

Development of a National Production Testday Model for New Zealand

B. L. Harris, A. M. Winkelman, D. L. Johnson and W. A. Montgomerie

*Livestock Improvement Corporation Ltd,
Private Bag 3016, Hamilton, New Zealand*

Introduction

Genetic evaluation for milk, fat and protein yields, using 270-day lactation yields in an across-breed animal model (AM), has been performed in New Zealand (NZ) since 1996 (Harris *et al.*, 1996). A new system has been developed to use the milk production herd-test data to calculate breeding values (BVs) for the dairy industry. In the current AM, a 270-day lactation yield record is calculated based on 1 to 10 testday records. The environmental effect of herd-testday is removed during 270-day lactation yield calculation (Johnson, 1996). However, this is done outside the animal model, without reference to the genetic level of the herd-testday contemporary group. A testday model (TDM) can simultaneously account for the fixed effects such as herd-testday contemporary groups, and the genetic, permanent environmental (PE) and temporary environmental (TE) random effects that affect the trait. The TDM allows the environments specific to each test day to be better taken into account, resulting in an improved accuracy of evaluation. A TDM can accommodate changes in the genetic and PE effects over time by fitting polynomials of time to these effects. In principle the TDM uses a model for each animal that allows each cow's lactation curve to have a different shape. The changes over time can be modelled by fitting Legendre polynomials (LPs) of days in milk (DIM) to the random effects in the model, thus creating a random regression (RR) model. LPs are a popular choice because they do not make prior assumptions about the shape of the lactation curve and the orthogonality of increasing orders of fit means that a higher order is not collinear with a lower order. TDM evaluations are more stable than evaluations using 270-day lactations because they account for genetic differences in maturity rate and persistency of lactation.

Many countries have implemented, or are moving towards, TDMs for genetic evaluation of herd-test data (de Roos *et al.*, 2003; Liu *et al.*, 2003; Schaeffer *et al.*, 2000). This document describes the statistical model used for the calculation of BVs and the results of the RR TDM applied to national data in NZ.

Data

Milk production data were extracted from Livestock Improvement's national dairy herd database for this analysis. Included in the data were all testday records stored electronically. Records from season 1986 through to season 2005 were included. Pedigree and breed data from the routine genetic evaluations for production were used. Ancestors were traced back to 1939. Animals born before 1960 were considered to be the base population. A total of 30,727,517, 25,757,202, 22,757,202 and 47,084,797 first, second, third lactation and fourth to sixth lactation testday records from 16,697,703 animals (cows and ancestors) were analysed. On average, cows had approximately 3.5 herd tests per lactation. Contemporary group was defined in the model as herd-year-season-testday within each of lactations 1, 2, 3 and 4-6. Season of calving was defined as spring or autumn. In NZ, the majority of cows (97%) calve in the spring. The average number of cows within a herd-year-season-lactation was 44.2. The breed compositions were approximately 56% Friesian, 17% Jersey, 22% Friesian-Jersey crosses and 1.5% Ayrshires with the remainder being other breeds and other crosses. The effect of induction of calving is included in the model. Less than 0.5% of first lactation cows and around 5% of later lactation cows were induced. The phenotypic averages and standard deviations by lactation are given in Table 1.

Genetic Parameter Estimation

Variance components were estimated using a RR TDM in which LPs of DIM were fitted to the sire and PE (cow) effects. Data for the analyses were obtained from Livestock Improvement's Sire Proving Scheme herds in seasons 1995 to 2004. The data set contained first-lactation records on 165,871 cows (progeny of 2427 sires) and subsequent lactations, up to lactation six, of these cows, where available. The breed composition of the cows was similar to that of the national population. Testday records consisted of milk, fat and protein yield measurements.

The RR analyses were done separately for each production trait (milk, fat, protein yields) using multiple-trait (MT) models in which lactations one through four were modelled as different traits for the sire effect, with lactations five and six treated as the same genetic trait as lactation four. Lactations one through six were modelled as separate traits for the PE effect. Third-order LPs (order is degree + 1) of DIM were fitted to the sire and PE (cow) effects and the covariance functions (CFs) for each of these effects were estimated using ASREML (Gilmour *et al.*, 2002). A sire model was used where the pedigree included the sire and maternal grandsire of each sire.

Starting values for the CFs used in the RR models were obtained by fitting LPs to the sire and phenotypic variances estimated within discrete segments of time using the approach of Kirkpatrick *et al.* (1990). MT sire models, where the 270-day lactation was subdivided into six traits based on 45-day intervals (first interval from 3 to 45 DIM), were used to estimate the sire and phenotypic variances. These analyses were done within lactation for the first three lactations for milk, fat and protein yields. Orders of fit of the LPs of one through five were investigated. The chi-squared statistic showed that there was no advantage of using a polynomial of order greater than four for the sire effect in all three production traits. However, as the shape of the curves of the variances over the lactation differed little when going from order 3 to order 4 LPs, the decision was made to use order 3 LPs for the RR analyses. Fitting LPs to the phenotypic variances involved fitting the variance associated with temporary environment (TE),

which was fitted using a third-order LP. The estimates of the associated variance function were used to model the heterogeneous TE variance for the RR analysis. The same variance structure for TE was used for lactations three and higher. The starting values for the CF associated with the PE effects were obtained by subtracting the CF for the sire effect from the CF for the phenotypic effect (excluding the TE effect).

Statistical Model for National Genetic Evaluation

The model for the national genetic evaluation of each production trait was a MT RR TDM where lactations 1, 2, 3 and lactations 4 through 6 are modelled as different genetic traits. Separate random PE effects were fitted for each of the 6 lactations. A testday production record, measured on day 'd' (3 to 270) of lactation, was modelled as:

$$y_{ijkmn} = htd_{ij} + \sum_{q=1}^4 ab_{ina} w_{inq}^a + ind_i + \sum_{s=1}^6 ht_{is} w_{ins}^h + \sum_{t=2}^5 \phi_t(d) s_{it} + \sum_{t=1}^3 \phi_t(d) a_{in} + \sum_{t=1}^3 \phi_t(d) p_{mn} + e_{ijkmn}$$

y_{ijkmn} is the kth record for ith trait of animal n ; i denotes trait 1 to 4,

htd_{ij} is the jth testday herd-year-season fixed effect for trait i , with season referring to spring or autumn calving period,

ind_i is the effect of induced calving of in lactation i ,

ab_{inq} is the linear regression coefficient of age in days at calving, nested within breed of animal n for trait i ,

w_{inq}^a is the qth covariate value for age at calving by breed proportion of animal n ,

ht_{is} is the linear regression coefficient for the sth heterosis effect for trait i ,

w_{ins}^h is the sth heterosis covariate value for animal n ,

$\phi_t(d)$ is the standardised LP of order t calculated at days in milk d ; $t = 2$ to 5 for the fixed effect, 1 to 3 for the random effects,

s_{it} is the stage of lactation fixed effect for order t of trait i

a_{in} is the random additive genetic effect for trait i of animal n ,
 p_{mn} is the random permanent environment effect for lactation m of animal n ,
 e_{ijkmn} is the random residual associated with record y_{ijkmn}

For the national milk production evaluation, genetic groups were assigned by breed, gender of missing parent, birth year and country of origin. Four breed classes were assigned genetic grouping, namely, Holstein-Friesian (HF), Jersey, Ayrshire-Red, and other breeds. Genetic groups were assigned in 5 year intervals from 1960 to 1980 then yearly, with the first birth year group being prior to 1960. Country of origin was defined as NZ, North American and Other. If a genetic group had fewer than 200 animals per group, birth years were clustered. No clustering occurred across breed, origin or gender genetic groups.

BVs were calculated for days 3 to 270 within each lactation. Results for the BV averaged over the 4 traits are reported. The eigenfunctions of the genetic covariance function (Kirkpatrick *et al.* 1990) from the RR analysis describe the aspects of the lactation curve that can be changed with selection. The second eigenfunctions show the trade-off between early and late lactation and are therefore associated with persistency of lactation. In this study the persistency BV was calculated as the sum of the BVs at each DIM of the lactation weighted by the associated value of the second eigenfunction.

Computational Strategy

The mixed model equations (MMEs) were solved using a preconditioned conjugate gradient solver (Stranden and Lidauer, 1999) and iteration on data with code reordering (Tsurata *et al.*, 2001). The MME for each of the yield traits contained approximately 420 million equations. The MME solver was run for 5000 iterations on a 2.6 Ghz AMD Opteron CPU with 24 Gb of RAM. The large number of iterations was done to ensure a converged system of equations, since these solutions will be the starting basis of future runs. Future runs will use prior solutions and will thus require considerably fewer iterations. Each iteration required 3.7 minutes to complete.

Table 1. Phenotypic means (μ) and standard deviations (σ) for the production testday records.

L [†]	Milk (l)		Fat (kg)		Protein (kg)	
	μ	σ	μ	σ	μ	σ
1	12.	4.4	0.6	0.1	0.4	0.1
	8	8	1	8	6	4
2	14.	5.3	0.7	0.2	0.5	0.1
	9	5	1	2	4	8
3	16.	5.9	0.7	0.2	0.5	0.2
	3	8	7	4	9	0
4-6	17.	6.3	0.8	0.2	0.6	0.2
	0	5	0	6	2	1

[†]L=Lactation

Results and Discussion

Table 2 provides summary statistics of the TDM yield and persistency BVs for milk, fat and protein yields, within breed, for sires enrolled in the national genetic evaluation program. The ranges of the yield BVs from the TDM model are similar to current values for these yield BVs from the national AM evaluation. The difference between the Jersey and HF breed average estimates is slightly higher in the TDM analyses than those obtained from the AM. This difference could be due TDM accounting for differences in maturity and lactation curve shape between Jersey and HF cows. The Jersey sires on average have the highest persistency breeding values for all three milk production traits, whereas the HF sires have the lowest. The correlations among TDM and AM BVs are given in Table 3 within and across breed for milk, fat and protein, respectively. The correlations show a high degree of agreement between the BVs from both models which is consistent with other studies that have compared TDM and AM BVs.

The hybrid vigour estimates from the TDM, averaged across 4 lactations, are provided in Table 4. Although the values for individual lactations increase in absolute value with increasing maturity, they are consistent in magnitude to the average values when expressed as a percentage of average production. The values presented in Table 4 are consistent with values previously reported by Harris and Kolver (2001) and Harris *et al.* (1996).

A national across-lactation TDM evaluation using RR appears to be computationally feasible. The TDM model results are in good agreement with the results from the current national AM evaluation. The proposed implement date for a national TDM evaluation is February 2007.

Table 2. Sire TDM BV summary statistics for the production testday records.

	μ	σ	Min	Max
Ayrshire				
Milk	345	317	-593	1318
Fat	27.4	14.0	-37.9	42.6
Protein	9.3	10.0	-19.8	41.9
Persist†	-92	87	-387	151
Holstein Friesian				
Milk	961	381	-531	2077
Fat	23.0	13.4	-29.4	66.9
Protein	25.3	12.1	-20.2	65.9
Persist†	-165	86	-458	185
Jersey				
Milk	-384	348	-1513	592
Fat	7.2	17.0	-51.3	60.4
Protein	-2.5	11.3	-42.3	36.5
Persist†	68	75	-409	193

†Persist=Milk persistency BV

Table 3. Correlations among TDM and AM BVs for the production traits.

Breed	Milk	Fat	Protein
All breeds	0.991	0.985	0.988
Ayrshire	0.960	0.971	0.957
HF	0.982	0.980	0.985
Jersey	0.977	0.989	0.979

Table 4. Hybrid vigour estimates from the TDM for production traits§.

Cross	Milk	Fat	Protein
HF x Ay	2.7	3.3	3.1
HF x J	4.5	6.1	5.7
J x Ay	4.8	6.3	5.7

†J=Jersey; HF=Holstein-Friesian; AY=Ayrshire

§ Averaged across lactations and expressed as a percentage of average production for the first-cross animal.

References

- De Roos, A.P.W, Harbers, A.G.F. & de Jong, G. 2003. Genetic parameters of testday somatic cell score estimated with a random regression model. *Interbull Bulletin* 31, 97-101.
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. & Thompson, R. 2002. *ASREML User Guide Release 1.0*, VSN International Ltd., Hemel Hempsted, HP1 1Es, UK.
- Harris, B.L., Clark, J.M. & Jackson, R.G. 1996. Across breed evaluation of dairy cattle. *Proceedings of the New Zealand Society of Animal Production* 56, 12-15.
- Harris, B.L. & Kolver, E.S. 2001. Review of Holsteinization in intensive pastoral dairy farming in New Zealand. *J. Dairy Sci.* 84 (E. Suppl.), E56-61.
- Johnson, D.L. 1996. Estimation of lactation yield from repeated measures of test day yields. *Proceedings of the New Zealand Society of Animal Production* 56, 16-18.
- Kirkpatrick, M., Lofsvold, D. & Bulmer, M. 1990. Analysis of inheritance, selection and evolution of growth trajectories. *Genetics* 124, 979-993.
- Liu, Z., Reinhardt, F., Bunger, J., Jaiter, J. & Reents, R. 2003. Genetic evaluation of somatic cell scores using a random regression test day model for a very large dairy cattle population. *Interbull Bulletin* 31, 88-91.
- Schaeffer, L.R., Jamrozik, J., Kistemaker, G.J. & Van Doormaal, B.J. 2000. Experience with a testday model. *J. Dairy Sci.* 83, 1135-1144.
- Stranden, I. & Lidauer, M. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. *J. Dairy Sci.* 82, 2779-2787.
- Tsuruta, S., Misztal, I. & Strandén, I. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. 2001. *J Anim Sci.* 79, 1166-1172.