

Multi-Trait Test-Day model for National Genetic Evaluation of Australian Dairy Cattle for Somatic Cell Count

K.V. Konstantinov^{1,2}, K.T. Beard^{1,2}, M.E. Goddard^{2,3} and J.H.J. van der Werf⁴

¹ Australian Dairy Herd Improvement Scheme, Melbourne, Victoria 3000, Australia

² Department of Primary Industries, Attwood Victoria 3049, Australia

³ University of Melbourne, Parkville, Victoria 3052, Australia

⁴ School of Rural Science and Agriculture, University of New England, Armidale, NSW, Australia

Abstract

A random regression test day model has been developed for genetic evaluation of somatic cell scores for Australian dairy cattle. The model includes herd-test-day, year-season, age at calving, heterosis and lactation curves modelled with Legendre polynomials as fixed effects, and random genetic and permanent environmental effects modelled with Legendre polynomials. Preconditioned Conjugate Gradient algorithm with iteration on data is implemented for solving the equation system. For reliability approximation, the method of Tier and Meyer (2004) is used. The genetic evaluation system was validated with Interbull validation method III by comparing proofs from a complete evaluation with those from an evaluation based on a data set excluding the most recent four years. The genetic trend estimate was in the allowed range and correlations between the two sets of proofs were very high. Additionally, the random regression model was compared to the previous genetic evaluation model. The correlations of proofs between both models were high for bulls with high reliabilities. The correlations of bulls decreased with increasing incompleteness of daughter performance information.

Introduction

For selection against mastitis in dairy cattle, Somatic Cell Counts (SCC) transformed into Somatic Cell Scores (SCS) are chosen as an indicator trait in most countries (Mark *et al.*, 2000). Genetic evaluations for SCS are calculated in several countries and international genetic evaluations have been implemented on a routine basis by Interbull (Mark *et al.*, 2002). Test day SCS are used in the model of analysis as single tests or aggregated into lactation measures. Although differences in methods and models exist between countries, many countries have implemented random regression (RR) test-day (TD) models for genetic evaluation of herd test data (de Roos *et al.*, 2002; Liu *et al.*, 2003; Harris and Winkelman, 2004; Negussie *et al.*, 2006). RR TD models can accommodate changes in the genetic and permanent environmental effects over time by fitting polynomials of time to these effects. The advantages of RR TD models over traditional multi-trait models have been well documented in the literature (van der Werf *et al.*, 1998;

Swalve, 1994; Jamrozik & Schaeffer, 1997; Schaeffer, 2003; Liu *et al.*, 2001).

In Australia, genetic evaluations for somatic cell score (SCS) were introduced in February 2002. Currently the genetic evaluations are based on test-day SCS using the method and model described by Wiggans and Goddard (1997). Because of the benefits of RR TD multiple trait (MT) models we have chosen to develop such a model for the genetic evaluation of Australian dairy cattle for SCC.

The objectives of this paper are:

- To present estimates of genetic and environmental parameters for test-day SCC in the first three lactations of Australian dairy cattle.
- To describe the results of the development and validation of the model for the genetic evaluation of SCS for Australian dairy cattle.
- To describe the results of the Australian genetic evaluation model for SCS

Estimation of Genetic Parameters

The data used for genetic parameter estimation was extracted from the Australian national database. The data set consisted of SCC from the first three lactations of cows which calved between 1985 and 2005. Cows with more than 5 test day SCC records in the first lactation and more than 4 test day SCC records in the second and third lactations were extracted from the data set. Later lactations were only included for cows with a first lactation record. After all edits, 243,303 cows with 3,162,939 test day records were used. The natural logarithm of SCC in the first three lactations were analysed, treating each lactation as a different trait. The RR sire model for this analysis was

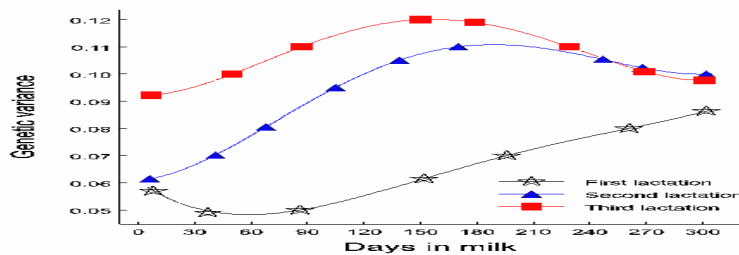
$$y = X\beta + Zs + Wc + e$$

where y is the vector of TD records for all animals; β is the vector of all fixed effects including herd-test-day (193,506, 193,789 and 189,115 effects), breeds which contained 13 subclasses of purebred and crossbred cows, 20 subclasses for age, 39 year-season subclasses and fixed regression on Legendre polynomial of order 5 on days in milk nested within parity; s , c and e are the vectors of random regressions

for sire, cow and residual effects, respectively. The matrices X , Z and W are the incidence and covariable matrices. The matrix Z contained coefficients of Legendre polynomial of order 3. The sire-maternal grand sire relationship matrix among all sires (2190) was included in the analysis. Residual variances were estimated for 10 different classes within parity: 5-30, 31-60, 61-90, 91-120, 121-150, 151-180, 181-210, 211-240, 241-270 and 271-305 days in milk. All calculations were carried out using the ASREML program (Gilmour *et al.*, 2002). Genetic variances and heritabilities for SCS at different days in milk were calculated from the estimated (co)variance matrices for the random regressions.

Daily genetic variances derived from covariance functions are presented in Figure 1. The variances increased with DIM in all lactations. In parity 1 the genetic variance decreased with DIM until DIM 60 and increased subsequently for the remaining part of the lactation. The peak in parities 2 and 3 was around DIM 150 and DIM 180. These findings are similar to those presented in earlier studies (Mrode and Swanson, 2003; Koivula *et al.*, (2004); Liu *et al.*, 2001; Negussie *et al.*, 2006).

Figure 1. Trajectory of estimated daily genetic variances for SCS in the first three lactations.



The trend of daily permanent environmental variances (Figure 2) and residual variances (table 1) was similar in all three lactations. The estimates had high

starting values and decreased with DIM. Slight increase is observed from 90 to 150 DIM. Mrode and Swanson (2003) and Negussie *et al.* (2006) reported similar results.

Figure 2. Trajectory of estimated daily permanent environment variances for SCS in the first three lactations

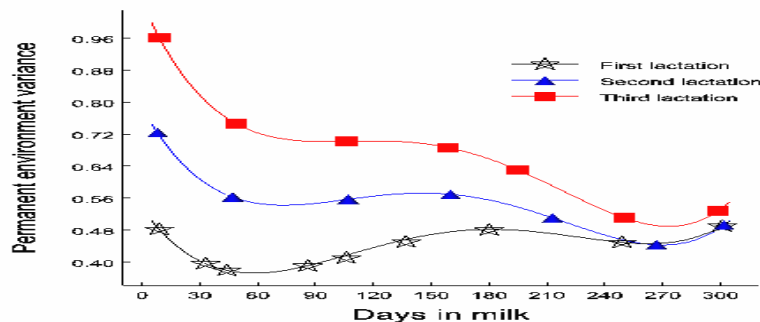


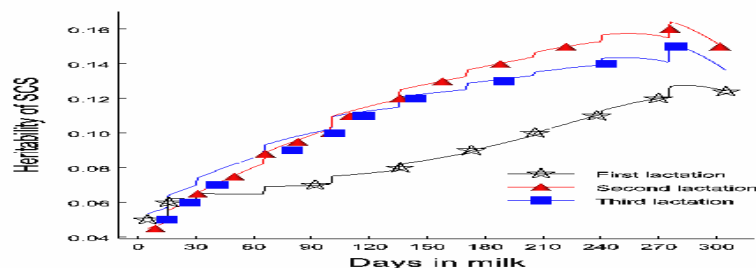
Table 1. Estimates of residual variances for different classes of days in milk.

Days in milk	Parities		
	1	2	3
5 – 30	0.613	0.602	0.700
31- 60	0.407	0.446	0.521
61 – 90	0.362	0.398	0.451
91 – 120	0.314	0.336	0.378
121 – 150	0.269	0.276	0.306
151 – 180	0.240	0.238	0.253
181 – 210	0.233	0.216	0.227
211 – 240	0.222	0.202	0.215
241 – 270	0.210	0.185	0.194
270 – 305	0.175	0.147	0.143

The estimates of the daily heritabilities, for different stages of lactation in all three parities, are in Figure 3. Generally, the estimates increased with days in milk in all lactations.

The heritabilities increased gradually, starting from 0.04 and ending to 0.16. These values are in a good agreement with the values reported in the literature.

Figure 3. Trajectory of estimated daily heritabilities for SCS in the first three lactations.



Genetic correlations within lactations for some selected days in milk are in table 2. The genetic correlations within each parity were highest between adjacent DIM, varying from 0.90 to 0.98. The decline in the genetic

correlation as DIM got further apart was greatest in the first parity. For instance, the correlation between day 45 and 285 in lactation 1 was 0.65, compared with 0.71 and 0.66 in parities 2 and 3, respectively.

Table 2. Heritabilities (on diagonal), genetic correlations (above diagonal) and phenotypic correlations (below diagonal) on selected days in milk for SCS in lactations 1, 2 and 3.

Parity		1					2					3				
	DIM	15	45	105	195	285	15	45	105	195	285	15	45	105	195	285
1	15	0.05	0.97	0.80	0.59	0.49	0.75	0.67	0.53	0.40	0.31	0.67	0.60	0.48	0.37	0.28
	45	0.43	0.06	0.92	0.76	0.65	0.83	0.79	0.69	0.58	0.45	0.76	0.72	0.63	0.53	0.41
	105	0.34	0.47	0.07	0.94	0.84	0.86	0.89	0.87	0.80	0.65	0.81	0.82	0.81	0.74	0.59
	195	0.24	0.40	0.58	0.10	0.96	0.81	0.89	0.92	0.91	0.80	0.76	0.82	0.86	0.85	0.72
	285	0.20	0.29	0.43	0.59	0.12	0.77	0.84	0.89	0.92	0.89	0.71	0.77	0.83	0.85	0.79
2	15	0.09	0.13	0.16	0.18	0.18	0.06	0.98	0.88	0.77	0.62	0.95	0.93	0.85	0.73	0.57
	45	0.10	0.15	0.19	0.23	0.23	0.52	0.07	0.96	0.88	0.71	0.95	0.97	0.93	0.84	0.66
	105	0.09	0.15	0.23	0.28	0.30	0.43	0.54	0.11	0.97	0.81	0.88	0.94	0.97	0.94	0.76
	195	0.09	0.15	0.23	0.31	0.34	0.31	0.45	0.63	0.14	0.92	0.76	0.85	0.94	0.98	0.87
	285	0.08	0.12	0.18	0.25	0.33	0.21	0.29	0.41	0.58	0.16	0.57	0.67	0.78	0.90	0.97
3	15	0.09	0.12	0.14	0.15	0.15	0.16	0.18	0.20	0.20	0.16	0.06	0.98	0.90	0.77	0.57
	45	0.09	0.13	0.16	0.18	0.19	0.17	0.21	0.25	0.26	0.21	0.59	0.08	0.97	0.87	0.66
	105	0.09	0.13	0.18	0.22	0.23	0.16	0.22	0.29	0.33	0.29	0.49	0.60	0.11	0.96	0.77
	195	0.08	0.13	0.18	0.24	0.26	0.15	0.22	0.31	0.38	0.37	0.35	0.49	0.66	0.13	0.90
	285	0.08	0.11	0.15	0.20	0.26	0.12	0.16	0.23	0.33	0.42	0.23	0.30	0.42	0.59	0.15

The phenotypic correlations (table 2, lower triangle) were similar in pattern to the genetic correlations. Again higher correlations were observed between adjacent DIM with gradual decline as DIM got further apart. The within parity correlations were lower than the corresponding genetic correlations.

The genetic correlations across lactations are shown in table 2. Similar to other studies (Haile-Mariam *et al.*, 2001; Mrode and Swanson (2003); Negussie *et al.* (2006)), at the same stage of lactation the correlations were highest and decreased gradually as DIM got further apart. For instance, the genetic correlation between DIM 45 in parity 1 and 2 was 0.79, but decreased to 0.45 between DIM 45 in parity 1 and DIM 285 in parity 2. The estimates of the genetic correlations for the same DIM in parities 1 and 2 and parities 1 and 3 were high and ranged from 0.67 to 0.92. For all lactations, the correlations were highest at the beginning of the lactations and decreased with an increase in DIM. The correlations between parity 1 and 3 were lower than the correlations between parity 1 and 2.

The phenotypic correlations between parities were low, varying from 0.08 to 0.42. The correlations between parities 2 and 3 were higher, compared to those between parity 1 and 2, and 1 and 3. The phenotypic correlations for DIM 285 were 0.33, 0.26 and 0.42 between parities 1 and 2, 1 and 3 and 2 and 3, respectively.

Statistical Model for National Genetic Evaluation

The model for the national genetic evaluation of SCS was multiple-trait (MT) random regression (RR) test-day (TD) model where each of the first three lactations was considered as a separate trait. The RR MT TD animal model for this analysis was

$$y = X\beta + Za + Wp + e$$

where y is the vector of TD records for all animals; β is the vector of all fixed effects including herd-test-day, heterosis effect modelled as covariable, age effect, year-season effect and fixed regression on Legendre polynomial of order 5 on days in milk nested within parity; a , p and e are the vectors of random regressions for animal, permanent environmental and residual effects, respectively. The matrices X , Z and W are the incidence and covariable matrices. The matrix Z and W contained coefficients of Legendre polynomial of order 3. The animal relationship matrix with phantom parent groups was included in the analysis. The number of groups varied from 70 to 78.

Computational Strategy

New software was developed in C++ with extensive use of Blitz library. The mixed

model equations (MME) were solved using preconditioned conjugate gradient (PCG) method (Shewchuk, 1994; Strandén and Lidauer, 1999) and iteration on data. The preconditioner matrices were block diagonal. Their inverses were stored in memory in sparse matrix format.

The convergence indicator for the method was the relative average difference between the right hand and left hand sides. For a linear system $\mathbf{Ax} = \mathbf{b}$, the convergence indicator was

$$c^n = \frac{\| \mathbf{b} - \mathbf{Ax}^{(n+1)} \|}{\| \mathbf{b} \|}$$

where $\| \mathbf{b} \|$ was the Euclidian norm of \mathbf{b} .

To allow comparisons between results from different data sets, we first investigated how small the values of \mathbf{c} should be to reach the accuracy of the solutions sufficient in

practical breeding work. Therefore we calculated a quasi-true ABV obtained by performing PCG iterations until \mathbf{c} became smaller than 10^{-10} . Intermediate ABV for various \mathbf{c} were obtained from corresponding solutions of MME. For each \mathbf{c} value the correlation between intermediate and the quasi-true ABV was calculated as well as the average difference and the standard deviation. Solutions were considered as converged if the standard deviation of the difference between quasi-true ABV and the intermediate ABV was less than 10^{-3} .

Table 3 presents means and standard deviations of the difference between quasi-true solutions and different intermediate solutions. The results from this analysis suggest that the accuracy of the solutions of 10^{-7} should be sufficient in practical breeding work. This accuracy has been used for all subsequent comparisons.

Table 3. Means and standard deviations of the difference between quasi-true solutions and intermediate solutions for 11,153 Holstein bulls.

Difference in c	Total ABV		Parity 1 ABV		Parity 2 ABV		Parity 3 ABV	
	mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.
10^{-10} vs 10^{-5}	0.5706	0.0050	0.4655	0.0044	0.6114	0.0054	0.6350	0.0060
10^{-10} vs 10^{-6}	0.1408	0.0011	0.1221	0.0010	0.1487	0.0011	0.1518	0.0013
10^{-10} vs 10^{-7}	0.0335	0.0004	0.0268	0.0003	0.0357	0.0005	0.0378	0.0006
10^{-10} vs 10^{-8}	0.0046	0.0001	0.0015	0.0001	0.0056	0.0002	0.0067	0.0002

National evaluation data

Several routine data sets were extracted from Australian Dairy Herd Improvement Scheme (ADHIS) database. The first two data sets contained the first three lactations for all cows up to February 2006 and May 2006.

For trend validation the extraction was for cows up to August 2002 and August 2006. Additionally a data set containing 77 herds was extracted and used for validation of repeatability procedure implemented into the software.

Cows had approximately 6 herd tests per lactation on average. Contemporary group was defined as herd-test-day within each lactation. The breed compositions were Holstein-Friesian, Jersey, Guernsey and Australian Red cattle.

National evaluation results

From practical point of view it is important to know the differences between estimated breeding values (EBV) from two subsequent runs. It is expected that the rank correlations between the two EBVs should be very high. A series of rank correlations between EBVs from February and May 2006 were calculated for different groups of animals and the results are presented in tables 4 to 6.

Rank correlations for all cows across breeds were very high (0.99) and for all bulls ranged from 0.93 to 0.98. Within breed rank correlations for cows and bulls were also high. The rank correlations for all Holstein bulls ranged from 0.90 to 0.98. These results suggest that some re-ranking has occurred in May 2006 EBVs. In order to establish in which group of sires the re-ranking occurred, we have divided

the sires into four groups: “proven” sires – bulls with more than 100 daughters; sires about to be published – bulls with $20 < \text{daughters} < 40$; “unpublishable” sires – bulls with $1 < \text{daughters} < 10$; and young bulls with no daughters.

The rank correlations for all “publishable” sires for major breeds are in table 4.

The correlations are very high (0.999) suggesting that there were no re-ranking within this group of sires. Similar picture was observed for sires about to become publishable (table 5). The rank correlations for sires with low number of daughters ranged from 0.98 to 0.99 (table 6). The lowest rank correlations (0.94 to 0.96) were found for young sires with no daughters.

Table 4. Rank correlations between breeding values from February 2006 and May 2006 for bulls with more than 100 daughters.

Type of breeding value	Holstein		Jersey	
	Correlation	Number of animals	Correlation	Number of animals
Total	0.9991	2,832	0.9991	550
I lactation	0.9986	2,832	0.9981	550
II lactation	0.9988	2,832	0.9989	550
III lactation	0.9985	2,832	0.9987	550

Table 5. Rank correlations between breeding values from February 2006 and May 2006 for bulls with $20 < \text{daughters} < 40$.

Type of breeding value	Holstein		Jersey	
	Correlation	Number of animals	Correlation	Number of animals
Total	0.9969	4,794	0.9955	1,456
I lactation	0.996	4,794	0.9928	1,456
II lactation	0.9968	4,794	0.9953	1,456
III lactation	0.9965	4,794	0.996	1,456

Table 6. Rank correlations between breeding values from February 2006 and May 2006 for bulls with $1 < \text{daughters} < 10$.

Type of breeding value	Holstein		Jersey	
	Correlation	Number of animals	Correlation	Number of animals
Total	0.988	20,919	0.9944	9,324
I lactation	0.9756	20,919	0.9861	9,324
II lactation	0.9884	20,919	0.9953	9,324
III lactation	0.9911	20,919	0.9958	9,324

The results from these analyses clearly indicate that the EBVs are stable from run to run. As expected the most affected groups of animals were the sires with low number of progeny and the young bulls.

RR TD model was validated using Interbull method 3 and EBVs from August 2002 and 2006. The results did not show a significantly different from zero effect from new daughters for all breeds.

The data from May 2006 with 77 herds only was used for validation of the method for calculation of approximate reliabilities implemented into the new system. We used the method published by Tier and Meyer (2004). “True” reliabilities were obtained from direct inversion of the MME coefficient matrix using sparse matrix inverse. Methods were compared by regressing one on another. Table 7 shows close correspondence between “true” and approximate reliabilities.

Table 7. Comparison of “true” and approximate reliabilities of estimated breeding values.

Mean reliabilities	
Inverse	0.288
Approximate	0.299
Regression of inverse on approximate	
Regression coefficient	1.040
Intercept	-0.024
R ² (%)	97.4
Total number of animals	24,577

Reliabilities are overestimated by only 1% on average by the approximate method of Tier and Meyer (2004). This is shown by the mean values, which are higher for the approximate method than those obtained from sparse inverse method. The intercept is zero and the linear coefficient is slightly > 1.0 , indicating that the higher the reliability the more it is overestimated. The regression model accounts for most (97.4%) of the variation. These results are in a good agreement with the results from Tier and Meyer (2004).

Conclusion

Random regression MT TD model performed well during the test runs. The correlations between the consecutive runs were high (0.98-0.99). There was no evidence for over or under estimation of the genetic trend in the population. The time for calculation of the breeding values is 30-35% less than the time needed for the current method. The method will be officially implemented in April 2008.

References

- De Roos, A.P.W., Harbers, A.G.F. & de Jong, G. 2002. Herd specific random regression curves in a test-day model for protein yield in dairy cattle. *Proc. 7WCGALP*, Communication #01-05.
- Gilmour, A.R., Cullis, B.R., Welham, S.J. & Thompson, R. 2002. ASReml User Guide. *NSW Agr. Biom. Bull.* 3.
- Haile-Mariam, M., Goddard, M.E. & Bowman, P.J. 2001. Estimates of Genetic Parameters for Daily Somatic Cell Count of Australian Dairy Cattle. *J. Dairy Sci.* 84, 1255-1264.
- Harris, B.L. & Winkelman, A.M. 2004. Test-Day Model for National Genetic

- Evaluation of Somatic Cell Count in New Zealand. *Proc. 2004 Interbull Mtg. Interbull Bulletin* 32, 101-104.
- Jamrozik, J., 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.
- Liu, Z., Reinhardt, F. & Reents, R. 2001. Parameter estimates of a random regression test day model for first three lactation somatic cell scores. *Interbull Bulletin* 26, 61-65.
- Liu, Z., Reinhardt, F. & Reents, R. 2003. Genetic evaluation for somatic cell scores using random regression test-day model for a very large dairy cattle population. *Interbull Bulletin* 31, 88-91.
- Mark, T., Fikse, F., Banos, G., Emanuelson, U. & Philipsson, J. 2000. Summary of Mace pilot runs for somatic cell count and mastitis. *Interbull Bulletin* 26, 43-52.
- Mark, T., Fikse, W.F., Emanuelson, U. & Philipsson, J. 2002. International genetic evaluation of Holstein sires for milk somatic cell and clinical mastitis. *J. Dairy Sci.* 85, 2384-2392.
- Koivula, M., Negussie, E. & Mantysaari, E.A. 2004. Genetic parameters for test-day somatic cell count at different lactation stages of Finnish dairy cattle. *Livest. Prod. Sci.* 90, 145-157.
- Mrode, R.A. & Swanson, G.J.T. 2003. Estimation of genetic parameters for somatic cell count in the first three lactations using random regression. *Livest. Prod. Sci.* 79, 239-247.
- Negussie, E., Koivula, M., Mantysaari, E.A. & Lidauer, M. 2006. Genetic evaluation of somatic cell score in dairy cattle considering first and later lactations as two different but correlated traits. *J. Anim. Breed. Genet.* 123, 224-238.

- Shewchuk, J.R. 1994. An introduction to the conjugate gradient method without the agonizing pain. *School of Computer Sci., Carnegie Mellon Univ. Pittsburgh, PA.*
- Swalve, H.H. 1994. Estimation of breeding values using test day models. Proc. INTERBULL Mtg., Ottawa. ON, Canada. August 5-6, 1994. INTERBULL, Uppsala, Sweden. *Interbull Bulletin 10*, 13 pp.
- Schaeffer, L.R. 2003. Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86, 35-45.
- Strandén, I. & Lidauer, M. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. *J. Dairy Sci.* 82, 2779-2787.
- Van der Werf, J.H.J., Goddard, M.E. & Meyer, K. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *J. Dairy Sci.* 81, 3300–3308.
- Wiggans, G.R. & 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.