

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

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

*Research Center in Biodiversity and Genetic Resources (CIBIO/ICETA), 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 are, respectively, the prediction error variance and the inbreeding coefficient for sire i , 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 (best linear unbiased estimator) 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 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;

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.

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).

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.

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). Except for males in A₁₅, trends were 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

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.

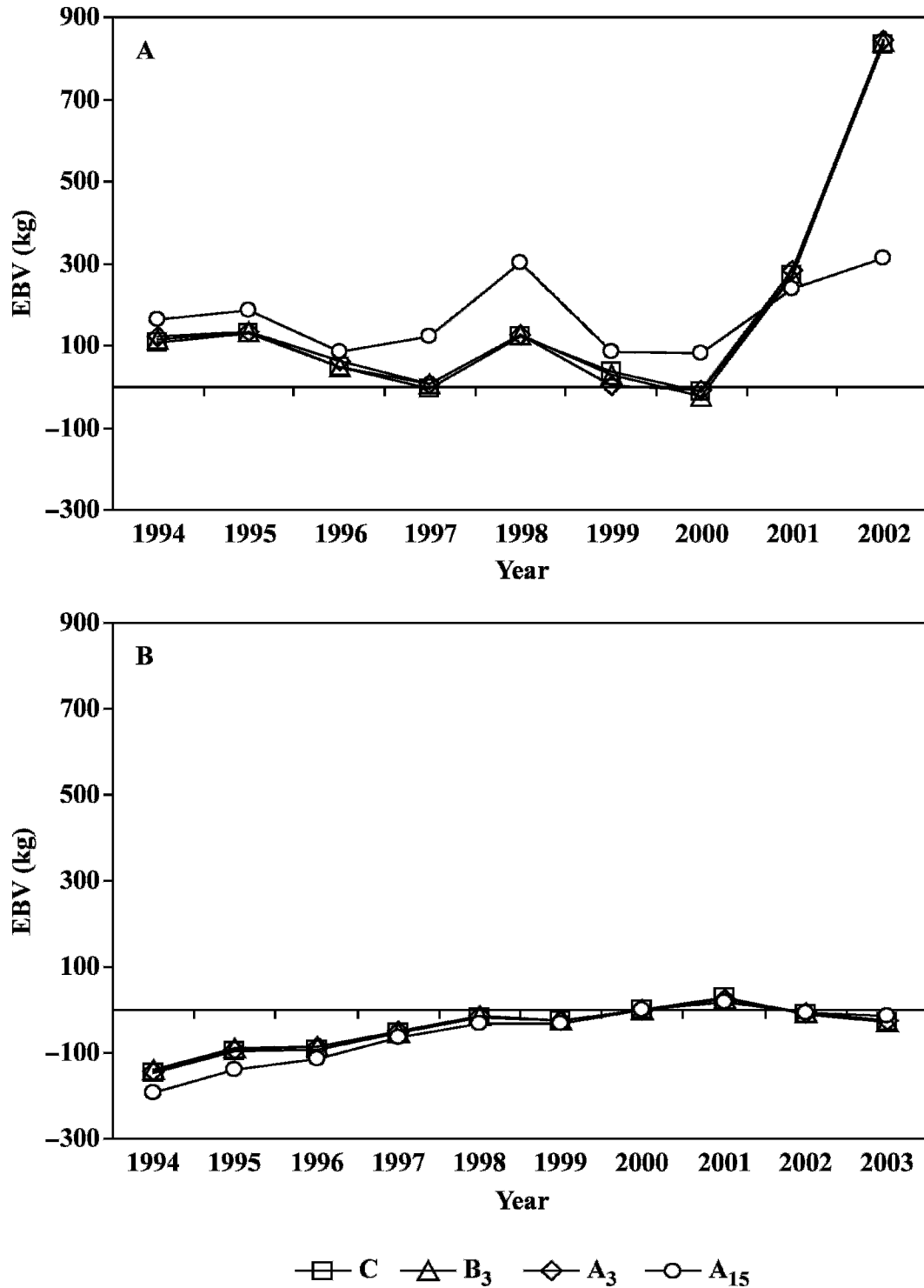


Figure 1. Milk genetic progress for sires (A) with 10 or more daughters and dams (B) by year of birth in data sets C, B₃, A₃, and A₁₅. In data set C, only herds with less than 3 observations per contemporary group (CG) were candidates to form new clusters, and management similarities among herds were 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.

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 evaluations 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 (Vila do Conde, Portugal) for providing the data set. J. Vasconcelos received financial support from Fundação para a Ciência e a Tecnologia (SFRH/BD/11397/2002) of the Ministry of Science and Technology, Portugal.

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. Proc. 8th World Congr. Gen. Appl. Livest. Prod., Belo Horizonte, Brazil. CD-ROM commun. no. 25-09.
- 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. Szwaczkowski. 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. *Proc. 8th World Cong. Gen. Appl. Livest. Prod.*, Belo Horizonte, Brazil. CD-ROM commun. no. 24-30.
- 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.