Preliminary Test Day genetic evaluations in the United Kingdom (UK) involving the use of Legendre Polynomials and Wilmink Function

R.A. Mrode and G. J. T. Swanson,
Animal Data Centre Ltd, Fox Talbot House, Greenways Business Park, Chippenham, Wiltshire

Introduction

Various methods for genetic evaluation using Test Day (TD) yields have been implemented or proposed. These include the multiple trait approach (Wiggans and Goddard, 1997) and the use of random regression models (RRM) involving the fitting of parametric curves or covariance functions or cubic splines. Initial research in the UK had focused on RRM examining the modelling of the lactation curve using parametric curves such as Wilmink (1987), Legendre polynomials and cubic splines (White et al, 1999). While the parametric curves generally fitted the lactation curve best, they produced negative genetic correlations (r_g) between yield in early and later lactations (Brotherstone et al, 1999). They observed that the models fitting orthogonal polynomials or cubic splines gave positive r_g across the entire lactation length but the former gave a better fit than the latter.

While further research is anticipated, the UK is currently working on implementing a RRM fitting Legendre polynomials to model the lactation curve for the multi-trait analysis of milk, fat and protein test day yields. Later lactation test day yields are handled as repeated observations with their variances scaled to those of first lactation. This paper presents some preliminary results from the analysis of the Jersey breed. It also compares results from fitting Legendre polynomials (LP) and the Wilmink curve (WL) to assess the impact of the negative r_g between yield in early and later lactations on evaluations.

Materials and Method

Data consisted of TD records for milk, fat and protein yields for up to five lactations of 24,871 Jersey cows calving since 1991. (See Table 1).

The RRM fitting Legendre polynomials to model the lactation curve was implemented in a multi-trait analysis of milk, fat and protein yields. TD records in later lactations were handled as repeated observations with their variances scaled to those of the first by the ratio of standard deviations of the appropriate daily yields. The model equation is assumed to be the same for all traits. For trait i, it is

$$Y_{imtljkn} = htd_{il} + \sum_{r=1}^{5} \beta_{imk} v_{tr} + Ø_{jtn} a_{ij} + Ø_{jtn} p_{ij} + e_{imtljkn}$$

Where $Y_{imtljkn}$ is the test day record n on trait i of cow j made on day t within herd-test-day-parity subclass l, for a cow from subclass k for calving season and age at calving subclass nested within lactation m; htd_{il} is the l^{th} herd-test date for trait i; β_{imk} are the fixed regression coefficients for lactation m of trait i specific to subclass k, v is a vector of the first five Legendre polynomials for the t^{th} day in milk; a_{ij} and p_{ij} are vectors of three random regressions for animal and permanent environmental effects respectively for animal j for trait i; $Ø_{jtn}$ is a vector of the first three Legendre polynomials for the test day n of animal j made on day t; and $e_{imtljkn}$ is random residual. Four difference residual variances were used in the analysis based on number of days in milk.

Gauss Seidel iteration with over-relaxation was employed to solve the mixed model equations, iterating on the data. Initially only first lactation milk, fat and protein were analysed. Later lactation TD were then included after scaling. Breeding values for 305-day yields for milk, fat and protein were computed for every animal, from the random regression coefficients after convergence. The breeding values were then expressed as predicted transmitting abilities (PTAs) and compared with the current official PTAs for the Jersey breed.

To compare results from LP and WL, two univariate analyses of TD milk yield were carried out using the above model. The first fitting Legendre polynomials and the second fitting the Wilmink function for both the fixed and random regressions. In Table 2 are the genetic correlations between TD yields over the lactation generated from parameters for the coefficients of LP and WL. The accuracy of both models was compared by calculating the mean and variance of prediction errors (y-ŷ) after equations have converged. In addition, in the last round of iteration, the vector of the random regression coefficients for cow j ($\hat{a}_j$) was partitioned into the contributions from parents ($\hat{a}_{qj}$) and observations ($\hat{a}_{oj}$) for cows with yield records but no progeny and written to a file. Thus for cow j, $\hat{a}_j = \hat{a}_{qj} + \hat{a}_{oj}$. The breeding value for 305-day yield and the proportionate contribution from parents and records to the breeding value were computed from $\hat{a}_j$, $\hat{a}_{qj}$ and $\hat{a}_{oj}$ respectively for cows. The correlations between these components from the model fitting LP and WL were calculated, and distribution of differences (LP-WL) were examined for cows with varying numbers of TD and different sections of the lactation curve.

Results and Discussion

The correlations between PTAs for 305-day yield computed from the multivariate RRM analysis with only first lactation milk, fat and protein TD yields and the official PTAs from a univariate animal model analysis are in Table 3. Also shown in Table 3 are the correlations between PTAs from the RRM using only first actation with those based on up to five lactations. The high correlation of 0.98 across all traits between PTAs from RRM based on only first lactation and those from up to 5 lactations indicate inclusion of later lactations did not result in a drastic re-ranking of bulls. This seems to indicate that the way later lactations were handled in this analysis by scaling their TD variance to those of the first and treating them as repeated measurements of the first lactation, is reasonable.

Considering bulls with at least 10 daughters, the correlations between PTAs from the RRM and the official evaluation ranged from 0.90 for milk to 0.92 for fat. This correlation increased to 0.93 – 0.94 when bulls with at least 10 daughters in the RRM and ±5 daughters in the official run were considered. This was essentially based on bulls with similar number of daughters. Similar correlations were reported between breeding values from RRM and the official runs in Canada (Jamrozik et al, 1997) and Germany (Reents et al, 1998).

Comparison of univariate analyses using LP and WL.

The mean prediction errors for both models were similar apart from the beginning of the lactation when the means were larger for LP. A similar trend was observed for the variance of prediction errors. However, in general LP and WL were very similar both in terms of accuracy of predicting daily yields.

Over all categories of bulls, in terms of number of daughters, the PTAs for 305-day yield from the RRM analysis of milk fitting either a LP or WL were highly correlated ($r = 0.99$). Similar high correlations (0.98-0.99) were obtained if PTAs for different sections of the lactation curve were considered. This implies both models gave similar bull rankings and results are in line with those of Jamrozik et al (1997) who compared several parametric curves.

PTAs for 305-day yield and components of PTAs, that is, the proportionate contribution from parents (PA) and records (RC) obtained from LP were correlated with those from WL

(Table 4) for cows at different stages of lactation. The means of differences between PTAs, PA and RC from LP and WL for 305-day yield are in Table 5. While the PTAs and PA for cows from LP were highly correlated with those from WL, the correlation was only 0.77 for RC for cows which were in milk in days 4 to 30. This indicates a difference between both models in terms of contribution of information from early TD to 305-day yield breeding values.

The mean differences in Table 5 similarly indicate large differences between both models in terms of the proportionate contribution of information from records for cows in milk up to 90 days. This difference between LP and WL may be due to the negative r_g between the early and later TD yields observed for the Wilmink function. This results in under-prediction of the proportionate contribution of information from early TD records to 305-day yield PTA for cows with few TD. Notably, the negative mean difference in PA would imply contributions from parents seem to "compensate" for this under-prediction from records in early lactation, hence the overall correlation between PTAs was 0.99. Contributions from PA would be estimated from daughters in different stages of lactation and so should not be influenced by the negative r_g observed with WL.

To demonstrate the above view, PTAs, PA and RC from LP were correlated with those from WL for different sections of the lactation curve (Table 6). In addition the mean differences between both models were computed for the corresponding sections of the lactation curve (Table 7). The mean differences for PTA, PA and RC from both models for 100-day yield were close to zero. Similarly, the correlation between RC from LP and WL was 0.89 for cows in milk 4-30 days. These indicate both models are similar in predicting PTAs for 100-day yields even for cows with TD only in early lactation. The r_g among TD in early lactation with the WL are positive, just as is the case with LP, or close to zero. However, considering PTAs and components of PTAs for 101-200 day and 201-305 day yields the correlation between RC from both models varies from 0.61 to 0.76 for cows in milk 4-30 days only and the mean of differences (LP-WL) ranged from 5.2 to 9.1. This is due to the negative correlations with the WL between yields in early lactation and yields in these sections of the lactation curve. Thus contribution from early records to PTAs for the later section of the lactation curve is lower compared with LP. For cows in milk for 90 days or more, the low correlation between RC from both models or large mean differences in RC is not observed. This is because the contribution of information from early TD is overwhelmed by information from the TD in the middle and end of the lactation, because of higher heritabilities of these TD and high genetic correlations among them.

Conclusion

The results indicate the RRM being implemented for analysis of TD records in the UK looks promising and the inclusion of later lactation TD as repeated observation after scaling seems reasonable.

The use of parametric curves such as WL compared with LP reduces the contribution of information from early TD records and increases parental contribution to evaluations for 200 to 305-day yield for cows with few test day records.

Acknowledgement

The authors thank S Brotherstone and I White for useful comments and supply of genetic parameters used in the analysis.

References

Brotherstone, S., White, I.M.S. & Meyer, K. 1999. Genetic modelling of daily milk yield using orthogonal polynomials and parametric curves. In press.

Jamrozik, J.O., Schaeffer, L.R. & Dekkers, J.C.M. 1997. Genetic evaluation of dairy cattle using test day yields and random regression model. J. Dairy Sci. 80, 1217-1226.

Jamrozik, J., Kistemaker, G.J., Dekkers, J.C.M. & Schaeffer, L.R. 1997. Comparison of possible covariates for use in a random regression model for analysis of test day yield. J. Dairy Sci. 80, 2550 – 2556.

Reents, R., Dopp, L., Schmutz, M. & Reinhardt, F. 1998. Impact of application of a test day model to dairy production traits on genetic evaluations of cows. Interbull Bulletin No 17, 49-54.

White, I.M.S., Thompson, R. & Brotherstone, S. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. 82, 632-638.

Wiggan, G. & 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.

Wilmink, J.B.M. 1987. Adjustment of test-day milk, fat and protein yields for age, season and stage of lactation. Livest. Prod . Sci. 16, 335-348.

Table 1. Distribution of cows and test day records by lactation in the Jersey Breed

	Lactation 1	**Lactation 2**	**Lactation 3**	**Lactation 4**	**Lactation 5**
Number of Cows	24,871	12,751	7,951	4,229	1,919
No of Test Days	197,192	102,089	60,363	30,081	12,113

Table 2. Genetic correlations between daily milk yield at days 7,14, 28, 42, 63, 98, 140, 182, 238, 280 and 301 estimated from Legendre polynomials (above diagonals) and Wilmink function (below diagonal)

7	**14**	**28**	**42**	**63**	**98**	**140**	**182**	**238**	**280**	**301**
-	0.99	0.96	0.91	0.82	0.69	0.59	0.52	0.45	0.41	0.38
0.52	-	0.98	0.94	0.87	0.76	0.66	0.60	0.53	0.48	0.45
0.05	0.88	-	0.99	0.95	0.87	0.79	0.73	0.67	0.61	0.56
-0.04	0.84	1.00	-	0.98	0.93	0.87	0.82	0.76	0.70	0.64
-0.08	0.81	0.99	1.00	-	0.98	0.94	0.91	0.86	0.79	0.73
-0.13	0.76	0.96	0.98	0.99	-	0.99	0.97	0.93	0.86	0.80
-0.19	0.68	0.90	0.93	0.96	0.99	-	0.99	0.97	0.91	0.85
-0.24	0.58	0.83	0.86	0.90	0.95	0.99	-	0.99	0.94	0.89
-0.30	0.44	0.70	0.74	0.79	0.86	0.93	0.98	-	0.98	0.95
-0.33	0.33	0.59	0.65	0.70	0.79	0.88	0.94	0.99	-	0.99
-0.35	0.28	0.54	0.60	0.66	0.75	0.85	0.92	0.98	1.00	-

Table 3. Correlations among PTAs from test day evaluations using Legendre polynomials and including first lactation only (Lac 1) or up to five lactations (Lac5) and official PTAs (OFF) for Jersey bulls

	Milk			Fat			Protein		
	Lac1 – Lac5	Lac1 – OFF	Lac5 – OFF	Lac1 – Lac5	Lac1 – OFF	Lac5 – OFF	Lac1 – Lac5	Lac1 – OFF	Lac5 – OFF
1093 (a)	0.98	0.88	0.88	0.98	0.89	0.89	0.98	0.87	0.88
345 (b)	0.97	0.90	0.90	0.98	0.92	0.92	0.98	0.91	0.91
178	0.98	0.94	0.93	0.98	0.94	0.93	0.98	0.94	0.94

(a) All bulls with at least 1 daughter in both analysis
(b) Bulls with at least 10 daughters in TD analysis
(c) Bulls with at least 10 daughters in TD and ±5 in official

Table 4. Correlations between PTAs and components of PTAs from fitting Legendre polynomials and Wilmink function for cows with different days in milk in first lactation

Number of Cows	**DIM**	**PA**	**RC**	**PTA**
969	4-30	0.99	0.77	0.99
896	31-60	0.98	0.94	0.99
855	61-90	0.98	0.97	0.99
894	91-120	0.98	0.98	0.99
992	121-150	0.98	0.99	0.99

DIM = days in milk, PA = parent average contribution, RC = contribution from test day records

Table 5. Means of differences in PTA and components of PTA for milk yield between fitting Legendre polynomials (LP) and Wilmink function (WL)

			Mean (LP - WL)	
Number of Cows	**DIM**	**PA**	**RC**	**PTA**
969	4-30	-8.0	14.4	6.4
896	31-60	-7.5	20.7	13.2
855	61-90	-4.9	15.3	10.3
894	91-120	4.1	4.8	9.0
992	121-150	9.7	5.0	14.7

Table 6. Correlations between PTAs and components of PTAs for different sections of the lactation curve from fitting Legendre polynomials and Wilmink function for cows with different test days

		1-100 days			**101-200 days**			**201-305 days**		
Number of cows	**DIM**	**PA**	**RC**	**PTA**	**PA**	**RC**	**PTA**	**PA**	**RC**	**PTA**
969	4-30	0.99	0.89	0.99	0.98	0.76	0.98	0.98	0.61	0.98
896	31-60	0.99	0.97	0.99	0.98	0.94	0.99	0.98	0.90	0.99
855	61-90	0.99	0.99	0.99	0.96	0.97	0.99	0.97	0.95	0.98

Table 7. Means of differences of PTAs and components of PTAs for milk yield for different parts of the lactation curve fitting Legendre polynomials or Wilmink function

		1-100 days			101-200 days			201-305 days		
Number of Cows	**DIM**	**PA**	**RC**	**PTA**	**PA**	**RC**	**PTA**	**PA**	**RC**	**PTA**
969	4-30	-0.7	0.4	-0.3	-3.8	9.1	5.3	-2.3	5.2	2.9
895	31-60	-0.7	2.2	1.5	-4.4	11.1	6.7	-2.4	7.3	4.9
855	61-90	0.6	0.7	1.3	-5.0	8.8	3.8	-1.8	5.5	3.7