

Multiple Trait Random Regression Test Day Model for Production Traits

J. Jamrozik^{1,2}, L.R. Schaeffer¹, Z. Liu² and G. Jansen²

¹ *Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science
University of Guelph, Guelph, ON, N1G 2W1, Canada*

² *Canadian Dairy Network, 150 Research Line, Suite 307
Guelph, Ontario, N1G 4T2, Canada*

Introduction

Research on test day (TD) models began in Canada in 1991. At that time fixed regressions (within age and season of calving) were used to account for the shape of the lactation curve of dairy cows, but the animal effect was used only to account for differences in height of these curves (Ptak and Schaeffer, 1993). Fixed regression TD model was applied to Canadian dairy goat data (Schaeffer and Sullivan, 1994) and to Canadian data on somatic cell score (SCS) (Reents et al., 1995a,b).

Schaeffer and Dekkers (1994) suggested the possibility of using random regressions in a linear model (Henderson, 1982) for analysing TD data. Single trait random regression models were applied to first lactation milk, fat and protein records, with different functions for fixed and random regressions (Jamrozik and Schaeffer, 1997; Jamrozik et al., 1997a,b). In the simulation study of Kistemaker (1997) random regression models were significantly better than an analysis of 305d yields in terms of correlation between estimated and true breeding values. A 2-3% increase in accuracy for bulls and 6-8% for cows was found for first lactation milk yield.

Changes occurring in Canadian milk recording system will require an application of multiple trait test day model in genetic evaluation for dairy production traits. Future milk recording programs would have TD records that have all yield values (milk, fat, protein, SCS) while other TD records might have only milk yield. Thus the multiple trait model for simultaneous analysis of all yields seems to be a logical choice. Also, preliminary work with milk yield in different lactations showed that lactation curves were different between first and second lactation, and again between second and

later lactations. Thus, yields in different lactations should also be considered as different traits.

The objective of this research was to develop the multiple trait, random regression TD model for genetic evaluation of dairy bulls and cows for production traits. Both animal genetic and permanent environmental effects were modelled by random regressions in this model. Some preliminary results are presented.

Model

Let the production record on cow j on a particular day in milk t in lactation n be denoted as

$$y_{nj} = \begin{bmatrix} y_{nj1} \\ y_{nj2} \\ y_{nj3} \\ y_{nj4} \end{bmatrix} = \begin{bmatrix} 24h \text{ Milk Yield, kg} \\ 24h \text{ Fat Yield, kg} \\ 24h \text{ Protein Yield, kg} \\ \text{SCS} \end{bmatrix}$$

One or more of these traits may be missing due to the sampling scheme, but milk yield is always assumed to be present. The 24h yield may be actual 24h weights, or estimated 24h yields based on AM/PM yields or samples from 3x milkings.

The model equation is assumed to be the same for all four traits. For trait h in lactation n it is

$$y_{hntijkl} = \text{HTD}_{hni} + \sum b_{hnm} z_{tm} + \sum a_{hnm} z_{tm} + \sum p_{hnm} z_{tm} + e_{hntijkl}$$

where

$y_{hntijkl}$	is record l on cow j made on day t in herd-test day i , for a cow belonging to subclass k for region, age and season of calving
HTD_{hni}	is fixed herd-test day effect
b_{hnmk}	are fixed regression coefficients for subclass k of region, age and season of calving
p_{hnmj}	are random regression coefficients for permanent environmental (PE) effect on cow j
a_{hnmj}	are random regression genetic coefficients specific to cow j
$e_{hntijkl}$	is residual effect for each observation
Z_{tm}	are covariates, assumed to be the same for fixed and random regressions

Wilink's function (Wilink, 1987) was chosen to describe the shape of lactation curves. The function is $W(t) = w_0z_{t0} + w_1z_{t1} + w_2z_{t2}$, where $z_{t0} = 1$, $z_{t1} = t$, $z_{t2} = \exp(-0.05t)$.

The full model includes TD records from all lactations, but traits within lactations are assumed to be separate traits, except for lactation 3 and later. Thus the model is a 4 trait model with separate effects for lactations 1, 2, 3+ specified within the equation of the model.

Because each y_{ntj} are separated in time, the residuals are assumed to be uncorrelated both within and between cows. The residual covariance matrix between traits on the same TD is R_{nt} . Different residual covariances are allowed for different lactations (n) and time period within lactation (t), defined as 5 to 45 DIM, 46 to 115 DIM, 116 to 265 DIM, and 266 to 305 DIM. The diagonal elements of R_{nt} will be appropriately weighted to reflect relative accuracy of 24h yields.

PE effects are modelled by random regressions so they are allowed to vary within lactation. Let p_j represent the 36 by 1 vector of PE regression coefficients for a cow j with covariance matrix P of order 36x36. The PE covariance matrix for all cows is $I \otimes P$. Environmental correlations between TD records on the same cow on different days are accounted for by PE effect.

Let a_j represent the 36x1 vector of random genetic regression coefficients for animal j . The covariance matrix for one animal is G of order 36 by 36. The covariance matrix for all animals is $A \otimes G$. Thus, the animal regression coefficients are genetically related between animals, between lactations, and between traits within lactation.

Through these correlations it is possible for a cow to have a single TD milk yield and to have genetic evaluations for all traits and lactations.

Material and methods

The model was applied to TD records on milk, fat, protein yields and SCS (on the \log_2 scale) in the first three lactations of Canadian Holsteins. Data were 11,544,946 TD records on 786,894 cows, calved after January 1st, 1988. Cows had to have at least one TD milk record in first lactation. First test in each lactation had to be recorded before 90 DIM. Only records collected between DIM 5 and 305 were included. Age at calving was restricted to 18 - 68 months. Cows were assigned to one of 19 subclasses for age at calving within lactation and one of two seasons of calving (April- September, October-March). Combined with region (Ontario, Maritimes and Quebec, Prairies, British Columbia) this gave 152 region-age-season of calving subclasses. Herd-test date effect (1,219,679 classes) was defined within parity, but records in second and third lactations in a herd on a given TD were assigned to the same HTD class. The total number of animals evaluated (including ancestors from pedigrees) was 1,520,096. Phantom parent groups (23 levels) were formed for missing parents on the basis of sex, birth year and origin (Canadian or foreign). Inbreeding was accounted for through A^{-1} matrix.

Genetic, PE and residual (co)variances were estimated on a subset of the current data. In total, 54,770 TD records with all traits recorded, on 3297 cows from Ontario and Quebec were used. The model for the estimation of parameters did not include phantom parent groups or inbreeding coefficients in relationships. Gibbs technique was applied to generate samples for posterior distributions of all variances and covariances (1452 parameters) of the model. After burn in, 20,000 samples were used to estimate means for all (co)variances of the model.

Mixed model equations for the multiple trait random regression model on the full data set (over 87 million equations) were solved by iteration on data with Gauss-Seidel and block iteration techniques. Two copies of the data set (sorted by HTD and cow, respectively) and inverted diagonal matrices for each block were read in each iteration. During the first run through the data solutions for

fixed regressions and HTDs were updated. PE and animal random regression coefficients were updated while reading the data sorted by cow. Fast input/output C routines were used to read data effectively. Pedigree data was stored in memory with animals ordered from youngest to oldest.

Breeding values for yields in 305d lactations were obtained from estimates of random regression coefficients of animal effects. They were expressed on a mature equivalent basis and weighted by lactation weights (1/3 for all lactations) to calculate a combined proof. EBVs for SCS were expressed as an average daily score within lactation and combined into an overall index (weights 0.25, 0.65, 0.10 for lactations 1, 2 and 3, respectively). Persistency was defined for milk yield in first lactation as $110 \times (\text{EBV}(280) - \text{EBV}(6))$ (Jamrozik and Schaeffer, 1997).

Estimated breeding values from the current model were compared with official Canadian evaluations from February 27. Some comparisons were also made between the current model and results obtained from a simplified version, with constant PE effect, applied to the same data set.

Results and discussion

The convergence of Gibbs chain for co(variance) components estimation was checked by visual inspection of plots of samples. Different values of prior and starting co(variances) gave similar estimates of means. The smallest number of independent samples for a single component was 38.

Estimates of genetic, PE and residual variances and heritabilities on selected DIM for first lactation milk yield are presented in Table 1.

Table 1. Estimates of genetic (G), permanent environmental (PE) and residual (E) variances, and heritabilities (h^2) for milk yield (kg) in lactations 1, 2 and 3 on selected DIM.

DIM	Lactation 1				Lactation 2				Lactation 3			
	G	PE	E	h^2	G	PE	E	h^2	G	PE	E	h^2
10	9.4	10.6	4.7	0.38	15.9	17.4	9.5	0.37	26.4	26.9	15.8	0.38
50	7.8	8.4	3.8	0.39	13.1	14.2	6.6	0.37	16.1	17.1	8.8	0.38
100	7.3	7.8	3.8	0.39	11.4	12.7	6.6	0.37	13.6	14.9	8.8	0.36
150	6.3	6.8	3.1	0.39	9.7	11.1	4.5	0.38	11.0	12.4	4.5	0.40
200	6.1	6.4	3.1	0.39	10.0	11.6	4.5	0.38	11.3	12.4	4.5	0.40
250	6.7	6.9	3.1	0.40	12.6	14.3	4.5	0.40	14.7	15.1	4.5	0.42
300	8.1	8.1	5.0	0.38	17.5	19.2	8.3	0.39	21.1	20.7	8.0	0.42

Similar values of heritabilities were obtained for other traits. SCS, however, had smaller daily h^2 (0.28-0.35).

Modelling PE effect with random regressions decreased genetic variances, increased environmental variance and gave more uniform heritabilities across DIM in lactations. The total phenotypic variances were comparable with estimates obtained for the model with fixed PE effect (not shown in this paper). Estimates of heritabilities for yield traits on a 305d basis were in the range of 0.5. Genetic correlations between lactations were smaller than most literature estimates from 305d models: 0.69, 0.61 and 0.73 between protein yields in 1st and 2nd, 1st and 3rd, and 2nd and 3rd lactations, respectively. Heritability of milk yield persistency in first lactation was 0.34.

Genetic correlation between persistency and 305d milk yield in first lactation was -0.10. Persistency was expressed as an average additional genetic yield between days 60 and 280 relative to an average cow with the same yield at day 60. However, other expressions of persistency based on estimated lactation curves from random regression model are also possible.

Estimated co(variance) components for all random effects were applied in a multiple trait random regression TD model for genetic evaluation of Holstein bulls and cows. Iterations were performed on an HP-UX 9000/800 workstation. The total number of iterations was 350 with the value of convergence criterion (relative squared difference between animal genetic solutions between two consecutive round) equal to 2.2×10^{-8} . CPU time per

round of iteration was 18min and 650MB of central memory were used to run the iteration program. Estimates of correlations between combined mature equivalent EBVs for 3429 official bulls (at least 12

daughters in 10 herds) from TD model and official February 27 Canadian evaluations are shown in Table 2.

Table 2. Correlations between EBVs from the random regression model and official February 27 evaluations for milk (M), fat (F), protein (P), SCS (S), and persistency (PRM) for milk yield in first lactation (3429 bulls).

	Random regression model				Official			
	F	P	S	PRM	M	F	P	S
M	.63	.89	-.10	.31	.93	.55	.84	-.01
F		.78	-.21	.43	.54	.93	.70	-.08
P			-.11	.35	.81	.69	.94	-.01
S				-.28	-.02	-.12	-.04	.92
PRM					.22	.36	.27	-.17
M						.57	.88	.04
F							.73	-.03
P								.04

Correlations between random regression model and official cow indexes (for 778,616 cows with at least 2 TD records) were 0.90 for overall milk, fat and protein yields. Modelling PE effect by random regressions did not change bull correlations in comparison with fixed PE random regression model. Cow EBVs, however, changed

significantly, from 0.82 to 0.90.

Standard deviations of indexes for bulls (>11 daughters, >9 herds) and cows (>1 TD record) from the random regression model and the official repeatability model were similar for all yield traits (Table 3).

Table 3. Standard deviations of EBVs from random regression model (RR) and Feb. 27 run (Official) for bulls and cows.

	Bulls (3429)		Cows (778,616)	
	RR	Official	RR	Official
Milk (kg)	823.6	811.4	697.2	665.6
Fat (kg)	30.7	29.9	26.6	25.8
Protein (kg)	24.9	23.8	20.6	19.9
SCS	0.20	0.18		
Persistency	170.7			

Canada is planning to implement the multiple trait random regression model in 1998. Several aspects of the model will require further research before final acceptance of TD methodology by the dairy industry. The most important areas are:

- 1) finalization of data edits to ensure completeness of TD data,
- 2) development of a preadjustment procedure to account for within herd heterogeneous variances,
- 3) blending method for including 305d information on older cows,
- 4) procedures for

calculating reliabilities for bull and cow proofs and 5) inclusion in MACE evaluation.

Conclusions

TD models allow milk recording to offer a variety of programs to producers. This is because the use of lactation curves in a multiple trait model can handle long intervals between tests and tests with missing fat or protein components. Estimation of breeding values for persistency is possible with a random regression TD model. The multiple trait random regression TD model with random regressions for the PE effect was applied for the Canadian Holstein population with available computational resources. The random regression model with fixed PE effect seemed to overestimate genetic variances and give more biased cow EBVs than the current model. Significant reranking of bull and cows was found in comparison with the present 305d repeatability animal model.

References

- Henderson, C.R., Jr. 1982. Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. *Biometrics* 38, 623.
- Jamrozik, J., Kistemaker, G.J., Schaeffer, L.R. and Dekkers, J.C.M. 1997. Comparison of possible covariates for use in a random regression model for analyses of test day yields. *J. Dairy Sci.* (in print)
- Jamrozik, J. and Schaeffer, L.R. 1997a. Estimates of genetic parameters for a test day model with random regressions for production of first lactation Holsteins. *J. Dairy Sci.* 80, 762.
- Jamrozik, J., Schaeffer, L.R. and Dekkers, J.C.M. 1997b. Genetic evaluation of dairy cattle using test day yields and a random regression model. *J. Dairy Sci.* 80, 1217.
- Kistemaker, G.J. 1997. *PhD Thesis*. University of Guelph.
- Ptak, E. and Schaeffer, L.R. 1993. Use of test day yields for genetic evaluation of dairy cattle. *Livest. Prod. Sci.* 34, 23.
- Reents, R., Dekkers, J.C.M. and Schaeffer, L.R. 1995a. Genetic evaluation for somatic cell score with a test day model for multiple lactation. *J. Dairy Sci.* 77, 2671.
- Reents, R., Jamrozik, J., Schaeffer, L.R. and Dekkers, J.C.M. 1995b. Estimation of genetic parameters for test day records of somatic cell score. *J. Dairy Sci.* 78, 2847.
- Schaeffer, L.R. and Sullivan, B.P. 1994. Genetic evaluation of dairy goats using test day yields. *Proc. 5th WCGALP*, Guelph, Canada, vol. 18, 182.
- Schaeffer, L.R. and Dekkers, J.C.M. 1994. Random regressions in animal models for test-day production in dairy cattle. *Proc. 5th World Congr. Genet. Appl. Livest. Prod.*, Guelph, vol. XVIII:443.
- 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.

Acknowledgments

Financial support from the Ontario Ministry of Agriculture , Food and Rural Affairs, the Cattle Breeding Research Council of Canada, and the Natural Sciences and Engineering Research Council is gratefully acknowledged.