

Application of a Multiple-Trait Herd Cluster Model for Genetic Evaluation of Dairy Sires from Seventeen Countries

N.R. Zwald, K.A. Weigel and R. Rekaya
University of Wisconsin, Madison 53706

Introduction

The number of dairy sires that are used for genetic improvement internationally is growing each year, and accuracy of genetic information for foreign animals is increasingly important. The last major modification of international dairy sire evaluations was in 1994 when the International Bull Evaluation Service (Interbull) implemented the multiple-trait across country evaluation (MACE) procedure. This allowed data from progeny in multiple countries to be combined simultaneously, and international EBV could be calculated from relatives' information for bulls with progeny in only one country. However this method treats progeny of a given sire differently based upon political boundaries regardless of the production environment in which these animals are actually performing. The MACE procedure considers lactation performance in small neighboring countries as different traits, and performance in large countries with varying management and climatic factors is considered a single trait. Therefore, it cannot properly account for genotype by environment interaction between herds.

This is the final paper from a large study. Previous papers have examined the production systems in seventeen leading dairy countries and identified management, climatic, and genetic factors that can explain genotype by environment interactions between herds. The present paper describes how such factors can be used to group herds into clusters or production environments for the purpose of international sire evaluation. In this approach, dairy sires no longer get an EBV for each individual country; they instead get an EBV for each unique production system. Data from progeny and other relatives in different clusters can be used to predict the performance of each sire in each production environment that exists globally. This method should increase the reliability and credibility of genetic information for dairy sires, and should ultimately increase genetic progress in each production environment.

The objectives of this study were twofold: 1) to determine a weighting scheme that would account for the correlations between descriptive herd variables, as well as the relative importance of each variable in defining a unique production environment, and 2) to apply this model to milk production data from seventeen leading dairy countries.

Materials and Methods

Data

The data consisted of 131,907,373 test day records from 16,403,413 first lactation cows in 233, 673 herds in Australia, Austria, Belgium, Canada, Czech Republic, Estonia, Finland, Germany, Hungary, Ireland, Israel, Italy, The Netherlands, New Zealand, South Africa, Switzerland, and the United States (Zwald et al., 2001). Herd means were calculated for each of the following variables: day of calving, age at calving, days to peak yield, fat: protein ratio, herd size, percent North American Holstein genes, peak yield, persistency, sire PTA milk, annual rainfall, standard deviation of milk yield, and maximum monthly temperature. Means for these variables were subsequently used to group herds into clusters or production systems for genetic evaluation purposes.

Weighting Variables

Because some variables are more important for explaining genotype by environment interaction than others, it was necessary to weight each of the thirteen management, climate and genetic factors prior to the clusters analysis. Weights were developed from the genetic covariance matrices between quintile groups as published by Zwald et al. (2001). Covariance functions were applied to these matrices and the difference in the sum of squared errors between the full-fit legendre polynomial regression and its corresponding intercept term was used as a weight. To avoid

double counting of correlated variables, a weighted factor analysis was used to develop unique contrasts of these variables. Seven factors had eigenvalues larger than 0.025, and these factors explained nearly 94% of the total variation. Scores for each of the seven factors were calculated and these scores were used in the subsequent cluster analysis. Each of the seven factor scores calculated for each herd was a composite of all thirteen herd variables based on the eigenvectors (weights) that were computed.

Model

A seven-trait BLUP sire model was used for this analysis with production in each cluster considered a separate trait. A heterogeneous variance adjustment was applied using the method of Wiggans and VanRaden (1991). The model is as follows:

$$y_{ijklmnop} = hys_i + Herd_j + Season_k + Breed_l + Age_m + MF_n + sire_o + e_{ijklmnop}$$

where

$y_{ijklmno}$	=	Lactation milk yield
hys_i	=	Random herd-year-season effect
$Herd_j$	=	Fixed herd effect
$Season_k$	=	Fixed season effect
$Breed_l$	=	Fixed breed composition effect
Age_m	=	Fixed age at calving effect
MF_n	=	Fixed milking frequency effect
$Sire_o$	=	Random sire effect
e_p	=	Random residual

Bayesian implementation via Gibbs sampling was used to estimate the covariance matrix corresponding to each cluster so that heritability within each cluster and genetic correlations between clusters could be examined. Sire EBV were calculated in each cluster and are presented.

Cluster Analysis

The FASTCLUS procedure in SAS was used for herd clustering, which performs a disjoint cluster analysis based on Euclidean distances. This iterative method guarantees that distances between all observations in the same cluster will be less than the distances between observations in different clusters. For each herd, the thirteen descriptive variables were combined into seven

factors using the weights described above, and the cluster analysis was based on the seven factor scores for each herd. The optimal number of clusters was based on the cubic clustering criterion, which compares the observed R-squared to the expected R-squared from a uniform distribution. (SAS Technical Report, 1983). High values of the cubic clustering criterion indicate more clearly defined clusters.

Results and Discussion

The relative importance of each variable in determining unique production environments is quantified in Table 1. Variables that were better predictors of genotype by environment interaction received the largest weights, and variables that were less informative received the smallest weights. Peak yield received the highest weight, because herds with different levels of peak yield had relatively low genetic correlations with each other. In contrast, fat: protein ratio received very little weight because herds with very different fat: protein ratio had high genetic correlations with each other, indicating that this variable did little to determine unique production environments. Eigenvalues of the weighted correlation matrix for the thirteen descriptive herd variables were computed. These eigenvalues consider both the correlation between variables and the relative weights from Table 1. Although these weights are difficult to decipher, inspection of these values can give an indication of the important variables in each factor. For example, factor 1 has a high positive weight on peak yield, and standard deviation of milk yield and a large negative weight on fat: protein ratio. Therefore, herds with a low fat: protein ratio, high peak yield, and high standard deviation of milk yield will have a high score for factor 1. Herds that are very large with a low percentage of North American genes will get a high score for factor 2, because of the large positive weight on herd size and the relatively large negative weight on North American Holstein percentage. Application of the cluster analysis to these factor scores led to seven herd clusters; each cluster contained from 4805 to 59,272 herds and 977,815 to 3,627,768 cows.

Means of the 13 descriptive herd variables for each cluster are shown in Table 2, and the representation of herds from each country or region in each cluster is shown in Table 4. Cluster 1 was comprised primarily of large herds

in Eastern Germany, the Czech Republic, Italy, and the United States. These herds were characterized by extended days to peak yield, a very large herd size and relatively low annual rainfall. Cluster 2 was the largest cluster, and it consisted of smaller herds that used a large proportion of North American Holstein genetics and had a moderate peak yield. This cluster included the majority of cows in the Netherlands as well as western Germany. Canada and Australia also had a large number of herds in this production environment. Cluster 3 was made up of small herds with low peak yield, extended days to peak yield, and high persistency. Australia, Germany, and the Czech Republic all had strong representation in this cluster. Cluster 4 largely consisted of herds from western Germany and Italy. These herds typically had the oldest age at first calving and a high standard deviation of milk yield. New Zealand herds make up over 97% of the cows in cluster 5. This cluster is comprised of low peak yields, low persistency, low percentage of North American genes, and a high fat: protein ratio. Cluster 6 contained mainly small herds from northern Europe, as well as a few herds from New Zealand. Cluster 7 consisted primarily of moderate sized herds in the United States and Canada with a very high peak yield, extended days to peak yield, and a high genetic level.

Heritability within clusters and genetic correlations between clusters are shown in Table 3. Heritability ranged from 0.24 to 0.42 and genetic correlations ranged from 0.59, between low input grazing herds of New Zealand (cluster 5) and the intensely managed herds with high peak yields located primarily in the United States and Canada (cluster 7), to 0.97, between Clusters 2 and 4. The relatively low genetic correlation between the grazing herds of New Zealand (Cluster 5) and the other clusters would suggest that the environment that cows are subject to in these herds is significantly different from other environments around the world. Other genetic correlations are similar to those between countries that are currently used by Interbull in the MACE procedure.

Conclusions

An alternative method of evaluating dairy sires based on production environments was proposed and applied to data from seventeen leading dairy countries. Intuitively, this model makes sense

because animals in similar environments are grouped accordingly, regardless of country borders. Thus genotype by environment interaction should be properly taken into account. Each of the thirteen descriptive herd variables was assigned a weight based on genetic correlations between lactation performances in herds at different levels. Herd size, peak yield, and temperature were found to be the best predictor of genotype by environment interaction between herds. Many of the descriptive variables were highly correlated, and to avoid double counting certain variables, a factor analysis was used to find seven uncorrelated factors. Seven clusters were formed using the management, genetic, and climatic information for each herd, and each of these clusters was considered as a separate trait for genetic evaluation purposes. Improvements could be made to this analysis by having specific temperature and rainfall estimates for each herd, as opposed to regional estimates. Traits other than milk production could also be evaluated to further distinguish between environments in the global dairy population.

Currently there are 27 separate countries (environments) that are members of Interbull. Therefore, each dairy sire receives 27 different EBV. It does not seem logical to assume that each country is a single environment that is independent from similar environments in other countries. By using the proposed model, the number of unique production environments and therefore the number of EBV for each bull would be reduced from 27 to seven. Each sire would receive a separate evaluation in each of the seven unique production environments based on progeny and relatives in that cluster. Evaluations would also consider progeny performance in other clusters, but that information would be discounted based on the genetic correlations between the clusters.

The approach to international sire evaluation presented herein would be an improvement over the current system of evaluating sires because it would properly account for management, climatic, or genetic differences between herds in large countries (e.g. tie-stall herds in Vermont versus large dairies in Arizona) by placing those herds in separate clusters, and it would also allow similarly managed herds in different countries to be in the same cluster and thus treated as the same environment for genetic evaluation purposes (e.g. grazing herds in Ireland and grazing herds in

Australia). This system should lead to higher reliability and credibility of international sire EBV. Implementation of this model would mean that an individual sire would no longer have a national EBV, and different sire EBV would be appropriate for different herds within a country.

References

- 1 Banos, G. 1995. Application of international sire evaluations. *J. Dairy Sci.* 78 (Suppl.1), 221 (Abstr.).
- 2 Cromie, A.R., Kelleher D.L., Gordon, F.J. & Rath, M. 1998. Proc. 1998 Interbull meeting., Rotorua, New Zealand. *Interbull Bulletin* 17, 100-104.
- 3 Kirkpatrick, M., Lofsvold, D. & Bulmer, M. 1990. *Genetics* 124, 979.
- 4 Lohuis, M.M. & Deckers, J.C.M. 1998. Proc. 6th World Congr. Genet. Appl. Livest. Prod.
- 5 Schaeffer, L.R. 1985. *J. Dairy Sci.* 77, 2671.
- 6 Wiggans, G.R. & VanRaden, P.M. 1991. *J. Dairy Sci.* 74, 4350.

Table 1. Weights for each of the thirteen descriptive herd variables, based on the genetic correlations for milk yield between quintile groups for each variable.

<u>Variable</u>	<u>Weight</u>
Peak yield	0.2442
Herd size	0.2386
Temperature	0.1567
Standard deviation of milk yield	0.0863
Percent North American Holstein	0.0742
Days to peak yield	0.0668
Rainfall	0.0479
Age at calving	0.0195
Percent of animals completing first lactation	0.0169
Persistency of yield	0.0136
PTAM of sires of cows currently in the herd	0.0130
Calving day	0.0115
Fat: protein ratio	0.0109

Table 2. Means of descriptive herd variables corresponding to each cluster.

<u>Variable</u>	<u>Cluster</u>						
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
Age at calving	27.5	28.5	28.5	28.8	24.4	27.4	27.2
Calving day	10	18	12	16	39	11	8
Days to peak yield	79.2	58.9	90.2	68.5	59.3	67.9	92.8
Fat: protein ratio	1.22	1.28	1.26	1.23	1.31	1.27	1.15
Herd size	86	9.5	11.4	11	25.3	9.5	19.8
Percent North American Holstein	60.2	81.9	76.3	76.8	28.2	29.5	88.7
Peak yield	25.1	26.1	22.6	25.9	17.3	23.7	32.3
Percent of animals completing first lactation	0.806	0.793	0.794	0.771	0.913	0.801	0.796
Persistency of yield	0.755	0.753	0.805	0.776	0.604	0.745	0.842
PTAM of sires of cows currently in the herd	173	196	166	229	-238	-231	700
Rainfall	26.4	31.5	30.4	30.9	53.2	25.6	36.7
Standard deviation of milk yield	1030	870	819	1259	579	825	1179
Temperature	24.8	23	24.4	23.9	21.6	22.8	28.2

Table 3. Heritability and Genetic correlation between the seven environments (clusters)

<u>Cluster</u>	<u>Cluster</u>						
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
1	0.33	0.92	0.91	0.94	0.68	0.95	0.91
2		0.39	0.96	0.97	0.64	0.91	0.95
3			0.35	0.95	0.68	0.91	0.92
4				0.39	0.61	0.92	0.94
5					0.24	0.75	0.59
6						0.42	0.85
7							0.28

Table 4. Number of cows from each country and region corresponding to each cluster.

<u>Country / Region</u>	<u>Cluster</u>						
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
AUSTRALIA TOTAL	98515	290787	327006	39921	28995	25222	29331
New South Wales	724	3017	34179	6034	27331	0	17881
Victoria	14934	255711	181318	25410	1664	12173	3843
Queensland	0	3627	56845	3689	0	404	1295
South Australia	75292	11581	32344	3905	0	5740	962
Western Australia	0	9477	12878	774	0	3877	5350
Tasmania	7565	7374	9442	109	0	3028	0
AUSTRIA	0	4239	1027	1421	124	251	53
BELGIUM	414	31250	13582	20680	0	1237	883
CANADA TOTAL	26766	469495	49442	123721	712	28159	330653
PEI / Newfoundland	0	5651	614	2481	0	42	2777
Nova Scotia	560	3310	75	1929	312	0	12820
New Brunswick	466	7943	825	2087	0	0	8980
Quebec	1644	219830	4101	33912	400	2345	100565
Ontario	5626	191655	27377	60748	0	9655	127974
Manitoba	952	9780	4486	4504	0	1842	5661
Saskatchewan	574	6736	3733	3059	0	2874	2793
Alberta	10184	15883	6788	7548	0	10959	14293
British Colombia	6760	8707	1443	7453	0	442	54790
SWITZERLAND	0	37645	2383	3819	9	783	336
CZECH REPUBLIC	75273	26780	30982	3136	0	16775	231
GERMANY TOTAL	608823	1245549	165774	556274	0	384675	5596
North West Germany	12182	996182	97691	317517	0	121473	4860
North East Germany	246597	7223	14396	51902	0	87213	0
South East Germany	350044	16364	18015	88304	0	97129	0
South West Germany	0	225780	35672	98551	0	78860	736
ESTONIA	3979	2011	5787	338	0	1092	0
FINLAND	0	156	588	2991	0	108022	32
HUNGARY	168656	1846	77507	12806	0	2705	2300
IRELAND	3463	31534	26756	3793	0	58446	739
ISRAEL	144063	0	78	0	0	63573	457
ITALY TOTAL	294571	168883	88372	382062	1228	49013	431303
Northern Italy	187940	61587	47948	230246	1228	15284	398890
Central-Northern Italy	62555	86335	15314	98788	0	19029	12738
Central-Southern Italy	11589	12472	9924	23338	0	3117	7510
Southern Italy	32487	8489	15186	29690	0	11583	12165
THE NETHERLANDS	14964	1156917	104655	160834	0	8705	30897
NEW ZEALAND TOTAL	123474	21959	40638	4349	1110630	180609	104
Northern North Island NZ	3975	680	3032	280	143484	0	0
Central North Island NZ	15740	0	590	138	698245	0	9
Southern North Island NZ	8917	495	507	65	114177	0	0
Northern South Island NZ	18512	6463	16281	1758	0	151736	0
Central South Island NZ	19417	8012	14587	860	48378	28392	0
Southern South Island NZ	56913	6309	5641	1248	106346	481	95
UNITED STATES TOTAL	467852	133473	137300	94317	10944	45464	2244577
Northeastern States	75361	16623	27062	31302	1456	3990	742076
Central States	21993	3160	14405	7469	191	3162	233709
North Central States	62863	110390	53919	44805	0	29310	901326
Northwestern States	45352	2485	16770	3780	0	2399	112700
South Central States	11678	196	7610	856	795	677	57474
Southeastern States	21401	90	5449	4411	8502	385	130131
Southwestern States	229204	529	12085	1694	0	5541	67161
<u>SOUTH AFRICA</u>	<u>36006</u>	<u>5244</u>	<u>39222</u>	<u>20586</u>	<u>305</u>	<u>3084</u>	<u>13168</u>
TOTAL	2066819	3627768	1111099	1431048	1152947	977815	3090660