# Multiple-Trait Random Regression Test-Day Model for all Lactations

M. Lidauer<sup>1</sup>, E. A. Mäntysaari<sup>1</sup>, I. Strandén<sup>1</sup> and J. Pösö<sup>2</sup>

<sup>1</sup>Agricultural Research Centre MTT, Animal Production Research, FIN-31600 Jokioinen, Finland <sup>2</sup>Finnish Animal Breeding Association, FIN-01301 Vantaa, Finland

#### Introduction

Estimated breeding values (EBV) for production traits of Finnish dairy cattle have been calculated using a single trait repeatability animal model since 1990 (Strandén and Mäntysaari, 1992). This has been a basis to a sound genetic progress, which places Finnish Ayrshire as one of the top red breeds worldwide.

During 1990's advances in computer technology and intensive research on test-day (TD) models have allowed Finland to upgrade the three lactations repeatability animal model to a multiple-trait multi-lactation random regression (RR) TD animal model. Introduction of the TD model was motivated by improvements in modeling of the herd environment, by advantages of multiple-trait approach for estimation of breeding values for protein and fat yields, by the desire of breeders to include all lactations and to obtain more detailed breeding values (separate EBV for first and later lactations as well as EBV for persistency).

Objectives of this presentation are to shortly describe the TD model developed, the data used in April 2000, the computational approach, and to compare the new breeding values with those from the previously used single trait repeatability model.

### **Material and Methods**

#### Data

The TD production data described are the same as used for the national breeding value estimation in April 2000. It had TD records of all lactations from cows calving after 1987. A TD record comprised of observations on daily milk, protein and fat yields, when available. The observations had to be recorded between days in milk (DIM) 8 to 365. The data set had 25,568,383 TD records from 1,073,413 cows. The cows' average number of TD observations per lactation was 9.4, 4.5 and 4.5 for milk, protein and fat yield, respectively. The pedigree data comprised of 1,526,943 animals including three breeds (79% Ayrshire, 20% Friesian and 1% Finncattle). Genetic differences between unknown parents and breeds were described by 106 phantom parent groups categorized by breed, timeperiod and selection path.

### Model

Following multiple-trait multi-lactation reduced rank RR TD model was used:

$$\begin{bmatrix} y_{\text{Fijklmnoq}} \\ y_{\text{Lijklmnopq}} \end{bmatrix} = \begin{bmatrix} age_{\text{Fi}} \\ age_{\text{Li}} \end{bmatrix} + \begin{bmatrix} dcc_{\text{Fj}} \\ dcc_{\text{Lj}} \end{bmatrix} + \begin{bmatrix} (ym)_{\text{Fkl}} \\ (ym)_{\text{Lkl}} \end{bmatrix} + \begin{bmatrix} \phi(\text{DIM}) \mathbf{b}_{\text{Fm}} \\ \phi(\text{DIM}) \mathbf{b}_{\text{Lm}} \end{bmatrix} + \begin{bmatrix} (hy)_{\text{Fnk}} \\ (hy)_{\text{Lnk}} \end{bmatrix} + \begin{bmatrix} e_{\text{Fijklmnoq}} \\ e_{\text{Fijklmnoq}} \\ e_{\text{Fijklmnoq}} \end{bmatrix} + \begin{bmatrix} e_{\text{Fijklmnoq}} \\ e_{\text{Fijklmnoq}} \\ e_{\text{Lijklmnopq}} \end{bmatrix} + \begin{bmatrix} e_{\text{Fijklmnoq}} \\ e_{\text{Fijklmnoq}} \end{bmatrix} + \begin{bmatrix} e_{\text{Fijklmnoq}} \\ e_{\text{Fijklmnoq}}$$

where  $y_{Fijklmnoq}$  are the first lactation TD observations of milk, protein, and fat yields, and  $y_{Lijklmnopq}$  are later lactation observations

correspondingly. Thus, number of traits in the phenotypic level is 6. The fixed effects are age at calving (age), days carried calf (dcc), test-

year  $\times$  test-month (ym), regression coefficients or the shape of the lactation curve ( $\phi$ (DIM) **b**), and the herd-year (hy) of the test. The lactation curves were within calving season (3), calving year and lactation (1,2, 3+). The covariables  $\phi(DIM) = [c_1 c_2 c_3 c_4 c_5], \text{ where } c_1, c_2, \text{ and } c_3$ represent a quadratic Legendre polynomial by DIM, and  $c_4$  and  $c_5$  are exponential terms exp( $p_1$ DIM) and exp(- $p_2$ DIM), where  $p_1$  is 0.05 for milk yield and 0.10 for the other traits, and  $p_2$ is 0.06, 0.01, and 0.35 [0.04, 0.20, and 0.35] for milk, protein, and fat yield of first [later] lactation, respectively.

The random effects of herd-test-month (htm) were separate for each of 6 traits similarly to the fixed terms. Observations of first [later] lactations were classified into 9 [16] age classes, 5 [5] dcc classes, 148 [139] ym classes, 36 [72] lactation curve classes and 227,606 [227,606] hy classes. Number of random htm classes was 2,736,535. The RR defining animal breeding values and nongenetic animal effects were defined across traits within lactation. The daily breeding value of cow o in lactation stage DIM in the first

lactation was  $\sum_{r=1}^{6} s(DIM)_{Fr} a_{or}$  where the

covariables in  $s(DIM)_F$  were defined within trait F and will be described in the next section. The later lactation breeding values had the covariables and coefficients of their own. The non-genetic animal effects had a similar structure as the genetic animal effects: functions with 6 terms were defined with covariables specific to trait and DIM for both lactation trait groups. The later lactations had a within lactation specific function  $\sum_{r=1}^{\infty} t(DIM)_{Lr} W_{op(r)}$  that modeled the across

lactation repeatability. The covariables in s(DIM) and t(DIM) were from the eigenfunctions representing the dominant eigenvalues in full fit covariance functions (CF) applied to all traits. Thus the genetic value of an animal is described by a vector of 12 RR coefficients, the non-genetic value of an animal across all lactations and traits is described by a vector of 12 RR coefficients plus 6 RR coefficients for each later lactation.

Let c be a vector of all htm effects; a a vector of all additive genetic animal effects; p

a vector of all animal environmental effects across lactations; w a vector of all animal environmental effects within each of later lactations; and e a vector of all residuals. Covariance matrix of these effects was assumed to be:

$$\operatorname{var}\begin{bmatrix} \mathbf{c} \\ \mathbf{a} \\ \mathbf{p} \\ \mathbf{w} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \otimes \mathbf{C} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \otimes \mathbf{D}_{a} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \otimes \mathbf{D}_{p} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \otimes \mathbf{D}_{w} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix},$$

where C is the (co)variance matrix for htm effects, A is the numerator relationship matrix,  $\mathbf{D}_{a}, \mathbf{D}_{p}, \text{ and } \mathbf{D}_{w}$  are variance matrices from CF, and **R** is the (co)variance matrix for the measurement errors.

# Variance components

The variance-covariance matrices  $\mathbf{D}_{a}$ ,  $\mathbf{D}_{p}$ ,  $\mathbf{D}_{w}$ and **R** were derived using a two step procedure and CF. First, multiple-trait (30×30) variancecovariance matrices for additive genetic ( $\mathcal{G}$ ) and for residual (R) effect were constructed from partial multiple trait REML analyses (see, Mäntysaari, 1999). In each matrix, each of the 6 biological traits were represented by five traits along the course of lactation (DIM intervals 5-20, 31-60, 121-150, 211-240, and 301-330). Next, the method for fitting CF for first lactation production traits (Mäntysaari, 1999) was extended to include later lactations. Rank of the CF were reduced: resulting coefficient matrices had rank of 12 for the genetic effect (D<sub>a</sub>), rank of 12 for the nongenetic animal environmental effect across all lactations and traits  $(\mathbf{D}_{p})$ , and rank of 6 for the non-genetic animal environmental effect within later lactations  $(\mathbf{D}_{w})$ . The corresponding matrices of eigenfunctions S, and T can be partitioned as:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}'_{M_{F}} & \mathbf{S}'_{P_{F}} & \mathbf{S}'_{F_{F}} & \mathbf{S}'_{M_{L}} & \mathbf{S}'_{P_{L}} & \mathbf{S}'_{F_{L}} \end{bmatrix}^{\prime},$$

where e. g.,  $\mathbf{S}_{M_{F}}$  contains the covariables for first lactation milk yield in the form

$$\mathbf{S}_{\mathbf{M}_{\mathrm{F}}} = \begin{bmatrix} \mathbf{s}_{(8)}_{\mathbf{M}_{\mathrm{F}}} & \mathbf{s}_{(9)}_{\mathbf{M}_{\mathrm{F}}} & \mathbf{s}_{(10)}_{\mathbf{M}_{\mathrm{F}}} & \dots & \mathbf{s}_{(365)}_{\mathbf{M}_{\mathrm{F}}} \end{bmatrix}$$

This facilitates the presentation of genetic or phenotypic (co)variances for any DIM and any trait as for example first lactation milk has

 $\sigma_{(i, j)_{M_F}} = \mathbf{s}_{(i)_{M_F}} \mathbf{D}_{aF} \mathbf{s}_{(j)_{M_F}}$ , where  $\mathbf{D}_{aF} = \mathbf{D}_a$ (1:6,1:6). The variances can be used to construct daily heritabilities, and genetic and phenotypic correlations across lactations and traits as given in Table 1.

## **Breeding values**

The TD model gave mixed model equations with 59.3 million unknowns. The equations were solved by preconditioned conjugate gradient method with iteration on data technique as proposed in Strandén and Lidauer (1999) and Lidauer et al. (1999). The solver program, called MiX99 (Lidauer and Strandén, 1999), can exploit parallel processing (Strandén, 1999). Iterations were performed on a Cycle SPARCengine Ultra AXmp with 2 Gb RAM and four processors (300 MHz each). The convergence statistics was the relative difference between left-hand and right-hand side. It had to be smaller than  $3.0 \times 10^{-7}$  in order to have convergence.

Breeding values for first [later] lactation 305-d yields were calculated from the estimated breeding value coefficients in the TD model:  $EBV_{Fi} = \mathbf{1'S}_{F}\hat{\mathbf{a}}(1:6)_{i}$  [EBV<sub>Li</sub> =  $\mathbf{1'S}_{L}\hat{\mathbf{a}}(7:12)_{i}$ ], where  $\mathbf{S}_{F}$  [ $\mathbf{S}_{L}$ ] contain covariables for DIM 8 to 312 and F [L] = milk, protein, and fat yield of first [later] lactation. EBVs were compared with corresponding EBVs obtained from the formerly used single trait repeatability animal model based on 305d-yields of first three lactations.

## **Results and Discussion**

Heritabilities for 305-d lactation yields constructed from daily (co)variances (DIM = 15, 45, ..., 285) were 0.42, 0.28, and 0.29 [0.34, 0.27, and 0.30] for milk, protein, and fat yield of first [later] lactations, respectively. These values were significantly higher, than those presented in Lidauer and Mäntysaari (1999). The difference was due to the reestimation of  $\mathcal{G}$  and  $\mathcal{R}$  using an animal model instead of a sire model. For milk the parameters suggest higher  $h^2$  than that have been used in animal model evaluations (0.30). Note that the genetic correlations between extreme DIMs were intermediate in first lactation but negative, although weak, in second lactation.

Convergence was reached after 458 rounds of iteration. CPU-time per round was 2.6 min. However, real time per round of iteration was 4.4 min. The increase in real time was due to swapping, because the operating system kept the data files in the core memory as well but there was not enough real memory to do this.

Standard deviations of EBVs by different groups of animals are given in Table 2. Test day model EBVs had higher standard deviations than those by the animal model. Increase was highest in EBVs of milk and protein, and not as high in EBVs of fat. As expected the increase in standard deviations of EBVs was higher for cows than for bulls. This reflects that the TD-model has most effect in improved estimation of breeding values for cows.

Correlations between average EBVs from the TD-model and EBVs from the animal model were high for active sires (0.94 to 0.97) but moderate high for cows (0.87 to 0.89). EBVs for first lactation had higher correlation with animal model results than the EBVs for later lactations. The observed correlations were somewhat higher than those reported by Reents et al. (1998) and similar to those by Jamrozik et al. (1997) and Van Doormaal and Kistemaker (1999).

# Conclusions

Evaluations based on reduced rank RR model were found feasible and reasonably robust. For a basic cow with 3 lactations a full rank model would have assigned 24 equations for the breeding values and 48 equations for the nongenetic animal effect. After rank reduction 12 breeding value coefficients and 24 non-genetic animal effects were fitted. Corresponding full RR multiple-trait model as in Jamrozik, et al. (1997) would lead into 27 breeding value equations and 27 non-genetic animal equations. Presumably, the correlation structures in TD models will lead into rather slow convergence in iteration. However, the convergence criteria we applied was quite conservative, although, based on first lactation model comparisons, the rank reduction speeds up the convergence considerably.

Improvements in accuracy of EBVs are not well described by examining the correlations between evaluations of bulls with large number of daughters in all lactations. For them, the 305 days yield based EBVs seem to average close to TD EBVs. In cow breeding value estimation, the TD model leads clear changes in rankings. This is mainly because of its ability to remove environmental biases. By far the largest change for users of sire evaluations comes from separation of breeding values for the first lactation and the later lactation. Correlations to old evaluations suggest that the repeatability model was more aimed to evaluate the first lactation EBVs. Another change due to TD evaluations is the future introduction of breeding values for persistency. In all occasions the Finnish dairy breeders, devoted to animal welfare and to use of total merit indices, have shown high interest on EBVs for the shape of lactation curve.

# References

- Jamrozik, J., Schaeffer, L.R., Liu, Z. & Jansen, G. 1997. Multiple trait random regression test day model for production traits. Vienna, Austria, Aug. 28-29, 1997. *Interbull Bulletin, No. 16*, 43.
- Lidauer, M. & Mäntysaari, E. 1999. Multiple trait reduced rank random regression testday model for production traits. Zurich, Switzerland, August 26-27, 1999. *Interbull Bulletin, No. 22*, 74.

- Lidauer, M. & Strandén, I. 1999. Fast and flexible program for genetic evaluation in dairy cattle. International workshop on high performance computing and new statistical methods in dairy cattle breeding. Tuusula, Finland, March 18-20, 1999. Interbull Bulletin, No. 20, 20.
- Lidauer, M., Strandén, I., Mäntysaari, E.A., Pösö, J. & Kettunen, A. 1999. Solving large test-day models by iteration on data and preconditioned conjugate gradient. *J. Dairy Sci.* 82, 2788.
- Mäntysaari, E.A. 1999. Derivation of multiple trait reduced rank random regression (RR) model for the first lactation test day records of milk, protein and fat. *Proc.* 50<sup>th</sup> Ann. *Mtg. EAAP*, Zurich, Switzerland, Aug. 22-26, 1999.
- 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. Rotorua, New Zealand, Jan. 18-19, 1998. *Interbull Bulletin, No. 17*, 49.
- Strandén, I. 1999. Parallel benefits in test day evaluations. International workshop on high performance computing and new statistical methods in dairy cattle breeding. Tuusula, Finland, March 18-20, 1999. Interbull Bulletin, No. 20, 26.
- Strandén, I. and Lidauer, M. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. *J. Dairy Sci.*, 82:2779.
- Strandén, I. & Mäntysaari, E. A. 1992. Animal model evaluation in Finland: Experience with two algorithms. J. Dairy Sci. 75, 2017.
- Van Doormaal, B.J. & Kistemaker, G.J. 1999. Application of the Canadian Test Day model. Zurich, Switzerland, August 26-27, 1999. *Interbull Bulletin*, No. 22, 100.

·								
		$M_{ m F}$	P <sub>F</sub>	$F_{\rm F}$	M <sub>L</sub>	PL	$F_L$	
	DIM	10 85 160 235 310	10 85 160 235 310	10 85 160 235 310	10 85 160 235 310	10 85 160 235 310	10 85 160 235 310	
$M_{\mathrm{F}}$	10 85 160 235 310	.18.77.64.57.49.43.29.97.88.62.37.64.33.97.76.31.54.62.29.89.24.38.50.61.21	.86.61.48.42.38.60.82.79.70.47.48.78.82.78.59.42.70.78.82.73.38.46.58.74.85	.65.44.35.32.32.43.62.60.53.35.35.60.63.59.42.33.54.59.60.52.33.35.43.53.62	.66.58.45.30.14.49.78.72.61.48.34.76.79.72.63.23.70.81.79.73.08.51.71.75.72	.60.50.34.17.02.37.60.52.40.28.24.57.59.51.41.16.52.62.58.50.08.39.57.58.54	.43.36.25.13.00.24.46.43.35.27.16.45.47.42.36.12.41.48.46.41.10.29.40.41.39	
$\mathbf{P}_{\mathrm{F}}$	10 85 160 235 310	.86.32.28.24.19.38.77.47.39.27.32.48.77.47.39.26.40.47.78.50.20.29.40.50.82	.14.75.63.57.52.39.17.97.87.62.35.56.20.96.74.31.48.56.19.90.25.37.49.59.17	.76.52.42.37.35.52.72.69.61.40.41.69.72.68.50.37.61.69.71.63.35.41.52.64.74	.55.34.21.10.00.39.51.43.35.29.25.50.50.46.43.14.44.52.52.51.03.29.45.50.51	.64.56.41.27.15.42.68.62.51.42.28.66.68.62.56.19.61.72.70.66.11.47.66.69.66	.55.42.28.15.04.34.54.49.41.34.23.53.54.50.46.17.48.56.55.54.12.35.49.51.51	
F <sub>F</sub>	10 85 160 235 310	.75.28.25.22.20.34.66.41.33.22.28.41.63.39.31.21.32.37.64.40.13.18.29.38.68	.73.30.26.21.17.30.69.41.33.22.25.42.67.39.33.19.33.40.68.42.13.21.32.42.72	.22.79.65.56.46.44.17.96.86.61.39.58.20.96.76.33.48.54.18.90.24.32.44.54.15	.35.24.19.13.07.24.39.37.32.27.16.39.41.38.34.10.35.42.40.37.05.23.34.35.32	.53.38.25.13.03.33.48.41.32.24.20.47.47.41.35.12.43.51.48.42.06.32.47.47.42	.69.58.41.24.09.47.73.67.55.44.31.72.75.68.60.20.67.78.75.69.09.52.70.71.68	
$M_L$	10 85 160 235 310	.33.17.13.08.03.20.34.32.27.18.17.33.35.34.28.12.28.34.36.33.07.20.29.34.34	.32.15.11.07.03.14.25.23.19.13.10.22.23.23.20.07.17.22.25.25.03.12.20.25.27	.28.13.10.05.01.10.20.17.13.06.08.17.18.16.12.07.13.16.18.17.04.08.14.18.19	.11.49.31.11.16.30.16.85.70.60.23.47.22.97.88.15.38.47.27.96.05.28.40.46.24	.90.46.28.07.17.31.78.62.47.37.21.60.72.67.57.06.47.70.72.6615.43.67.73.74	.56.25.19.07.11.15.55.43.34.29.13.42.51.50.44.07.34.49.51.4804.31.44.47.49	
$P_L$	10 85 160 235 310	.34.16.12.09.06.18.27.24.20.13.14.24.26.26.22.09.19.25.28.27.04.13.21.27.30	.36.16.13.09.06.17.26.24.20.14.14.24.26.26.23.10.20.25.29.29.06.15.24.29.32	.31.13.10.06.02.13.22.19.15.09.11.19.20.19.17.08.15.19.22.22.04.10.17.22.25	.67.21.16.09.00.25.61.30.23.17.18.31.59.31.27.10.24.31.60.34.00.17.27.33.62	.16.47.32.12.10.33.12.85.70.61.27.43.17.97.89.19.37.46.21.97.10.31.43.51.22	.84.35.23.06.15.35.74.60.49.44.20.59.71.71.69.05.46.67.72.7411.39.58.66.73	
F <sub>L</sub>	10 85 160 235 310	.31.13.11.09.08.15.22.20.15.09.11.19.20.19.15.06.14.18.20.1800.07.14.18.20	.33.14.12.10.08.14.24.21.17.11.11.20.22.21.18.06.15.20.22.22.02.10.17.22.24	.37.16.12.08.05.18.28.26.21.14.15.26.27.26.22.11.21.26.28.27.06.14.22.28.30	.69.22.19.15.10.30.60.32.24.17.23.31.53.28.22.14.21.26.50.26.03.11.19.24.50	.67.22.17.11.04.22.61.30.24.18.16.29.59.30.27.08.22.30.60.3302.15.26.33.63	.25.51.29.08.14.29.13.86.70.61.21.44.18.96.89.12.35.43.22.97.02.25.37.44.21	

Table 1. Heritabilities (on diagonal), genetic (above diagonal) and phenotypic (below diagonal) correlations from reduced rank covariance function for first lactation milk (M<sub>F</sub>), protein (P<sub>F</sub>) and fat (F<sub>F</sub>), and later lactations milk (M<sub>L</sub>), protein (P<sub>L</sub>) and fat (F<sub>L</sub>) test-day yields by a sample of days in milk (DIM)

Table 2. Standard deviations (kg) of estimated breeding values for milk (M), protein (P), and fat (F) yields for first lactation (FIRST), later lactations (LATER), average of first and later lactations (0.5(FIRST+LATER)) obtained from the multiple-trait multi-lactation random regression testday animal model (TD-MODEL), and obtained from the previously used single trait repeatability animal model (ST-R-AM) by different groups of animals; Ayrshire (AY) and Friesian (FR) bulls born 1991-1993 with at least 60 daughters, and Ay and Fr cows born in 1995 with at least 4 test-day observations. Number of animals in parenthesis

	TD-MODEL							ST-R-AM				
	First			]	LATER			.5(First+Later)				
	М	Р	F	М	Р	F	М	Р	F	М	Р	F
AY bulls (335)	415.	10.4	16.4	496.	14.5	21.2	438.	11.9	17.9	414.	11.0	17.4
FR bulls (132)	426.	11.0	18.9	477.	13.3	22.4	434.	11.6	20.1	407.	11.1	19.4
AY cows(67,252)	401.	9.3	14.1	436.	11.5	17.5	406.	10.0	15.3	334.	9.6	15.0
FR cows (20,804)	449.	10.7	14.5	491.	12.9	18.4	457.	11.4	16.0	370.	10.7	15.2

Table 3. Correlations between estimated breeding values for milk (M), protein (P), and fat (F) yield for first lactation (FIRST), later lactation (LATER), first and later lactations average (0.5(FIRST+LATER)) obtained from the multiple-trait multi-lactation random regression testday animal model, and corresponding breeding values obtained from the previously used single trait repeatability animal model by different groups of animals; Ayrshire (AY) and Friesian (FR) bulls born 1991-1993 with at least 60 daughters, and Ayrshire and Friesian cows born in 1995 with at least 4 test-day observations. Number of animals in parenthesis

	First				LATER			.5(FIRST+LATER)			
	М	Р	F	Μ	Р	F	М	Р	F		
AY bulls (335)	0.96	0.93	0.95	0.90	0.89	0.90	0.97	0.96	0.97		
FR bulls (132)	0.95	0.94	0.97	0.89	0.87	0.93	0.96	0.94	0.97		
AY cows (67,252)	0.86	0.85	0.88	0.83	0.82	0.85	0.87	0.87	0.89		
FR cows (20,804)	0.88	0.87	0.88	0.84	0.84	0.85	0.88	0.88	0.89		