

## IDENTIFICATION OF PRODUCTION ENVIRONMENTS AND PRESENCE OF GxE INTERACTIONS IN URUGUAY USING HOLSTEIN HERDS RECORDS

H. Naya<sup>1</sup>, J.I. Urioste<sup>1</sup> and J.Franco<sup>2</sup>

<sup>1</sup> Depto. Producción Animal y Pasturas, Fac. de Agronomía, 12900 Montevideo, Uruguay

<sup>2</sup> Depto. Biometría y Estadística, Fac. de Agronomía, 12900 Montevideo, Uruguay

### INTRODUCTION

Many applications of mixed models in genetic evaluation of dairy cattle assume constant residual and genetic variances across production environments. However, extensive research (Hill *et al.*, 1983 ; Dodenhoff and Swalve, 1998 ; Ibáñez *et al.*, 1999 ; Urioste *et al.*, 2001) has found evidence of intra-herd heterogeneity of variances for milk traits. Some authors (*e.g.* Ibáñez *et al.*, 1996) have grouped dairy herds in clusters characterized by similar production levels. Zwald *et al.* (2001) have posed climate, management and genetic level measures to describe dairy production systems in different countries. Recent research (Fikse *et al.*, 2001 ; Kearney *et al.*, 2001 ; Zwald and Weigel, 2001) suggests the existence of genotype by environment interactions in milk yield, both between and within countries. The objectives of this study are: (i) to identify and characterize different production systems for the Uruguayan Holsteins participating in Milk Recording Schemes, in terms of production level and management factors; (ii) to assess the presence of genetic by environment interactions by estimation of genetic (co)variances, heritabilities and genetic correlations between milk yield in the different production systems.

### MATERIALS AND METHODS

Information was obtained from 248646 lactations, corresponding to 119255 cows from 497 dairy herds participating in two National Recording Schemes and fulfilling the editing requirements in the 2001 Uruguayan national genetic evaluation (see Urioste *et al.*, 2001 for details). Production and pedigree data were supplied by the Asociación Rural del Uruguay (ARU) and by the Instituto Nacional para el Mejoramiento Lechero (INML). A new file was constructed, containing informative variables by herd.

Average peak yield (**PYLD**), calculated as the average of maximum test day yield for each cow, average herd milk yield at 305 days (**YLD**) and average herd standard deviation of milk yield (**SDYLD**) provided information about production intensity on each farm. General measures of the management system were average number of lactations per year, taken between 1996 and 2000 (**ANLAC**), as a measure of the herd size, and percentage of lactations completed to 305 days (**PCL**) and average calving age (**ACAGE**) as expressions of herd culling policies. The following variables were thought of as to capture differences between herds in nutritional management before and after calving: average dry period (**DRYP**), days to peak yield (**DTPK**) and persistency of lactation (**PERS**), the latter measure calculated as the quotient between the seventh and the second test day yield. Variation in reproductive management policies was assessed through two variables: average calving day (**ACD**), where consecutive numbers were assigned to each day (1 to January 1<sup>st</sup>, 365 to December 31<sup>st</sup>), and seasonal calving (**SEAC**), calculated as the difference between the number of calvings in

spring (August to October) and fall (March to May), divided by the total sum of spring and fall calvings. This measure varies between  $-1$  (all calvings in the fall) and  $1$  (the opposite).

Climate (average maximum daily temperature and average relative humidity in the warmest month of the year) and genetic level variables (herd average Predicted Transmitting Ability and percentage of imported Holstein genes) were previously investigated, but not used in the present study. The variable YLD was also discarded, due to the extreme high correlation (0.97) with PKYLD.

Identification of different production environments was performed with cluster analysis techniques, using the variables defined above. All analyses were conducted with SAS (2000). The original variables were standardized (proc STANDARD) and a hierarchical cluster analysis was performed (proc CLUSTER with the method of minimum variances within group of Ward). The appropriate number of clusters was decided on the basis of the pseudo F statistic. Next, a canonical discriminant analysis (Mardia *et al.*, 1979) was conducted (proc CANDISC), to derive a set of new canonical variables, linear combination of the original ones, which maximize in a reduced number of dimensions (two or three) the relation of the variation between clusters with respect to the variation within clusters. Additionally, we used the step discriminant analysis (proc STEPDISC) for the grouping interpretation.

The genetic characterization of the groupings was made by estimating genetic, permanent environmental and residual variances (and corresponding heritabilities) for milk yield at 305 days in each 'environment'. To detect genotype by environment interactions (GxE), it was assumed that milk yield in the different environments were different traits (Lynch and Walsh, 1998), and that genetic correlations deviating substantially from one were a strong indication of this. The residual covariance structure assumed no covariances between environments. Information from 240066 lactations and 133630 animals in the pedigree file was used for this purpose. The statistical model for each yield variable included the fixed effects of herd-year-month of calving and age-parity classes, and animal and permanent environment as random effects. Pairwise analysis of traits was done with the ASREML software (Gilmour *et al.*, 1997). Cluster four had a very few observations (1784 lactations) and was not analyzed.

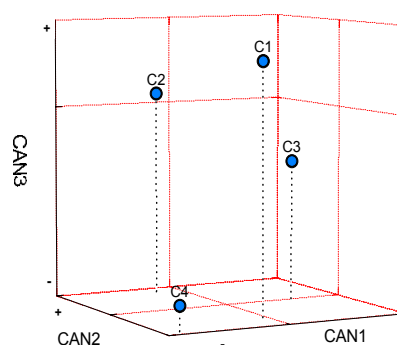
## RESULTS AND DISCUSSION

Four distinct groups were identified with the cluster analysis, being cluster 4 clearly different from the others. Correlation coefficients between the original and the canonical variables are presented in Table 1. The original variables were ordered after their entry in the step discriminant analysis. The canonical variable 1 (**Can1**) was positively associated with DTPK and PERS (variables describing the lactation curve), and negatively with ACD and SEAC, which describe seasonal patterns of calving. **Can2** was strongly correlated with DRYP (nutritional management before calving), ACAGE (culling policy) and PYLD. The canonical variable **Can3** was associated with production variables (PYLD and SDYLD) and also with SEAC and in some extent with ACAGE. Figure 1 shows the position of the cluster centroids in the three dimensions. Cluster 4 was very different of the rest in all canonical variables. Can1 separated Cluster 3 from Clusters 1 and 2, Can2 set apart Cluster 1 from Cluster 2, and Can3 distinguished between Clusters 1 and 3.

**Table 1. Correlation coefficients between the original and the canonical variables<sup>A</sup>**

	can1	can2	can3
<b>DTPK</b> (days)	<b>0.681</b>	0.270	-0.180
<b>DRYP</b> (days)	-0.251	<b>0.523</b>	0.018
<b>SEAC</b>	<b>-0.542</b>	-0.335	-0.446
<b>PYLD</b> (kg)	0.110	-0.452	<b>0.499</b>
<b>ACAGE</b> (months)	-0.157	<b>0.460</b>	0.375
<b>SDYLD</b> (kg)	0.140	-0.286	<b>0.660</b>
<b>PERS</b>	0.598	0.213	-0.206
<b>PCL</b> (%)	-0.125	-0.048	0.151
<b>ANLAC</b>	0.192	-0.100	0.219
<b>ACD</b>	-0.552	-0.165	-0.271

<sup>A</sup>Variables ordered after importance in the discriminant analysis

**Figure 1. Cluster centroid positions in the three canonical axes**

Differences between clusters, expressed in the original variables, are illustrated in Table 2. Cluster 1 was characterized by high milk production, short dry period and days to peak, and balanced calving season, suggesting an all-year-round calving system, with improved nutritional management before and after calving. Cluster 2 was the lowest in milk yield, with the longest dry period and oldest cows. Cluster 3 included herds intermediate in production, its main features being a distinct fall calving and higher persistency of the lactation curve. A few, small herds with low production, very short days to peak yield and spring concentrated calvings were grouped in Cluster 4.

**Table 2. Mean of management and production variables for the four identified clusters**

	CLUSTER 1 N=174	CLUSTER 2 N=171	CLUSTER 3 N=121	CLUSTER 4 N=31
PYLD (kg.)	24.1	20.8	21.9	20.6
YLD (kg.)	5270.6	4456.5	4854.9	4409.4
SDYLD (kg.)	1025.5	860.8	890.6	716.7
ANLAC	93.7	63.1	92.7	22.4
PCL (%)	41.3	39.7	34.8	40.5
ACAGE (months)	50.1	54.9	48.5	46.1
DRYP (days)	77.9	89.1	78.2	81.3
DTPK (days)	74.8	77.8	91	62.6
PERS	0.783	0.795	0.859	0.733
ACD	173.6	175	152.4	226.1
SEAC	-0.195	-0.264	-0.425	0.561

The genetic characterization of Clusters 1 to 3 is shown in Table 3. The pooled heritability estimates (0.25 to 0.31) were in the usual range of milk yield, but genetic variances increased

with production level and genetic correlations varied between 0.74 and 0.93, indicating the presence of genotype by environment interactions in milk yield for the Uruguayan Holstein population. These results suggest that sires may rank differently when progeny tested in different environments and thus obtain biased estimates of genetic merit.

**Table 3. Pooled estimates of genetic variances (diagonal) and covariances (lower triangle), heritabilities (diagonal, bold type within parenthesis) and genetic correlations (upper triangle) between environments (clusters) for milk yield**

	N° sires	N° records	CLUSTER 1	CLUSTER 2	CLUSTER 3
CLUSTER 1	1466	106776	176782 <b>(0.31)</b>	0.83	0.93
CLUSTER 2	1236	75098	122014	122309 <b>(0.27)</b>	0.74
CLUSTER 3	799	58192	151452	100974	150813 <b>(0.25)</b>

### CONCLUSIONS

Multivariate clustering techniques appeared to be useful tools for detecting distinct production environments from milk recording databases. Further, GxE interactions can be detected. Deeper studies adding new variables, such as climate (as suggested by Ravagnolo *et al.*, 2000) or other variables, are in order.

### ACKNOWLEDGEMENTS

ARU and INML are thanked for supplying the data for this study.

### REFERENCES

- Dodenhoff, J. and Swalve, H.H. (1998) *Livest. Prod. Sci.* **53** : 225-236.
- Fikse, W.F., Rekaya, R. and Weigel, K.A. (2001) *J. Dairy Sci.* **84 (Suppl. 1)** : 214.
- Gilmour, A.R., Thompson, R., Cullis, B.R. and Welham, S.J. (2000) "ASREML Manual". New South Wales Department of Agriculture, Orange, 2800 Australia.
- Hill, W.G., Edwards, M.R. Ahmed, M.-K.A. and Thompson, R. (1983) *Anim. Prod.* **36** : 59-68.
- Ibáñez, M.A., Carabaño, M.J., Foulley, J.L. and Alenda, R. (1996) *Livest. Prod. Sci.* **45** : 137-147.
- Ibáñez, M.A., Carabaño, M.J. and Alenda, R. (1999) *Livest. Prod. Sci.* **59** : 33-49.
- Kearney, J.F., Schutz, M.M., Boettcher, P.J. and Weigel, K.A. (2001) *J. Dairy Sci.* **84 (Suppl. 1)** : 214.
- Lynch, M. and Walsh, B. (1998) "Genetics and Analysis of Quantitative Traits". Sinauer Associates, Inc., Massachusetts, USA.
- Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) "Multivariate analysis", London, Academic Press, Inc.
- Ravagnolo, O., Misztal, I. and Hoogenboom, G. (2000) *J. Dairy Sci.* **83** : 2120-2125.
- SAS Institute Inc. (2000) SAS Online Doc®, version **8**.
- Urioste, J.I., Gianola, D., Rekaya, R., Fikse, W.F. and Weigel, K.A. (2001) *Anim. Sci.* **72** : 259-268.
- Zwald, N.R. and Weigel, K.A. (2001) *J. Dairy Sci.* **84 (Suppl. 1)** : 214.
- Zwald, N.R., Weigel, K.A., Fikse, W.F. and Rekaya, R. (2001) *J. Dairy Sci.* **84** : 2530-2534.