

# Multiple Trait Reduced Rank Random Regression Test-Day Model for Production Traits

*M. Lidauer and E. A. Mäntysaari*

*Agricultural Research Centre MTT, Animal Production Research,  
FIN-31600 Jokioinen, Finland*

## Introduction

At the present the random regression (RR) test-day (TD) model is referred to be the most precise model for estimation of production traits in dairy cattle. It accounts for environmental variation within lactation and allows better description of effects due to pregnancy and maturity. Genetic variance among different TD yields can be modeled and allows estimation of breeding values for persistency. Multiple trait analysis increases accuracy of estimated breeding values (EBV) of less frequently measured traits. Hence, a multiple trait RR TD model should reduce biases in EBV's and increase accuracy. However, implementation of RR models into national dairy cattle evaluation systems is a challenge due to computational limitations and difficulties in estimation of (co)variance components.

New developments for iteration on data technique (Strandén and Lidauer, 1999) overcome some computational constraints in solving large RR models. For the estimation of (co)variance components different methods have been proposed, each dictating a different way of modeling the random effects. Jamrozik et al (1997) introduced a model with 12 traits and three RR coefficients per trait for both non-genetic, and genetic animal effect. Wiggans and Goddard (1998) considered each test month of three biological traits and two lactations as different trait, yielding a multiple trait model with 60 traits. To make it computationally feasible, they reduced the model by canonical transformation to six most significant contrasts. Introduction of covariance functions (CF) (Kirkpatrick et al., 1990) opened new possibilities in (co)variance component estimation for RR models (e.g., Van der Werf et al., 1998; Tijani et al., 1999).

Mäntysaari (1999) fitted CF to multiple trait (co)variance components of first parity milk, protein and fat yield, where each biological trait was represented by TD yields from five different lactation periods. Describing sufficiently the (co)variance structure of genetic and non-genetic animal effect required five regression coefficients for each effect.

Aim of this study was to estimate breeding values from first parity TD records of all Finnish dairy cows using the RR (co)variance parameters as derived by Mäntysaari (1999). Breeding values for 305-d lactation yields and for persistency were derived from the estimated breeding value coefficients and compared with official proofs.

## Material and Methods

### *Data*

The data set consisted of 9,329,505 first parity TD records of 986,540 cows from the national dairy cattle population. TD records were from Ayrshire, Friesian, and Finncattle cows that calved between January 1988 and April 1999. Records from day 5 to 365 of cows which had at least an observation on milk yield, were included. The pedigree comprised of 1,414,887 cows and 10,059 bulls of all three breeds and genetic differences between unknown parents were described by 106 phantom parent groups.

### *Model*

The following multiple trait RR TD model based on reduced rank CF was applied:

$$\begin{bmatrix} y_{1ijklmnop} \\ y_{2ijklmnop} \\ y_{3ijklmnop} \end{bmatrix} = \begin{bmatrix} \text{age}_{1i} \\ \text{age}_{2i} \\ \text{age}_{3i} \end{bmatrix} + \begin{bmatrix} \text{dcc}_{1j} \\ \text{dcc}_{2j} \\ \text{dcc}_{3j} \end{bmatrix} + \begin{bmatrix} \text{yrm}_{1kl} \\ \text{yrm}_{2kl} \\ \text{yrm}_{3kl} \end{bmatrix} + \begin{bmatrix} \sum_{d=1}^4 \phi_t^d b_{1m}^d \\ \sum_{d=1}^4 \phi_t^d b_{2m}^d \\ \sum_{d=1}^4 \phi_t^d b_{3m}^d \end{bmatrix} + \begin{bmatrix} (\text{herd} * \text{yr})_{1nk} \\ (\text{herd} * \text{yr})_{2nk} \\ (\text{herd} * \text{yr})_{3nk} \end{bmatrix} + \begin{bmatrix} (\text{herd} * \text{yrm})_{1nkl} \\ (\text{herd} * \text{yrm})_{2nkl} \\ (\text{herd} * \text{yrm})_{3nkl} \end{bmatrix} + \\
\sum_{d=1}^5 \begin{bmatrix} \mathbf{q}_{1t}^d \\ \mathbf{q}_{2t}^d \\ \mathbf{q}_{3t}^d \end{bmatrix} [\mathbf{a}_o^d] + \sum_{d=1}^5 \begin{bmatrix} \mathbf{t}_{1t}^d \\ \mathbf{t}_{2t}^d \\ \mathbf{t}_{3t}^d \end{bmatrix} [\mathbf{p}_o^d] + \begin{bmatrix} e_{1ijklmnop} \\ e_{2ijklmnop} \\ e_{3ijklmnop} \end{bmatrix},$$

where  $y_{Zijklmnop}$  is the TD observation  $p$  for trait  $Z = 1 = \text{milk yield}$ ,  $Z = 2 = \text{protein yield}$ , and  $Z = 3 = \text{fat yield}$ , made in herd  $n$ , in the year  $k$ , of the month  $l$ , on a cow  $o$ , which belongs to the age class  $i$  and to the days carried calf class  $j$  and produced on the TD  $t$ ;  $\text{age}_{Zi}$ , is age at calving effect with 9 classes;  $\text{dcc}_{Zj}$  is days carried calf effect with 5 classes;  $\text{yrm}_{Zkl}$ , is test-year\*test-month effect with 136 classes;  $b_{Zm}^1$ ,  $b_{Zm}^2$ ,  $b_{Zm}^3$ , and  $b_{Zm}^4$  are fixed regression coefficients describing the shape of the seasonal lactation curve  $m$  for each trait  $Z$ , where calving season  $m$  was October to February, March to June, or July to September;  $\phi_t = [1 \ x_t \ x_t^2 \ w_t]'$ , with  $x_t$  and  $x_t^2$  being the first and second order Legendre polynomial, and  $w_t$  being  $\exp(-0.05t)$  for TD  $t$ , as given by Wilmlink (1987);  $\text{herd} * \text{yr}_{Znk}$  is the herd\*test-year effect with 236,412 classes. Random effects are:  $\text{herd} * \text{yrm}_{Znkl}$  is the herd\*test-year-test-month effect;  $\mathbf{a}_o^d$  is a vector of five RR coefficients describing the genetic effect of animal  $o$ ;  $\mathbf{p}_o^d$  is a vector of five RR coefficients of non-hereditary animal effects modeling the environmental covariances among measurements along lactation and across traits of animal  $o$ ; and  $e_{Zijklmnop}$  is the measurement error.  $\mathbf{q}_{Zt}^d$  and  $\mathbf{t}_{Zt}^d$  contain covariables associated with  $\mathbf{a}_o^d$  and  $\mathbf{p}_o^d$ , respectively, where covariables for a particular trait  $Z$  and TD  $t$  are obtained from eigen functions as explained below.

### Reduced rank covariance functions

Additive genetic ( $\mathbf{G}$ ) and residual ( $\mathbf{R}$ ) variance-covariance matrix, with each biological trait

represented by five separate traits along the course of lactation (period 5-20, 31-60, 121-150, 211-240, and 301-330 days in milk), were constructed from 130 partial multiple trait REML analyses (Pösö et al., 1999; in preparation).  $\mathbf{G}$  and  $\mathbf{R}$  were used to fit CF as explained in Mäntysaari (1999):

$$\mathbf{G}_{15 \times 15} = \Phi \mathbf{K}_a \Phi', \text{ and}$$

$$\mathbf{R}_{15 \times 15} = \Phi \mathbf{K}_p \Phi' + \mathbf{R}_{i3 \times 3} \otimes \mathbf{I}_{5 \times 5}.$$

Rank of CF was reduced by using the five largest eigenvalues of  $\mathbf{K}_a$  and  $\mathbf{K}_p$  and replacing  $\Phi$  by eigen functions:

$$\mathbf{G}_{15 \times 15} \approx \Phi \mathbf{V}_a \mathbf{D}_a \mathbf{V}_a' \Phi', \text{ and}$$

$$\mathbf{R}_{15 \times 15} \approx \Phi \mathbf{V}_p \mathbf{W} \mathbf{V}_p' \Phi' + \mathbf{R}_{i3 \times 3} \otimes \mathbf{I}_{5 \times 5},$$

where  $\mathbf{D}_a$  and  $\mathbf{W}$  are matrices with five largest eigenvalues of  $\mathbf{K}_a$  and  $\mathbf{K}_p$ , respectively, and  $\mathbf{V}_a$  and  $\mathbf{V}_p$  contain the corresponding eigen vectors;

$$\Phi_{15 \times 12} = \mathbf{I}_{3 \times 3} \otimes \Lambda_{5 \times 4}, \text{ with}$$

$\Lambda_{5 \times 4} = [\phi_1, \dots, \phi_5]'$  and  $\phi_t$  is the same as given above. The diagonality of  $\mathbf{W}$ , however was lost because after rank reduction the  $\mathbf{W}$  and  $\mathbf{R}_i$  were refitted (Mäntysaari 1999). Extending  $\Lambda$  to the first 365 days in milk ( $\Lambda_{365 \times 4} = [\phi_1, \dots, \phi_{365}]'$ ) allows direct approximation of additive genetic and residual multiple trait variance-covariance matrix for each TD as a trait:

$$\mathbf{G}_{1095 \times 1095} \approx \mathbf{\Phi}_{1095 \times 12} \mathbf{V}_{a12 \times 5} \mathbf{D}_a \mathbf{V}'_{a5 \times 12} \mathbf{\Phi}'_{12 \times 1095} = \mathbf{QD}_a \mathbf{Q}', \text{ and}$$

$$\mathbf{R}_{1095 \times 1095} \approx \mathbf{\Phi}_{1095 \times 12} \mathbf{V}_{p12 \times 5} \mathbf{WV}'_{p5 \times 12} \mathbf{\Phi}'_{12 \times 1095} + \mathbf{R}_{i3 \times 3} \otimes \mathbf{I}_{365 \times 365} = \mathbf{TWT}' + \mathbf{R}_i \otimes \mathbf{I}.$$

The covariable matrices  $\mathbf{Q}$  and  $\mathbf{T}$  can be partitioned by traits, e.g.,

$$\mathbf{Q} = [\mathbf{Q}'_1 \quad \mathbf{Q}'_2 \quad \mathbf{Q}'_3]',$$

where e. g.,  $\mathbf{Q}_1$  contains the covariables for milk yield in the form

$$\mathbf{Q}_1 = [\mathbf{q}_{1,1} \quad \mathbf{q}_{1,2} \quad \mathbf{q}_{1,3} \quad \dots \quad \mathbf{q}_{1,365}]'.$$

Heritabilities, genetic and phenotypic correlations for a sample of TD's, calculated from the reduced rank CF, are given in Table 1.

Table 1. Heritabilities (on diagonal), genetic (above diagonal) and phenotypic (blow diagonal) correlations for milk, protein, and fat yields by different days in milk (DIM). Values given in 1/100

	DIM	Milk							Protein							Fat						
		5	55	105	155	205	255	305	5	55	105	155	205	255	305	5	55	105	155	205	255	305
Milk	5	<b>14</b>	78	68	63	61	59	53	93	75	69	67	65	61	54	84	69	46	35	30	28	27
	55	47	<b>16</b>	98	96	92	81	60	79	77	75	74	72	66	56	69	59	45	37	34	31	29
	105	48	62	<b>20</b>	99	96	85	65	72	73	73	74	73	69	60	62	54	45	40	38	38	36
	155	46	59	64	<b>22</b>	96	90	72	71	70	72	74	76	74	67	59	51	46	43	43	44	45
	205	40	52	59	62	<b>20</b>	96	82	72	67	70	74	78	80	77	57	48	45	45	47	52	57
	255	29	42	49	55	59	<b>18</b>	95	72	60	65	71	79	87	89	52	41	41	45	51	61	71
	305	16	30	37	44	51	58	<b>16</b>	66	47	52	61	72	86	95	43	29	33	39	49	64	80
Protein	5	84	53	52	48	41	30	18	<b>10</b>	79	79	81	83	81	74	95	85	73	66	63	61	58
	55	44	77	47	46	42	34	24	40	<b>10</b>	99	95	89	78	61	76	70	52	42	37	35	33
	105	46	50	82	53	49	41	31	45	51	<b>13</b>	99	95	84	68	78	74	61	53	50	48	45
	155	45	51	55	85	54	48	39	47	48	54	<b>14</b>	98	91	77	79	76	67	62	60	60	58
	205	39	48	53	55	85	53	47	42	43	50	55	<b>13</b>	97	87	78	73	69	67	68	71	71
	255	28	40	45	50	54	87	55	31	35	42	49	54	<b>12</b>	96	71	64	65	66	71	78	84
	305	15	29	35	42	49	56	89	17	25	32	40	48	55	<b>11</b>	59	49	54	59	67	79	90
Fat	5	71	47	48	46	40	29	16	70	41	46	46	41	30	16	<b>16</b>	97	86	77	71	63	49
	55	38	63	42	39	34	26	17	40	66	44	42	37	29	19	26	<b>7</b>	94	86	79	68	49
	105	38	43	65	43	39	32	23	42	43	68	45	41	34	26	33	51	<b>9</b>	98	94	85	66
	155	37	42	45	66	43	38	31	40	40	45	70	45	40	33	38	43	47	<b>12</b>	99	92	76
	205	31	37	42	44	67	43	39	34	35	41	45	70	45	40	36	35	41	45	<b>12</b>	97	85
	255	20	30	35	40	44	70	47	24	27	34	40	44	72	48	25	26	32	39	43	<b>9</b>	95
	305	9	21	27	33	40	47	73	12	19	26	32	40	47	75	10	18	24	30	37	44	<b>6</b>

The model included 57 effects and yielded in mixed model equations with 20,180,163 unknowns. TD observations were scaled for all three traits to have equal phenotypic standard

deviations. The equations were solved by preconditioned conjugate gradient method, using iteration on data technique as proposed by Strandén and Lidauer (1999). The

preconditioner matrix comprised of diagonal blocks of the coefficient matrix, analogous with the structure as explained in Lidauer et al. (1999). The data and the LDL' decomposition of the preconditioner were read once per iteration round. A general BLUP software (Lidauer and Strandén, 1999) suitable for parallel computing (Strandén, 1999) was applied. Iterations were performed on a Cycle SPARCEngine Ultra AXmp with 2 Gb RAM, using four processors (300 MHz each). The user system (Solaris 5.2) avoided I/O operations by keeping the data and preconditioner files in the available core memory. The convergence criteria was defined as relative difference between left-hand and right-hand side weighted by the preconditioner matrix being smaller than  $10^{-12}$ .

Breeding values for 305-d yields were calculated from estimated breeding value coefficients:  $EBV_{Zi} = \mathbf{1}'\mathbf{Q}_Z\hat{\mathbf{a}}_i$ , where  $\mathbf{Q}_Z$  contains covariables for the first 305 TD's, and  $Z$  = milk, protein, fat. Two different breeding values for persistency of milk yield were derived. The first was as given by Jamrozik et al. (1997) and was breeding value for TD 280 minus breeding value for TD 60;  $P_{280-60i} = (\mathbf{q}_{280} - \mathbf{q}_{60})\hat{\mathbf{a}}_i$ . The second was the standard deviation of the breeding values for TD 1 to 305 ( $P_{SD}$ ) as suggested by Sölkner and Fuchs (1987). Breeding values were compared with official proofs from July 1999. For official evaluations a single trait repeatability model, which includes first three lactations was used.

## Results and Discussion

Convergence of solutions was reached after 215 rounds of iteration. The more commonly used convergence indicator, relative squared difference between solutions of two consecutive rounds, was  $2.4 \times 10^{-10}$ . CPU-time per round of iteration was 45 seconds and total wall clock time needed for pre-processing and solving was 4.0 hours.

Inspection of estimated breeding value curves confirmed that the model is capable for accounting different shaped lactation curves (Figure 3 - 6). Concavity of breeding value curves increased over time (Figure 3). As expected, standard deviations of breeding values for particular TD's were higher for TD's with higher heritabilities. Correlations between EBV's of 491 active bulls (born in 1990-92, with at least 60 daughters) from the RR model and from official evaluations were 0.95, 0.93, and 0.95 for milk, protein, and fat yield, respectively. Standard deviations of EBV's from the RR model were larger for milk but same and smaller for protein and fat yield compared with those from the official run (Table 3). From a RR model we would expect higher standard deviations for EBV's. However, for the official proofs observations on second and third lactation were included as well. More important might be the lower heritabilities for the RR model. If converted to 305-d yields these were 0.30, 0.23, and 0.22 for milk, protein, and fat, respectively, whereas for official evaluation used heritability was 0.30 for all three traits. Low heritabilities originated from multiple trait parameters (Mäntysaari 1999). One reason might be that a sire model was used for estimation of variance components, but it can be speculated whether heritabilities for RR TD models are conceptually lower, as indicated also in other studies (e.g., Gengler et al., 1999).

Persistency breeding value  $P_{280-60}$  was uncorrelated with milk yield for bulls but negatively correlated for cows. Persistency breeding value  $P_{SD}$  was positively (unfavorable) correlated with milk yield for both bulls and cows (Table 2). Both persistency measurements yielded significantly different ranking of animals because  $P_{280-60}$  accounts for the slope only but not for the concavity of curve, and vice versa for  $P_{SD}$ . Because persistency of phenotypic lactation curve is what we are interested in, it should be investigated how persistency breeding values relate to the shape of the phenotypic curve.

Table 2. Correlations between breeding value coefficients and breeding values for milk yield (M), protein yield (P), fat yield (F), persistency for milk yield as EBV for day 280 minus EBV for day 60 ( $P_{280-60}$ ) and as standard deviation of EBV's for all first 305 days ( $P_{SD}$ ), obtained from a multiple trait random regression (RR) test-day model. Upper triangle values from bulls born 1990-1992, with at least 60 daughters; lower triangle values from cows born 1996 with at least one test-day record

	Estimated breeding values					RR breeding value coefficients				
	M	P	F	$P_{280-60}$	$P_{SD}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$
M		.80	.59	-.01	.46	.86	.23	-.25	.32	.02
P	.83		.74	-.12	.38	.88	-.27	-.10	.10	.11
F	.34	.60		-.05	.31	.89	-.09	.59	-.14	-.23
$P_{280-60}$	-.39	-.38	-.06		.00	-.05	.65	.06	-.35	.80
$P_{SD}$	.65	.57	.37	-.29		.39	.17	-.21	-.39	-.01
$\hat{a}_1$	.82	.88	.82	-.28	.58		.03	.23	.13	-.13
$\hat{a}_2$	.06	-.42	-.21	.51	.05	-.13		.02	.00	.20
$\hat{a}_3$	-.57	-.36	.49	.36	-.39	-.05	.05		-.18	-.37
$\hat{a}_4$	.28	-.01	-.46	-.35	-.35	-.07	.09	-.42		-.28
$\hat{a}_5$	-.11	.06	-.24	.68	-.17	-.20	-.02	-.22	-.18	

Table 3. Standard deviations (kg) of EBV's for milk, protein, fat, and persistency of milk as EBV for day 280 minus EBV for day 60 ( $P_{280-60}$ ) and as standard deviation of EBV's for first 305 days ( $P_{SD}$ ), from multiple trait random regression (RR) test-day model and official evaluations July 1999 (Official) by bulls (born 1990-1992 with at least 60 daughters) and cows (born in 1996). Number of animals in parenthesis

	Ayrshire bulls (535)		Friesian bulls (129)		Ayrshire cows (57 669)	
	RR	Official	RR	Official	RR	Official
Milk <sub>305d</sub>	406.8	390.3	398.0	362.0	301.2	278.0
Protein <sub>305d</sub>	10.1	10.1	10.3	9.6	7.1	8.0
Fat <sub>305d</sub>	15.3	17.0	17.4	17.9	10.4	12.7
$P_{280-60}$	1.00		0.96		0.61	
$P_{SD}$	0.25		0.24		0.20	

## Conclusions

Multiple trait reduced rank random regression test-day model was applied to first lactation test-day yields of the Finnish dairy cattle population. For each animal three different breeding value curves were sufficiently described by five breeding value coefficients. Standard deviations of 305-d yield EBV's from the random regression model were not so high as expected. This was most likely due to the lower heritabilities used for the random regression model. The used solving algorithm

yielded in considerable savings in computing time.

## References

- Gengler, N., Tijani, A., Wiggans, G.R., Van Tassel, C.P. & Philpot, J.C. 1999. Estimation of (co)variances of test day yields for first lactation Holsteins in the United States. J. Dairy Sci. 82, 225.
- Jamrozik, J. Schaeffer, L.R., Liu, Z. & Jansen, G. 1997. Multiple trait random regression test day model for production traits. Proc.

- INTERBULL Annu. Mtg., Vienna, Austria, Aug. 28-29, 1997. INTERBULL Bulletin. 16, 43-47.
- Kirkpartick, M., Lofsvold, D. & Bulmer, M. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics*. 124, 979-993.
- 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. In press.
- 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.* Accepted.
- 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.
- Sölkner, J. & Fuchs, W. 1987. A comparison of different measures of persistency with special respect to variation of test-day milk yields. *Livest. Prod. Sci.* 16, 305-319.
- 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. In press.
- Strandén, I. & Lidauer, M. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. *J. Dairy Sci.* Accepted.
- Tijani, A., Wiggans, G.R., Van Tassel, C.P., Philpot, J.C. & Gengler, N. 1999. Use of (co)variance functions to describe (co)variance for test day yield. *J. Dairy Sci.* 82, 226.
- 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 evaluations of yield traits in the United States. *J. Dairy Sci.* 80, 1795-1800.
- Wilmink, J.B.M. 1987. Adjustment on test-day milk, fat and protein yields for age, season and stage of lactation. *Livest. Prod. Sci.* 16, 335-348.

Fig. 1: Fixed lactation curves by calving season

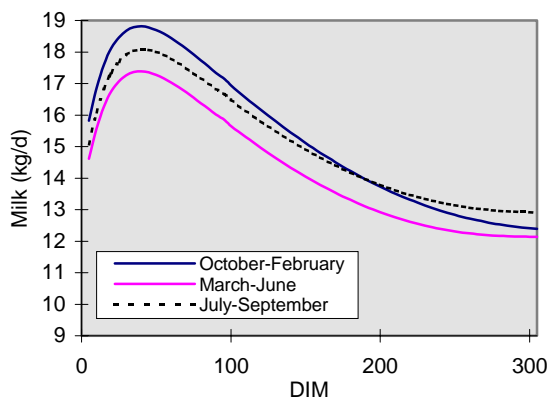


Fig. 2: Genetic groups for dams of cows born in 1990 by breed

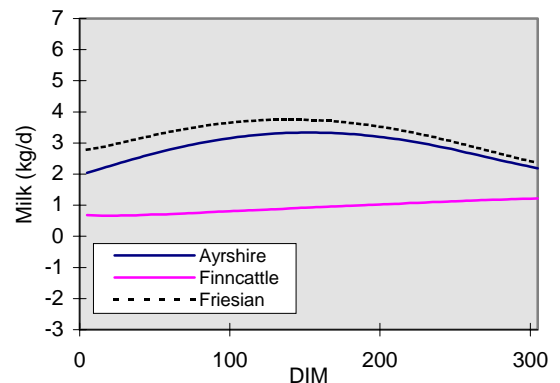


Fig. 3: Average breeding value curves of cows by year of birth

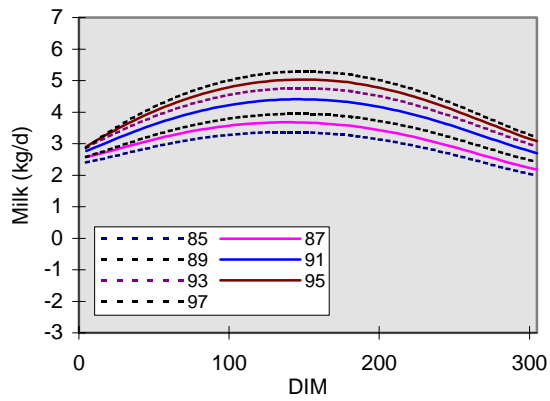


Fig. 4: Breeding value curves of top bulls (>20 daughters)

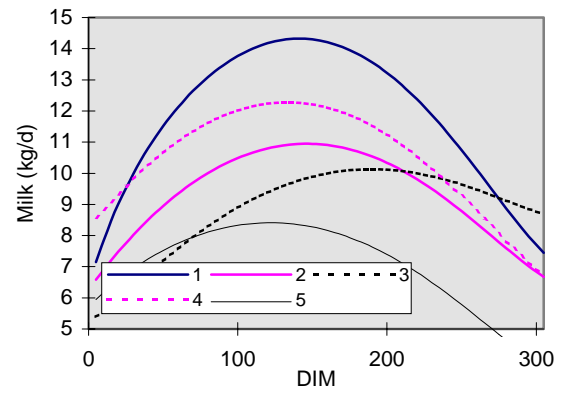


Fig. 5: Breeding value curves of top bulls (>20 daughters)

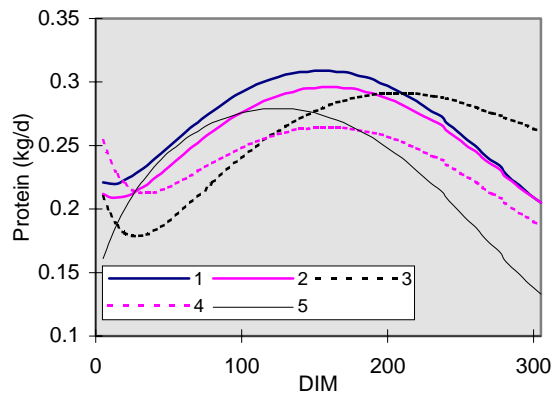


Fig. 6: Breeding value curves of top bulls (>20 daughters)

