy and Trus, 1993):

It was obtained from $X'ZAZ'X$ matrix, that measures the sum of genetic relationship between and within MG. Average relationship between and within MG were obtained by dividing diagonal elements of $X'ZAZ'X$ matrix by the square of the number of records in the MG and the off-diagonal elements by the product of the number of records in each of the MG considered. The resulting matrix, $\bar{A}$, can be interpreted as the components of the genetic drift (co)variances. The average GDV for each MG with all other was calculated as the average of all variance differences with respect to that MG.

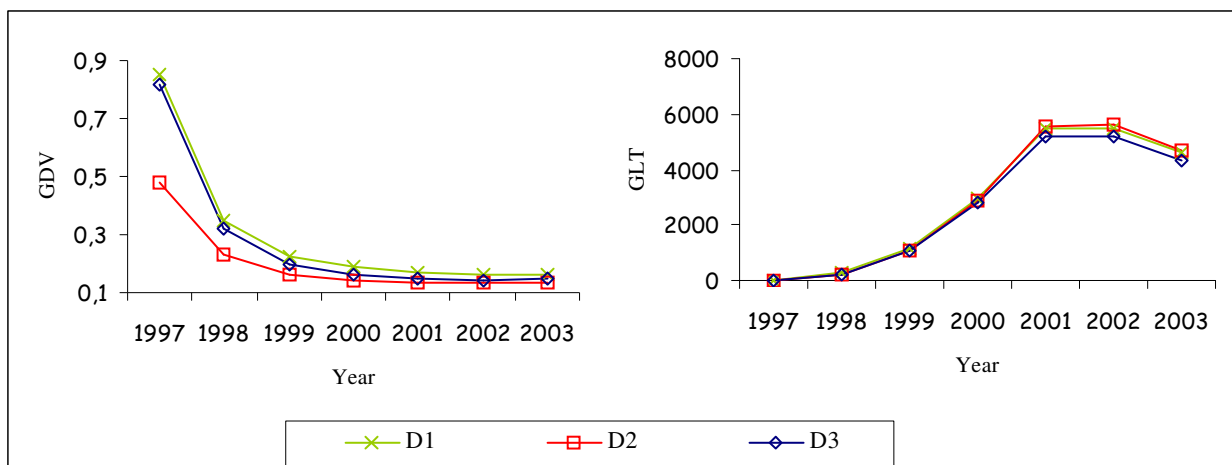
Both methods were applied to 3 datasets: **D1**-containing all herds, **D2**-same as D1, but the small herds (with contemporary groups having less than 4 observations) were clustered based on the simultaneous similarity between phenotypic mean and STD (TD average at 1st lactation) and **D3**-same as D1, but small-herds were removed from the analysis.

RESULTS AND DISCUSSION

Large values of GLT and small values of GDV are preferred, indicating a high level of GC. Results from GLT and GDV for all datasets (Fig. 1) indicate an increase in GC from 1997 to 2001, staying approximately constant towards the end of the period, with a slight decrease in 2003. Table 1 shows the correlations for GLT between all datasets. The observed high correlation obtained in this study suggests that clustering had a negligible effect on GC. The GDV method gave similar

results (Table 1), confirming that clustering may be a reliable procedure to joint herds with few observations, without significant changes on accuracy of prediction.

Fig. 1. Annual average degree of genetic connectedness between management groups (herd/year) on the basis of total number of direct genetic links (GLT) and genetic drift variance (GDV) ^a.



^aD1= contains all herds,

D2= same as D1, but small herds were clustered based on the simultaneous similarity between phenotypic mean and STD,

D3= same as D1, but small herds were removed from the analysis.

The computation of GDV for this study was very costly in CPU-time, especially because of the large number of animals considered in the additive relationship matrix. Kennedy and Trus (1993) obtained a high correlation between GDV and the average PEVd, which, for the purpose of this study, makes the GDV the standard method. Correlations between GDV and GLT were negative, as expected, and ranged from -0.70 to -0.68 . This relatively high correlation supports the use of GLT as a viable method to compute GC in large datasets.

Table 1. Correlations between genetic connectedness in the three datasets (D1, D2 and D3) computed by two methods: total number of direct genetic links (GLT - below diagonal) and genetic drift variance (GDV – above diagonal).

	D1	D2	D3
D1	–	0,9997	0,9997
D2	0,9998	–	0,9996
D3	0,9999	0,9998	–

CONCLUSIONS

Clustering small herds based on the simultaneous similarity of the phenotypic mean and STD, had a negligible effect on GC of the Portuguese Holstein cattle. These results suggest that changes in accuracy of prediction or the systematic use of sires between clusters, may be minimized, but needs confirmation. Although cluster techniques may be a feasible procedure to joint herds with few observations, further studies are necessary to validate their use on genetic evaluations. The next step of this project will evaluate and compare the application of different clustering techniques, their impact on EBV rankings and possible consequences of herds changing clusters between genetic evaluations.

REFERENCES

- Fries, L. A. 1998. Proc. 6th World Cong. Genet. Appl. Livest. Prod., Armindale, NSW, Australia. 27:449-500.
- Foulley, J. L., E. Hanocq, D. Boichard. 1992. Genet. Sel. Evol. 24:315-330.
- Hanocq and Boichard. 1999. Genet. Sel. Evol. 31(2):163-176.
- Kennedy, B. W., and D. Trus. 1993. J. Anim. Sci. 71:2341-2352.
- Roso, V. M., F. S. Schenkel, S. R. Miller. 2004. Canadian J. Anim. Sci. 84(1):37-47.



Effect of Small-Herd Clustering on the Genetic Connectedness of the Portuguese Holstein Cattle Population

J. Vasconcelos^{1,2}, and J. Carvalheira^{1,2}

¹Centro Investigação em Biodiversidade e Recursos Genéticos (CIBIO-ICETA)
Universidade do Porto, Rua Padre Armando Quintas, 4485-661 Vairão, PORTUGAL,
²ICBAS - Universidade do Porto



Accuracy of genetic evaluations is directly related with connectedness between management groups.

Does clustering small-herds (with unacceptable small number of observations in contemporary groups) **affects genetic connectedness and thus, EBV accuracy?**



Data provided by the Portuguese National Association of Dairy Cattle Breeders (ANABLE).

194 980 animals, with records produced from 1997 to 2003.

3 datasets were produced: D1-containing all herds, D2-same as D1, but with small herds (with contemporary groups having less than 4 observations) clustered, based on the simultaneous similarity between phenotypic mean and STD (TD average at 1st lactation) and D3-same as D1, but with small-herds removed from the analysis.

The estimation of PEVd in a data set of this dimension, is computationally infeasible.

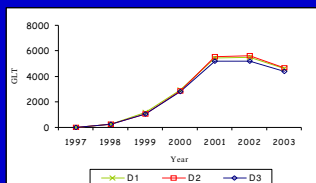
Two alternative methods were applied to compute the degree of genetic connectedness:

GLT (Total number of direct genetic links between management groups; Roso *et al.*, 2004)

Management group defined in this work as herd/year.

GLT was calculated by summing the total number of common sires and dams in pairs of management group, forming genetic links.

For each management group, the overall number of genetic links with all other management group was calculated.

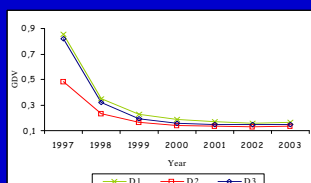


GDV (Genetic Drift Variance; Kennedy and Trus, 1993)

It was obtained from **X'ZAZX** matrix, which measures the sum of genetic relationship between and within management group.

$\bar{\lambda}$ (Average relationship between and within MG) = **X'ZAZX** divided by the squared number of records in the MG, for diagonal elements

$\hat{\lambda}$ = **X'ZAZX** divided by the product of the number of records in each of the management group considered, for off-diagonal elements



$\hat{\lambda}$ can be interpreted as the components of the genetic drift (co)variances.

The average GDV for each management group with all other was calculated as the average of all variance differences with respect to that management group.

	D1	D2	D3
D1	–	0,9997	0,9997
D2	0,9998	–	0,9996
D3	0,9999	0,9998	–

Table 1. Correlations between genetic connectedness in D1, D2 and D3 computed by the two methods: GLT (below diagonal) and GDV (above diagonal).

The computation of GDV for this study was very costly in CPU-time, especially because of the large number of animals considered in the additive relationship matrix.

Kennedy and Trus (1993) obtained a high correlation between GDV and the average PEVd, which, for the purpose of this study, makes the GDV the standard method. Correlations between GDV and GLT were negative, as expected, and ranged from -0.70 to -0.68. This relatively high correlation supports the use of GLT as a viable method to compute GC in large datasets.

Clustering had a negligible effect on genetic connectedness.

Clustering may be a reliable procedure to joint herds with few observations, without significant changes on accuracy of prediction.

These results suggest that changes in accuracy of prediction on the effect of systematic use of sires between clusters, may be minimized.

The next step of this project will evaluate and compare the application of different clustering techniques, their impact on EBV rankings and possible consequences of herds changing clusters between genetic evaluations.

56th Annual EAAP Meeting, Uppsala,
5 - 8 June 2005
Abstract no. 424; Session G4.44



Acknowledgments

JV is financed by a FCT grant
SFRH/BD/11397/2002

Anexo VII*

Consequências da Eliminação de Pequenos Rebanhos da Avaliação Genética de Bovinos Leiteiros em Portugal

*Vasconcelos, J., A. Martins, A. Ferreira e J. Carvalheira. 2005. Consequências da Eliminação de Pequenos Rebanhos da Avaliação Genética de Bovinos Leiteiros em Portugal. XV Congresso de Zootecnia. Vila Real, 3 a 5 de Novembro de 2005, p.160-163.

Consequências da Eliminação de Pequenos Rebanhos da Avaliação Genética de Bovinos Leiteiros em Portugal

J. Vasconcelos^{1,2}, A. Martins³, A. Ferreira³, e J. Carvalheira^{1,2}

¹CIBIO-ICETA, Universidade do Porto. ²ICBAS, Universidade do Porto. ³ABLN, Vila do Conde.
joanav@mail.icav.up.pt

Resumo

A existência de rebanhos com número insuficiente de observações em cada grupo contemporâneo, limita a possibilidade de utilização do rebanho-ano-mês como efeito fixo no modelo de contrastes. Com o objectivo de analisar as consequências da utilização de bases de dados truncadas nas avaliações genéticas Nacionais, foi feita a eliminação progressiva de pequenas explorações, de modo a evidenciar variações e possíveis tendências nos componentes de variância, rigor de predição dos valores genéticos e alterações no rank genético dos animais. Verificou-se que esta eliminação diminui a oportunidade de selecção de alguns reprodutores e a intensidade de selecção, sendo também uma forma de tratamento preferencial de reprodutores machos. Provoca ainda um reordenamento acentuado no rank dos animais, especialmente machos. Conclui-se que ao eliminar rebanhos, incrementamos o enviesamento na avaliação genética.

Palavras-chave: Modelo de contrastes, grupos contemporâneos, efeito rebanho-ano-mês, rebanhos pequenos, avaliação genética.

Introdução

Os modelos de contrastes (MC), ao permitirem a partição dos efeitos ambientais sistemáticos em períodos temporais de curta duração (*e.g.* um mês), conseguem diminuições consideráveis nas variâncias ambientais, permitindo uma maior expressão da componente genética. Todavia, o número de observações em cada grupo contemporâneo (**GC**) torna-se crítico, especialmente em explorações de menor dimensão, onde o número de animais pode ser insuficiente, limitando a possibilidade de utilização do rebanho-ano-mês (**RAM**) como efeito fixo do modelo. O RAM pode, em alternativa, ser considerado como efeito aleatório (Schaeffer e Sullivan, 1994), porém, a necessidade de estimar mais componentes de variância (**CV**) em modelos já por si complexos, pode tornar o processo computacionalmente inviável (Wade, 1990). Por outro lado, a única forma de evitar o enviesamento provocado pela utilização não aleatória de touros por rebanho, é assumir os efeitos ambientais sistemáticos como fixos no modelo (Henderson, 1975; Meyer, 1987; Schaeffer, 1987; Van Vleck, 1987). Há, portanto, vantagens em assumir o efeito dos GC como fixo. Das medidas já propostas, incluem-se técnicas de formação de "clusters" (Schmitz et al., 1991; Carvalheira, 2000; Strabel e Szwaczkowski, 1999). Por outro lado, Swalve (1995), definiu os GC com base no RAM, impondo um limite mínimo de 60 lactações anuais por exploração, o que garantiria pelo menos 3 contrastes por GC. A eliminação de explorações com número reduzido de observações, implica a perda de parte da informação que pode ser significativa no caso da base de dados (**BD**) portuguesa, cuja percentagem de

pequenas explorações é muito elevada. Torna-se assim importante, antes de avaliar em detalhe as consequências da aplicação de técnicas de formação de "clusters", estudar o impacto que a eliminação de pequenos rebanhos terá na avaliação genética Nacional. O objectivo deste estudo centra-se na análise da variação de alguns parâmetros genéticos da população Holstein Nacional, de acordo com a aplicação de diferentes critérios para a determinação do tamanho mínimo dos GC do MC adoptado para fazer a avaliação genética dos caracteres quantitativos nesta espécie.

Material e Métodos

O modelo auto-regressivo difere dos restantes MC por assumir que a expressão da produção de leite ao longo da vida produtiva da fêmea corresponde a uma única característica quantitativa, controlada pelo mesmo grupo de genes, representada no modelo por uma estrutura de covariância não genética, onde as medições repetidas são consideradas como variáveis longitudinais, seguindo processos auto-regressivos de 1ª ordem dentro e entre lactações (Carvalho *et al.*, 1996, 1998, 2002). O modelo adoptado neste estudo, pode ser descrito da seguinte forma:

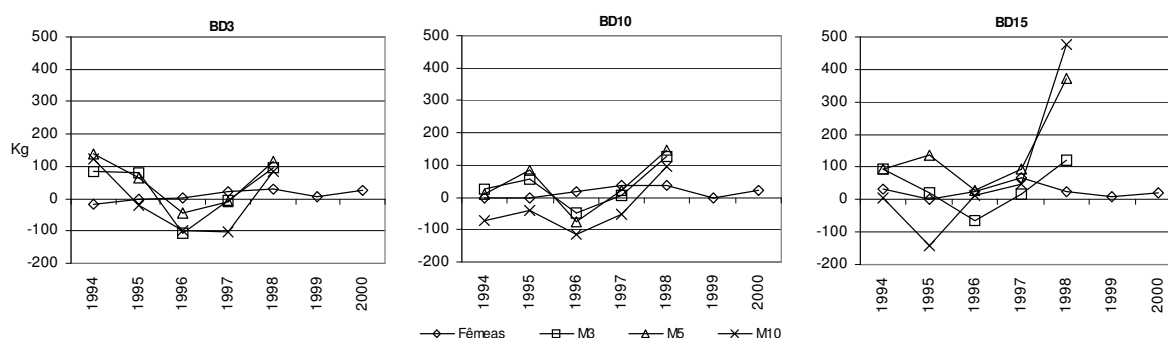
$$y_{ijkLmn} = \text{RAM}_i + \text{Idade}(\mathbf{R})_j + \text{DEL}(\mathbf{R})_{k(L)} + \mathbf{a}_m + \mathbf{p}_{m(L)} + \mathbf{t}_{n(mL)} + r_{ijkLmn},$$

onde, y_{ijkLmn} representa o valor do contraste; RAM_i mede o efeito fixo do rebanho-ano-mês do contraste; $\text{Idade}(\mathbf{R})_j$ mede o efeito fixo hierarquizado por rebanho, correspondente à $j^{\text{ésima}}$ classe de idade ao parto; $\text{DEL}(\mathbf{R})_{k(L)}$ mede o efeito fixo das vacas na mesma classe de dias-em-lactação para cada rebanho e lactação; $\mathbf{a}_m$ representa o efeito aleatório do animal; $\mathbf{p}_{m(L)}$ representa o efeito aleatório do ambiente a longo prazo, medindo a correlação criada pela vaca ao produzir em mais do que uma lactação; $\mathbf{t}_{n(mL)}$ representa o efeito aleatório do ambiente a curto prazo, que mede a correlação entre os contrastes da mesma vaca para cada lactação. Ambos $\mathbf{p}_{m(L)}$ e $\mathbf{t}_{n(mL)}$ seguem processos auto-regressivos de 1ª ordem; r_{ijkLmn} é o termo residual aleatório.

A BD de contrastes do efectivo leiteiro Nacional, com produções entre Junho de 1994 e Dezembro de 2004, foi cedida pela Associação para o Melhoramento dos Bovinos Leiteiros. Durante a edição dos dados [descrita em detalhe em Vasconcelos *et al.* (2001)], foram validadas 3.537.305 observações, distribuídas por 1.310 explorações, com base na produção de 214.618 vacas, filhas de 12.154 touros. A partir da BD editada, foram criadas 3 BD com base nos seguintes critérios: **BD3** - onde explorações com menos de 3 observações por nível de efeitos fixos no modelo foram eliminadas; **BD10** - onde o limite mínimo foi de 10 observações por nível de efeitos fixos e; **BD15** - contendo apenas explorações com pelo menos 15 observações por nível de efeitos fixos. A eliminação progressiva de pequenas explorações, teve como objectivo evidenciar variações e possíveis tendências nos parâmetros, analisando a viabilidade da utilização de BD truncadas nas avaliações genéticas Nacionais. Os parâmetros em estudo incluíram comparações entre CV e autocorrelações (estimados pelo método DFREML-SIMPLEX; Nelder e Mead, 1965), heritabilidades, rigor de predição dos valores genéticos e alterações no rank genético dos animais presentes.

Resultados e Discussão

Os valores obtidos para os CV, autocorrelações e heritabilidades nas 3 BD em estudo, são similares aos previamente estimados em Portugal (Carvalho et al., 2002), usando o mesmo modelo animal. Verificou-se uma ligeira tendência para o aumento nos CV e heritabilidades associada positivamente ao tamanho dos GC que define as 3 BD. A ausência de significância estatística ($P > 0,3$) entre estas diferenças assegura a aleatoriedade do processo de amostragem no sentido de que as 3 BD representam a mesma população. Por este motivo, decidiu-se utilizar a média aritmética entre os 3 estimadores para o cálculo dos restantes parâmetros analisados.



O rigor de predição (RP), definido como a correlação teórica entre os valores genéticos (VGA) verdadeiros (nunca observados) e os VGA estimados, é o principal critério de selecção num programa de melhoramento animal. Apesar dos elevados valores de RP obtidos (médias de 0,868, 0,861 e 0,860 para a BD3, BD10 e BD15, respectivamente), verificou-se uma tendência negativa no RP dos machos à medida que aumenta a restrição do número mínimo de observações por nível de efeitos fixos de 3 para 15. O número total de filhas por touro torna-se assim, num importante factor para a estimativa do RP, fazendo alterar a variância da predição do erro. A eliminação de pequenas explorações tem, portanto, consequências negativas, diminuindo a oportunidade de selecção de alguns reprodutores e a intensidade de selecção. A exclusão de rebanhos mais pequenos é também uma forma de tratamento preferencial de reprodutores machos. Ao conservar apenas explorações de grande dimensão, que geralmente investem mais na genética, estamos, de certo modo, a avaliar maioritariamente touros de elevado potencial genético, obtendo progressos genéticos acima dos reais.

Figura 1. Progresso genético anual do efectivo Nacional Holstein para a produção acumulada de leite aos 305 dias, nas 3 bases de dados em estudo (**BD3**, **BD10** e **BD15** = 3, 10 e 15 ou mais observações por grupo contemporâneo, respectivamente; **M3**, **M5** e **M10** = machos com pelo menos 3, 5 e 10 filhas respectivamente).

Este aspecto é claramente demonstrado na Figura 1. Enquanto que o progresso genético nas fêmeas é praticamente igual nas 3 BD, o mesmo não se verifica nos machos, onde o impacto do tratamento preferencial é bastante acentuado. Na sequência destas alterações, a exclusão de GC induz a um reordenamento no rank de VGA dos animais, especialmente no caso dos machos. Com o objectivo de verificar este reordenamento, foi calculada a percentagem de

animais do top 100 do VGA em comum nas 3 BD. Relativamente à BD3, a percentagem de animais em comum com a BD10 e com a BD15 é inferior a 21%. As alterações no rank são menos evidentes nas fêmeas do que nos machos, pois nestas a avaliação é baseada na sua própria produção, enquanto que nos machos a estimativa do seu VGA é baseada na avaliação do genótipo das suas filhas, cujo número depende do tamanho mínimo dos GC adoptado. Estas alterações foram confirmadas pelos valores de correlação de rank dos VGA obtidos nas 3 BD para os animais em comum. No caso das fêmeas, os elevados valores de correlação (entre 0,82 e 0,83) sugerem pequenas alterações no VGA de uma BD para outra. Para os machos, os valores de correlação foram mais baixos (entre 0,21 e 0,32), evidenciando o facto da exclusão de GC provocar um reordenamento acentuado no rank dos VGA.

Conclusões

Ao eliminar explorações, truncando a BD Nacional, estamos a incluir um erro na avaliação genética que pode ter consequências dramáticas na selecção de reprodutores, influenciando negativamente o progresso genético da população. Agrupar pequenos GC apresenta-se como uma das soluções viáveis para o problema, pois apenas deste modo se pode garantir simultaneamente a participação de todos os rebanhos e um número mínimo de animais contemporâneos. Quanto maior for a diversidade dentro dos grupos, menor será a variância da predição do erro e consequentemente, maior o rigor de predição. O passo seguinte neste projecto será o estudo de técnicas de agrupamento de GC com base no RAM.

Referências

- Carvalho, J., 1996. Doctoral Dissertation, Cornell Univ., Ithaca, NY.
- Carvalho, J., 2000. X Congresso de Zootecnia, EZN, Vale de Santarém, 2000:103.
- Carvalho J., R. W. Blake, E. J. Pollak, R. L. Quaas, and C. V. Duran-Castro, 1998. J. Dairy Sci. 81:2738-2751.
- Carvalho J., E. J. Pollak, R. L. Quaas and R. W. Blake, 2002. J. Dairy Sci. 85:2040-2045.
- Henderson, C. R., 1975. J. Anim. Sci. 41:760-770.
- Meyer, K., 1987. Livest. Prod. Sci. 17:95.
- Nelder, J. A. and R. Mead, 1965. Computer J. 7: 308-313.
- Schaeffer, L., 1987. J. Dairy Sci. 70: 661.
- Schaeffer, L., and B. Sullivan, 1994. Proc. 5th World Cong. Gen. Appl. Livest. Prod. 18:182-185.
- Schmitz, F., R. W. Everett, and R. L. Quaas, 1991. J. Dairy Sci. 74: 629-636.
- Strabel, T., and T. Szwaczkowski, 1999. J. Anim. Breed. Genet. 116, 379-386.
- Swalve, H. H., 1995. J. Dairy Sci. 78: 929-938.
- Van Vleck, L., 1987. J. Dairy Sci. 70: 2456.
- Vasconcelos, J., M.F.Petim-Batista, A. Silvestre, A. Ferreira, A. Martins, J. Colaço e J. Carvalho, 2001. XI Congresso de Zootecnia, Madeira, 2001:105-106.
- Wade, K. M., 1990. Doctoral Diss., Cornell Univ. Ithaca, NY, pp 117.



Consequências da Eliminação de Pequenos Rebanhos da Avaliação Genética de Bovinos Leiteiros em Portugal

J. Vasconcelos^{1,2}, A. Martins³, A. Ferreira³, e J. Carvalheira^{1,2}

¹Centro Investigação em Biodiversidade e Recursos Genéticos (CIBIO-ICETA) Universidade do Porto, Rua Padre Armando Quintas, 4485-661 Vairão, PORTUGAL, ²ICBAS - Universidade do Porto; ³ABLN, Vila do Conde



A existência de rebanhos com número insuficiente de observações em cada grupo contemporâneo, limita a possibilidade de utilização do rebanho-ano-mês como efeito fixo no modelo de contrastes.

Quais as consequências da eliminação de pequenas explorações da avaliação genética



O modelo adoptado neste estudo, pode ser descrito da seguinte forma:

$$Y_{ijklmn} = \text{Valor do contraste} = \text{RAM}_i + \text{Idade(R)}_j + \text{DEL(R)}_{k(L)} + \alpha_m + \text{P}_{n(L)} + \tau_{n(L)} + \text{Termo residual aleatório}$$

RAM_i: Efeito fixo do rebanho-ano-mês
 Idade(R)_j: Efeito fixo da idade ao parto
 DEL(R)_{k(L)}: Efeito fixo dos dias-em-lactação
 α_m: Efeito aleatório do animal
 P_{n(L)}: Efeito aleatório do ambiente e curto prazo
 τ_{n(L)}: Efeito aleatório do ambiente e longo prazo
 Termo residual aleatório

Tabela 1. Componentes de variância (kg²) e autocorrelações para a produção diária de leite no efectivo Holstein Nacional para as primeiras 3 lactações (L₁, L₂ e L₃) nas 3 bases de dados (BD3, BD10 e BD15).

Parâmetro	BD3	BD10	BD15
Variância genética	7,9363	8,2256	8,7843
Variância ambiental	3,7009	3,9786	4,4478
Variância ALP ^a	<0,0001	<0,0001	<0,0001
Autocorrelação ALP	<0,0001	<0,0001	<0,0001
Variância ACP ^b (L ₁)	12,5742	13,5512	13,0997
Autocorrelação ACP (L ₁)	0,8412	0,8250	0,8294
Variância ACP (L ₂)	24,7861	25,8795	26,2600
Autocorrelação ACP (L ₂)	0,8327	0,8164	0,8200
Variância ACP (L ₃)	29,3067	30,4231	30,5559
Autocorrelação ACP (L ₃)	0,8241	0,8044	0,8003
Heritabilidade (L ₁)	0,3278	0,3194	0,3336
Heritabilidade (L ₂)	0,2179	0,2160	0,2224
Heritabilidade (L ₃)	0,1930	0,1930	0,2006

^aALP = Efeito Ambiental a Longo Prazo.
^bACP = Efeito Ambiental a Curto Prazo.

Os Valores de componentes de variância (CV) obtidos foram similares aos previamente estimados em Portugal, usando o mesmo modelo animal.

Verificou-se uma ligeira tendência para o aumento nos CV e heritabilidade, associado positivamente ao tamanho dos grupos contemporâneos que definem as 3 BD. A ausência de significância estatística (p>0,3) entre estas diferenças valida o processo aleatório de amostragem, no sentido das 3 BD representarem a mesma população.

Por este motivo, decidiu-se utilizar a média aritmética entre os 3 estimadores para o cálculo dos restantes parâmetros analisados.

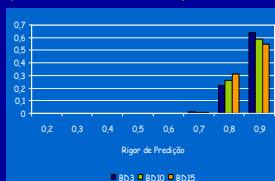


Figura 1. Frequência de valores do rigor de predição (RP) estimados em machos com 15 ou mais descendentes, em pelo menos 5 explorações, nas 3 bases de dados em estudo.

Apesar dos elevados valores de rigor de predição (RP) obtidos (médias de 0,868, 0,861 e 0,860 para a BD3, BD10 e BD15, respectivamente), verificou-se uma tendência negativa no RP dos machos da BD3 para a BD15.

O número total de filhas por touro é um importante factor para a estimativa do RP, fazendo alterar a variância da predição do erro.

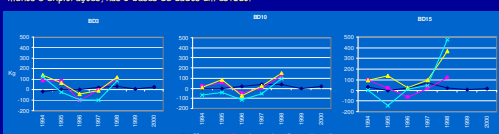


Figura 2. Progresso genético anual do efectivo Nacional Holstein para a produção acumulada de leite aos 305 dias, nas 3 bases de dados em estudo (M3, M5 e M10 = machos com pelo menos 3, 5 e 10 filhas respectivamente).

Resultados e Discussão

A exclusão de rebanhos pequenos é também uma forma de tratamento preferencial de reprodutores machos. Ao conservar apenas explorações de grande dimensão, que geralmente investem mais na genética, estamos, de certo modo, a avaliar maioritariamente touros de elevado potencial genético, obtendo progressos genéticos acima dos reais (Figura 2).

Enquanto que o progresso genético nas fêmeas é praticamente igual nas 3 BD, o mesmo não se verifica nos machos, onde o impacto do tratamento preferencial é bastante acentuado, resultando num incremento artificial do seu valor genético.

Tabela 2. Lista dos geneticamente 10 melhores animais nas 3 bases de dados em estudo. *animais em comum nas 3 bases de dados representados da mesma cor.

Ranking	Machos			Fêmeas		
	BD3	BD10	BD15	EBV3	EBV10	EBV15
1	875531	616143	169117	832024	872987	501393
2	26676	709771	309137	782647	783242	596715
3	709771	364910	1022843	526378	834592	834592
4	633788	875531	216819	565022	501393	773560
5	200133	111330	364291	653428	565022	856449
6	713843	530611	155730	804910	653428	832024
7	227807	215927	388763	477853	1067455	597146
8	878655	620028	531645	501393	570613	475229
9	67548	67548	713687	866800	969399	373476
10	1085933	146361	77111	703292	490978	792701

A Tabela 2 dá uma indicação deste reordenamento, evidenciando com a mesma cor os animais em comum nas 3 BD, para o Top10 de cada sexo.

Tabela 3. Coeficientes de correlação entre o ranking genético para machos (acima da diagonal) e fêmeas (por baixo da diagonal), nas 3 bases de dados em estudo.

	BD3 ^a	BD10	BD15
BD3	--	0,32	0,21
BD10	0,83	--	0,28
BD15	0,82	0,83	--

As correlações de ranking entre os valores genéticos obtidos (Tabela 3) confirmam o reordenamento, sendo mais marcado nos machos do que nas fêmeas.

Com correlações de ranking tão baixas, torna-se claro que a exclusão de rebanhos deve ser limitada ao mínimo e o tamanho dos grupos contemporâneos deve ser garantido recorrendo a outros procedimentos (e.g., técnicas de agrupamento).

Conclusões

Os resultados obtidos neste estudo comprovam que a exclusão de rebanhos de pequena dimensão da avaliação genética pode ter consequências dramáticas na selecção de reprodutores, influenciando negativamente o progresso genético da população.

Se por um lado é necessário garantir um número mínimo de observações em cada grupo contemporâneo, por outro lado, há que maximizar o número de filhas e de rebanhos que contribuem para a avaliação dos touros.

Agrupar rebanhos pode ser uma solução para este problema.

Agradecimentos

Trabalho parcialmente financiado pela FCT através da bolsa: SFRH/BD/11397/2002

FCT Fundação para a Ciência e a Tecnologia

Anexo VIII*

Effects of Clustering Dairy Herds for Genetic Evaluations using Different Descriptors to
Define Similarities between Production Environments

*Vasconcelos, J., F. Santos, R. Barroso, A. Martins, A. Ferreira, J. Carvalheira. 2006. Effects of Clustering Dairy Herds for Genetic Evaluations using Different Descriptors to Define Similarities between Production Environments. 8th World Congress on Genetics Applied to Livestock Production. 2006. Proc. 8th World Cong. Gen. Appl. Livest. Prod., CD-ROM communication n° 24-30.

EFFECTS OF CLUSTERING DAIRY HERDS FOR GENETIC EVALUATIONS USING DIFFERENT DESCRIPTORS TO DEFINE SIMILARITIES BETWEEN PRODUCTION ENVIRONMENTS

J. Vasconcelos^{1,2}, F. Santos¹, R. Barroso¹, A. Martins³, A. Ferreira³, and J. Carvalho^{1,2}

¹CIBIO, Research Center in Biodiversity and Genetic Resources - University of Porto, Rua Padre Armando Quintas-Crasto, 4485-661 Vairão, Portugal

²Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Portugal

³ABLN, Vila do Conde, Portugal.

INTRODUCTION

In genetic evaluations, animal performances are compared within contemporary groups (**CG**) of assumed similar environmental conditions (Schmitz *et al.*, 1991). The autoregressive test day animal model (**AR**) fits multiple lactation data and defines the systematic environmental effects using 3 fixed factors: herd-test-dates (**HTD**), age at calving within herd and days-in-milk within herd and lactation (Carvalho *et al.*, 2002). For each level of these effects, a minimum number of observations (e.g., 3) are required in order to minimize residual variances and thus, prediction bias. With this criterion, small and recent farms (about 40% of all dairy farms in Portugal) may be excluded from the evaluations unless we devise some procedure to join them while minimizing the impact that the change in the environmental and genetic structure of the data may bring to the genetic evaluations. Some authors suggested clustering these herds either joining HTD classes within the herd (Schmitz *et al.*, 1991; Swalve 1995; Carabaño *et al.*, 2004), or between herds (Carvalho, 2000), or a combination of both (Strabel and Szwaczkowski, 1999). If management information is also a required output of the evaluation process, the 12 month classes per year should be maintained and, therefore, clustering should be performed among herds (Carvalho, 2000).

Recent genetic evaluations in Portugal have been done using this latter approach. However, the criteria for management similarities between herds have been based only on the phenotypic daily mean (milk, fat and protein) and STD at first lactation. Also, the clustering process was only stopped when all the formed clusters had the minimum number of obs. per level of fixed effects, meaning that parameters such as size and the maximum number of herds forming a new cluster were not controlled. The major implication of this approach was that the final number of herds forming a cluster depended only on the tree organization defined by the distance matrix.

Therefore, the aim of the present work was to study the consequences of including new descriptors of production environments to define similarities between herds using principal components (**PC**) analyses. To improve the control of the clustering process, additional parameters were introduced in order to minimize the number of herds per cluster (maximizing the final number of comparison groups), while ensuring the minimum number of obs. per CG.

MATERIAL AND METHODS

The Portuguese National Association of Dairy Cattle Breeders (**ANABLE**) provided the database. After editing (details in Vasconcelos *et al.*, 2005a), the data comprised 3,679,869 test day (**TD**) records, produced by 222,480 Holstein cows in the first three lactations, in 1,315 herds, recorded between June 1994 and May 2005. Forty percent (517) of these herds had at least one CG with less than 3 observations. Two complete data sets were built differing on the methodology used to group the farms. For data set one (**D1**), these farms were grouped according to their similarities based only on the phenotypic mean and STD at first lactation, as described before. The clustering methodology used on data set two (**D2**) was more complex and was as follows:

Descriptive variables for D2. Fifteen descriptive variables (**DV**) were considered as potential indicators of production environments for each herd: herd size (**HS**) – number of cows calving during the study period; region (**R**); age at first calving (**AFC**); peak yield (**PY**) - average of the maximum TD yield of each cow in the herd; peak yield variability (**PYV**) - standard deviation of PY; production level (**PL**) - average daily milk yield; and production variability (**PLV**) - standard deviation of PL. The last 4 variables (PY, PYV, PL and PLV) were calculated for all 3 lactations in each herd.

PC analyses. DV variables were grouped into a reduced set of variables using PC analyses after the estimation of all missing values by qualitative PC analysis using the maximum total variance method (PROC PRINQUAL; SAS, 1989). This procedure also standardized the original DV to zero mean and unit variance. Only components, which individually explained more than 7% of the total variance in all variables, were used for further studies. The analyses resulted in 3 PC vectors explaining 0.9999 % of the total variance (PROC PRINCOMP; SAS, 1989), which were retained for subsequent cluster analysis.

Cluster analyses. The first step consisted in building the herd's distance matrix based on the 3 PC computed previously. The method applied to define similarity between pairs of herds was the Ward's minimum-variance (PROC CLUSTER; SAS, 1989). Herds were then clustered according to the following criteria:

- Herds were sorted according to the distances, so that adjacent herds in the list represent the closest ones in terms of similarity in management.
- Cluster formation had to start with herds having less than 3 obs. in at least one CG.
- These herds were joined following the branches of the tree, which reflects the shortest distance between them.
- After each agglomeration, the program recounts all obs. at each CG and stops the process if all of them have the minimum number. If not, it will continue the clustering process, firstly checking if there are herds available at lower levels in the tree or, otherwise, it will go up for the parent node and starts looking for candidates on the other descending branch.
- If there is more then one herd at the same distance, priority to form a new cluster was given to the smaller one. This criterion permitted the formation of smaller clusters.

The clustering process was performed using software specifically written for this study (Santos, F., 2006, personal communication). After the clustering process, both data sets were analyzed using the AR animal model and results were compared to detect differences between breeding values (genetic rank and progress) and accuracies of prediction.

RESULTS AND DISCUSSION

A first run on PC analysis revealed that R, after HS, did not have a significant contribution for the total variance and therefore, was discarded from the DV list. This was in agreement with the results presented by König et al., (2005). The remaining DV were reduced to 3 PC variables explaining almost 100 % of the total environmental variation (64.3 %, 28.6 % and 7.1 %, respectively). The most important DV contributors were production variables (PY, PL, PYV and PLV). Windig et al. (2005) also found that variables related to average production per cow were the most important environmental variables since they can explain herd ranking according to intensity of production.

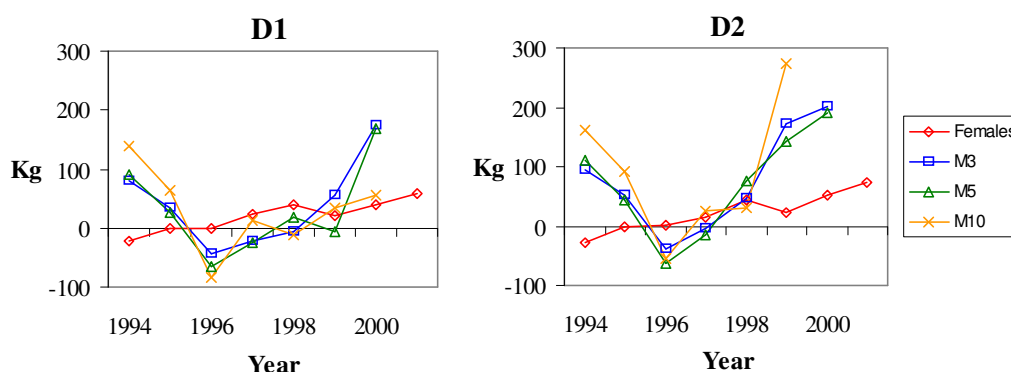
The simultaneous minimization of the number of herds and TD records forming a cluster was one of the aims of this study. Considering that the algorithm applied in D2 uses as main criteria to join for herds, distance and not size, 197 big farms (none in D1) were also used to form clusters. Although the higher number of herds involved in cluster formation in D2 (Table 1), the average number of TD records per cluster were significantly smaller, representing more comparison groups used in the evaluations.

Table 1. Number of herds and clusters in data sets D1 and D2.

	D1	D2
No. of herds	1315	1315
No. of clusters	14	186
Herds clustered	517	714
Average TD/Cluster	38,270	5,594

Breeding value rank correlation was high (0.953 for males and 0.950 for females) suggesting that the clustering procedure per se, did not have a direct influence on genetic rank. The high percentage of animals in common in the top100 list (86 sires with more than 10 daughters and 67 females) is a confirmation of this result. Comparisons between the prediction error variance and the effective progeny number also resulted in non-significant differences between the two clustering techniques ($p > 0.4$ and $p > 0.6$, respectively). Studying the effect of clustering on genetic connectedness, Vasconcelos et al. (2005b) also arrived to a similar conclusion.

On the other hand, trends of realized genetic progress by year of birth for the two data sets (Figure 1) shows that males in D2 had a faster genetic progress relatively to D1. Genetic progress for females was practically the same. This result indicates that breeding values of males in D2 were higher, probably because the greater number of comparison groups in this data set allowed for a better distinction between their daughters. These results also imply that more research is necessary to confirm that there are no systematic



associations between sires and environments.

Figure 1. Genetic progress by year of birth for data sets D1 and D2. Male trends are for sires with more than 3, 5 and 10 daughters (M3, M5 and M10, respectively).

CONCLUSION

Defining herd management by reducing a large number of environmental descriptors using principal components analyses was a viable approach to cluster herds with insufficient contemporaries per level of fixed effects in the analytical model. More clusters permitted more comparison groups and, therefore, a better contrast between daughters, which resulted in faster genetic progress per year of sires. More research is needed to evaluate possible systematic associations between sires and environments.

REFERENCES

- Carabaño, M. J., et al. (2004) *J. Anim. Sci.* **82** : 3447-3457.
 Carvalheira, J. (2000) *X Congresso de Zootecnia* p103.
 Carvalheira J., Pollak, E. J., Quaas R. L. and Blake, R. W. (2002) *J. Dairy Sci.* 85:2040-2045.
 König, S., Dietl, G., Reader, I., and Swalve, H. H. (2005) *J. Dairy Sci.* 88: 4087-4096.
 SAS Institute Inc., SAS/STAT® User's Guide, Version 6, Volume 1. (1989) p 535.
 Schmitz, F., Everett, R. W. and Quaas, R. L. (1991) *J. Dairy Sci.* 74 : 629-636.

- Strabel, T. and Szwaczkowski T. (1999) *J. Anim. Breed. Genet.* 116 : 379-386.
- Swalve, H. H. (1995) *J. Dairy Sci.* 78 : 929-938.
- Vasconcelos, J., Ferreira, A., Martins, A. and Carvalheira, J. (2005a) XIV Congresso de Zootecnia 160-163.
- Vasconcelos, J. and Carvalheira, J. (2005b) EAAP – 56th Annual Meeting. Abstract no. 424.
- Windig, J. J., Calus, M. P. L. and Veerkamp, R. F. (2005) *J. Dairy Sci.* **88** : 335-347.



Effects of Clustering Dairy Herds for Genetic Evaluations Using Different Descriptors to Define Similarities Between Production Environments



J. Vasconcelos^{1,2}, F. Santos¹, R. Barroso¹, A. Martins³, A. Ferreira³ and J. Carvalheira^{1,2}

¹CIBIO, Centro de investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, PORTUGAL, ²ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, 4099-003 Porto, Portugal; ³ABLN, Vila do Conde, Portugal

In genetic evaluations, animal performances are compared within contemporary groups (CG) of assumed similar environmental conditions.

The autoregressive test day animal model (AR) fits multiple lactation data and defines the systematic environmental effects using 3 fixed factors: HTD, Age(H) and DIM(H) _{ijklmn}:

$$Y_{ijklmn} = \text{HTD}_{ij} + \text{Age(H)}_k + \text{DIM(H)}_{kl} + a_m + P_{n(l)} + t_{n(ml)} + e_{ijklmn}$$

Fixed Effect of Age at Calving Nested within Herd

Fixed Effect of DIM Nested with Herd and L

Random animal eff.

Random long term environmental effects

Random residual eff.

Fixed Effect of Herd-Test-Date

For each level of these effects, a minimum number of observations (e.g., 3) are required in order to minimize residual variances and thus, prediction bias.

Random effect of short term environmental effects

The aim of the work was to study the consequences of including new descriptors of production environments, to define similarities between herds using principal components (PC) analyses.

Two datasets were produced:

D1 -farms were grouped according to their similarities based only on the phenotypic mean and STD at first lactation.

D2 -farms were grouped according to 15 descriptive variables: herd size, region, age at first calving, peak yield, peak yield variability, production level, and production variability (calculated for all 3 lactations).

The PC analyses resulted in 3 PC vectors explaining 0.9999 % of the total variance (PROC PRINCOMP; SAS, 1989), which were retained for subsequent cluster analysis.

To improve the control of the clustering process, additional parameters were introduced to minimize the number of herds per cluster, while ensuring the minimum number of obs. per CG.

Herds distance matrix was based on the 3 PC (Ward's minimum-variance method).

Criteria to cluster herds:

- Herds sorted according to the distances, so that adjacent herds in the list represent the most similar ones in terms of management.
- Cluster formation had to start with herds having less than 3 obs. in at least one CG.
- These herds were joined following the branches of the tree, which reflects the shortest distance between them.
- After each agglomeration, the program recounts all obs. at each CG and stops the process if all of them have the minimum number. If not, it will continue the clustering process, firstly checking if there are herds available at lower levels in the tree or, otherwise, it will go up for the parent node and starts looking for candidates on the other descending branch.
- If there is more than one herd at the same distance, priority to form a new cluster was given to the smaller one. This criterion permitted the formation of smaller clusters.

The DV were reduced to 3 PC variables explaining almost 100 % of the total environmental variation (64.3 %, 28.6 % and 7.1 %, respectively). The variables related to average production per cow were the most important environmental variables since they can explain herd ranking according to intensity of production.

Defining herd management by reducing a large number of environmental descriptors using principal components analyses was a viable approach to cluster herds with insufficient contemporaries per level of fixed effects in the analytical model.

Although the higher number of herds involved in cluster formation in D2 (Table 1), the average number of TD records per cluster were significantly smaller, representing more comparison groups used in the evaluations.

More clusters permitted more comparison groups and, therefore, a better contrast between daughters, which resulted in faster genetic progress per year of sires.

Table 1. Number of herds and clusters in data sets D1 and D2.

	D1	D2
Nº Herds	1315	1315
Nº Clusters	14	186
Herds clustered	517	714
Average TD/Cluster	38,270	5,594

After the clustering process, both data sets were analyzed using the AR animal model and results were compared.

Breeding value rank correlation was high (0.953 for males and 0.950 for females) suggesting that the clustering procedure per se, did not have a direct influence on genetic rank.

The high percentage of animals in common in the top100 list (86 sires with more than 10 daughters and 67 females) is a confirmation of this result.

Comparisons between the prediction error variance and the effective progeny number also resulted in non-significant differences between the two clustering techniques ($p > 0.4$ and $p > 0.6$, respectively).

Trends of realized genetic progress by year of birth shows that males in D2 had a faster genetic progress relatively to D1, probably because the greater number of comparison groups in D2, allowing for a better distinction between their daughters.

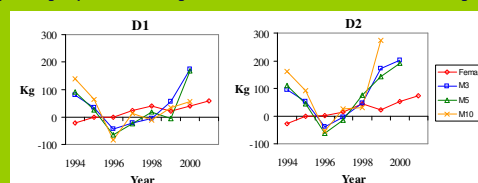


Figure 1. Genetic progress by year of birth for data sets D1 and D2. Male trends are for sires with more than 3, 5 and 10 daughters (M3, M5 and M10, respectively).

Anexo IX*

Assessment of heterogeneity of residual variances in an autoregressive test-day
multiple lactations model

*Vasconcelos, J., and J. Carvalheira. 2007. Assessment of heterogeneity of residual variances in an autoregressive test-day multiple lactations model. EAAP – 58th Annual Meeting, Dublin – Ireland. 26 – 29 of August 2007, Session 10, Poster 39.

Assessment of Heterogeneity of Residual Variances in an Autoregressive Test-Day Multiple Lactations Model. J. Vasconcelos, A. Martins, A. Ferreira and J. Carvalheira.

The autoregressive test-day animal model (ARTD) is a class of TD models where the within animal covariance follows 1st order autoregressive processes. Animal correlations are fitted within and across lactations (L) using short (STE) and long (LTE) term environmental effects. TD yields are treated as random variables with heterogeneous (HE) variance for each L and describe STE effects. LTE effect accounts for the animal correlation across L. Results from a previous study suggested that the usual assumption of a homogeneous (HO) residual variance (RV) component may not hold. To assess HE of RV across L, 5 random samples of the Portuguese Holstein milk database were extracted and analyzed by ARTD models differing on the RV structure (HO vs. HE). Likelihood ratio tests indicated a better fit ($P < .01$) for models with HE RV. Heritabilities were similar, averaging .32, .21 and .19 for 1st, 2nd and 3rd L, respectively. Correlations between the animal's EBV, PEV, accuracy of prediction (ACC) and rank were high ($>.99$) suggesting that HE residuals will not change the genetic rank of sires and cows. Differences among EBV were on average not significant ($P > .05$). On the other hand, differences among PEV and ACC ($P < .001$) favoured the HE RV model, implying greater potential for genetic progress. These advantages (better fit and higher ACC) are going to be explored in future genetic evaluations in Portugal.

Assessment of Heterogeneity of Residual Variances in an Autoregressive Test-Day Multiple Lactations Model



J. Vasconcelos, A. Martins, A. Ferreira and J. Carvalheira



Introduction

• Numerous studies (Meuwissen and Van der Werf, 1993; Ibáñez et al., 1999; Rekaya et al., 2000; López-Romero et al., 2003) have reported heterogeneous (HE) genetic, residual and permanent environmental variances for production traits.

• Ignoring heterogeneity of variances may affect genetic progress (Robert-Granié et al., 1999; Ibáñez et al., 1999; Rekaya et al., 2000; López-Romero et al., 2003).

• Most studies using TD animal models recommend that HE sources of variance should be accounted with especial concern to residual variances (Rekaya et al., 2000; López-Romero et al., 2003).

• López-Romero et al. (2003) studying first lactation TD records with a random regression model, found significant reductions in the residual variances (RV) at the beginning and at the end of the lactation when a 5th order polynomial sub model was fitted to the permanent environment (PE) component.

• They concluded that allowing for heterogeneous RV only improves the adequacy of the model to the observed or predicted data for models fitting a low order to the PE sub model.

Objectives

• To assess heterogeneity of RV across L, 5 random samples of the Portuguese Holstein milk database were extracted and analyzed by autoregressive test-day animal (ARTD) models differing on the RV structure (HO vs. HE).

Data

Table 1. Characteristics of the data sets

	Sample				
	1	2	3	4	5
No. of Animals	13,270	10,743	9,509	10,239	9,408
Animals w/records	7,625	5,757	5,155	5,161	4,055
TD observations	127,407	97,342	91,010	78,014	88,163
1st Lactation:					
Animals w/records	5,542	4,843	4,355	4,189	4,257
TD observations	56,811	43,317	37,002	34,537	37,348
Average TD/Animal	9	9	9	8	9
2nd Lactation:					
Animals w/records	5,098	3,858	3,581	3,256	3,519
TD observations	42,916	32,995	30,675	26,514	30,362
Average TD/Animal	8	9	9	8	9
3rd Lactation:					
Animals w/records	3,120	2,644	2,555	2,254	2,441
TD observations	27,680	22,000	22,343	16,743	20,453
Average TD/Animal	8	8	8	8	8

The ARTD Model

• The autoregressive test-day/animal model (ARTD) is a class of TD models where the within animal covariance (PE effects) follows 1st order autoregressive processes and are fitted within and across lactations (L) using short (STE) and long (LTE) term environmental effects (Carvalheira et al., 2002).

• STE effects are fitted with HE variance for each L and describe animal correlations between TD within each L.

• LTE effect accounts for the animal correlation across L.

• Results from a previous study suggested that the usual assumption of a homogeneous (HO) residual variance component may not hold (Carvalheira, 2006).

• Multiple lactations single trait animal model,

$$y_{ijklmn} = \text{HTD} + \text{Age}_i + \text{DIM}_{Lk} + a_m + p_{Lm} + t_{Lmn} + e_{ijklmn}$$

- HTD - Hand TD,
- Age - Age at calving,
- DIM - days in milk (80 classes),
- a - Genetic,
- p - LTE, Long term environmental effects, AR(1),
- t - STE, Short term environmental effects, AR(1),
- e - residuals, Temporary environmental effects.

Results and Discussion

Table 2. Likelihood Ratio Tests (Heterogeneous vs. Homogeneous RV)

Sample	LRT	P
1	77.599	< 0.0005
2	9.510	< 0.01
3	124.879	< 0.0005
4	37.691	< 0.0005
5	62.531	< 0.0005

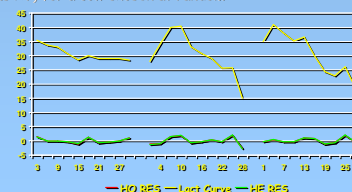
• Likelihood Ratio Tests indicate a better fit for models with Heterogeneous RV.

Table 3. (Co)variance components (kg²), autocorrelations and heritabilities

Parameter	Average Components	
	Heterogeneous RV	Homogeneous RV
var(G)	7.942	7.901
var(LTE)	<0.001	<0.001
corr(LTE)	<0.001	<0.001
var(e1)	3.572	3.967
var(e2)	4.448	
var(e3)	4.989	
var(STE1)	12.970	12.793
corr(STE1)	0.805	0.823
var(STE2)	25.534	25.837
corr(STE2)	0.834	0.822
var(STE3)	28.862	29.518
corr(STE3)	0.837	0.812
Heritability(L1)	0.324	0.322
Heritability(L2)	0.209	0.211
Heritability(L3)	0.190	0.192

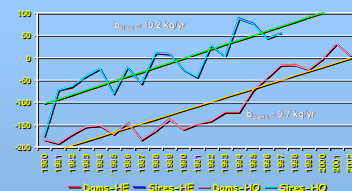
• Heritabilities were similar for the 2 models and the range was within the ones reported in the literature.

Figure 1. First 3 Lactations and Residuals (Heterogeneous vs. Homogeneous RV) for a cow chosen at random



• Residuals in both models never exceeded ± 3 kg.

Figure 2. Genetic Progress (Heterogeneous vs. Homogeneous RV)



• Genetic progress was similar between models with sires and dams having a positive and almost parallel slope.

Table 4. Correlations (Heterogeneous vs. Homogeneous RV)

	Heterogeneous			
	EBV	PEV	ACC	Rank
EBV	0.999	-	-	-
PEV	-	0.999	-	-
ACC	-	-	0.999	-
Rank	-	-	-	0.999

• Correlations between the animal's EBV, PEV, ACC and Rank were high suggesting that HE RV will not change the genetic rank of sires and cows.

Table 5. Average Differences (Heterogeneous vs. Homogeneous RV)

Parameter	N	Mean	STD	P
EBV	53,169	-0.0003	0.0344	= 0.06
PEV	337	-0.0021	0.0077	< 0.0001
ACC	337	0.0009	0.0007	< 0.0001

• Differences between EBV were on average not significant. On the other hand, differences among PEV and ACC favoured the HE RV model, implying greater potential for genetic progress.