

# Univariate and multivariate parameter estimates for milk production traits using an animal model. I. Description and results of REML analyses

PM Visscher<sup>1,\*</sup>, R Thompson<sup>2</sup>

<sup>1</sup> *University of Edinburgh, Institute of Cell, Animal and Population Biology,  
West Mains Road, Edinburgh EH9 3JT;*

<sup>2</sup> *AFRC Institute of Animal Physiology and Genetics Research,  
Roslin EH25 9PS, Midlothian, UK*

(Received 5 August 1991; accepted 3 June 1992)

**Summary** – Using production records in lactations 1–3 from 100 large Holstein–Friesian pedigree herds, parameters for milk, fat and protein yield in lactations 1–3 were estimated with REML using an animal model. The number of records for each lactation was 38 811, 26 223 and 16 542 for lactation 1, 2 and 3 respectively. Heritabilities for the 3 yield traits were similar: approximately 0.36 in lactation 1 and 0.30 in lactations 2 and 3. Genetic correlations between yield traits in lactations 1 and 2, for example between milk production in first and second lactations, were approximately 0.86. Genetic correlations between yield traits in lactations 2 and 3 were near unity. Genetic correlations between yield traits within lactations ranged from 0.58, for milk and fat yield in lactation 3, to 0.91, for milk and protein yield in lactation 1. Genetic correlations between yield traits between lactations ranged from 0.55, for milk yield in lactation 1 and fat yield in lactation 2, to 0.85, for milk yield in lactation 2 and protein yield in lactation 3. Environmental correlations between traits within lactations were approximately 0.95, and approximately 0.40 across lactations.

**dairy cattle / animal model / maximum likelihood / multivariate analyse / multi-trait / multi-lactation**

**Résumé** – Utilisation du modèle animal pour l'estimation des paramètres univariates et multivariates concernant les caractères de production laitière. I. Description et résultats des analyses selon le maximum de vraisemblance restreint (REML). A partir des données obtenues pendant les lactations 1 à 3 dans 100 grands troupeaux Holstein–Friesian inscrits, les paramètres de production de lait, de matière grasse et de matière protéique pour les lactations 1 à 3 ont été estimés par maximum de vraisemblance restreint

\* Present address: Livestock Improvement Unit, Department of Food and Agriculture, 166-176 Wellington Parade, East Melbourne, Melbourne 3002, Australia

*(REML) selon le modèle animal. Le nombre des données pour chaque lactation a été de 38 811, 26 223 et 16 542 pour les lactations 1, 2 et 3 respectivement. Les héritabilités des 3 critères de production ont été similaires: approximativement 0,36 en première lactation et 0,30 en seconde et troisième lactation. Les corrélations génétiques entre les caractères de production en lactations 1 et 2, par exemple la production laitière, ont été approximativement de 0,86. Les corrélations génétiques entre critères de production aux lactations 2 et 3 ont été pratiquement égaux à 1. Les corrélations génétiques entre les critères de production intralactation ont varié de 0,58, pour la production laitière et la production de matière grasse en lactation 3, à 0,91 pour la production laitière et la production de protéine en lactation 1. Les corrélations entre les productions à différentes lactations ont varié de 0,55, pour la production laitière en lactation 1 et la production de matière grasse en lactation 2, à 0,85 pour la production laitière en lactation 2 et la production de protéine en lactation 3. Les corrélations non génétiques entre caractères pour une même lactation ont été approximativement de 0,95 et celles correspondant à des lactations différentes ont été approximativement de 0,40.*

**bovin laitier / modèle animal / maximum de vraisemblance / analyse multivariée / multicaractère / multilactation**

## INTRODUCTION

Dairy cattle sire evaluation in many countries is carried out using best linear unbiased prediction (BLUP) (Interbull, 1988), while cows are usually evaluated separately using a selection index type approach (eg Hill and Swanson, 1983). Recently there has been a shift towards a joint evaluation of cows and bulls, using a so-called animal model (AM). Some countries have implemented an AM national evaluation for single traits (Wiggans *et al*, 1988a, b; Ducrocq *et al*, 1990; Jones and Goddard, 1990), others are in the process of doing so. Assumptions about the covariance structure of observations analysed with a linear model are often simplified to make computations feasible. For example, the USA (Wiggans *et al*, 1988a), France (Ducrocq *et al*, 1990; Bonaiti and Boichard, 1990) and Australia (Jones and Goddard, 1990) use a modified repeatability model for which a genetic correlation of unity is assumed between performances across lactations and some (pre)scaling is applied to later lactation records to account for higher phenotypic variances of traits in later lactations. Later lactation records are given lower weightings by adjusting the error structure of the observations, and milk, fat, and protein yield are analysed separately using this modified repeatability model. The potential loss in efficiency of selection by making these assumptions depends on the true, unknown, covariance structure of the data, and on the breeding goal. By estimating relevant (co)variances and assuming a particular combination of traits to select for, the potential loss in efficiency of selection by using simplified covariance structures may be quantified.

For estimating (co)variance components it seems desirable to use the same model as is, or soon will be, used for the prediction of breeding values, *ie* an animal model. Few (co)variance estimates from AM analyses have been reported; Swalve and Van Vleck (1987) analysed milk yield in lactations 1–3, and Van Vleck and Dong (1988) performed a multivariate analysis on milk, fat and protein yield in the first lactation.

The aims of this study were: 1) to estimate parameters for milk (M), fat (F) and protein (P) yield in lactations 1, 2 and 3 ( $L_1$ ,  $L_2$ ,  $L_3$ ); 2) to investigate the implications of the estimates for prediction of breeding values when simplified assumptions are made regarding covariances structures. This part of the study is reported separately (Visscher *et al*, 1992).

Estimates of correlations between different traits in different lactations, for example between milk yield in lactation 1 ( $M_1$ ) and fat yield in lactation 2 ( $F_2$ ) have not been reported before. In the notation used, the number following M, F or P refers to lactation number, and the combination above,  $M_1$  and  $F_2$ , may be written as  $M_1F_2$ . Similarly, a multivariate (MV) analysis on  $M_1$ ,  $F_1$  and  $P_1$  may be written as analysing  $M_1F_1P_1$ .

## MATERIAL

First, second and third lactation production records for the period 1979–1987 from 100 large pedigree herds were extracted from the Milk Marketing Board's production files. Herds were selected on the number of heifers present in 1987, *ie* data were extracted from those herds which had the largest number of first lactation cows in 1987. Later lactation records, *ie* second or third, were included only from cows for which the previous lactations were present. All cows were pedigree Holstein–Friesian (HF). Some summary statistics of the data are presented in table I.

**Table I.** Summary statistics of data.

		<i>Lactation</i>		
		<i>1</i>	<i>2</i>	<i>3</i>
No of records		38 811	26 223	16 542
No of animal effects		58 689	42 835	28 919
No of sires		2 357	1 948	1 565
Mean (kg)	M	5 291	6 143	6 643
	F	208.8	239.7	257.8
	P	173.0	201.4	215.5
SD <sup>1</sup> (kg)	M	1 111	1 335	1 372
	F	44.6	53.0	55.7
	P	34.6	41.5	42.7

<sup>1</sup> Standard deviation.

## METHODS

Residual maximum likelihood (REML; Patterson and Thompson, 1971) was used to estimate (co)variances, using programs based on software written by Meyer (1988, 1989). Fixed effects in the mixed linear model were herd–year–seasons (HYS) and month of calving. Seasons were defined as 4-month periods, corresponding to the definition used for the current UK sire evaluations. Proportion of HF ancestry

in the cow, age at calving and lactation length were fitted as covariables. All animal effects, including those of proven sires, were treated as random; this may cause a (downward) bias in the estimates, since comparisons between proven sires contribute to the estimate of genetic variance. For these data, the possible bias in parameter estimates from treating all sire effects as random was investigated elsewhere (Visscher and Thompson, 1992).

The following analyses were carried out:

1) univariate analyses for each of M, F and P in lactations 1–3. If culling takes place on performance in previous lactations, the parameter estimates from univariate analyses on later lactations will be biased. Comparing variance components from these univariate analyses with components from models that (partly) take account of selection may give some indication about what kind of selection (if any) has acted on these data;

2) analyses using a repeatability model for each of M, F and P in lactations 1 and 2. For this model it was assumed that the genetic correlation of performance between lactations was unity and that variances were constant across lactations. A permanent environmental effect was fitted as an additional random effect for these analyses. Comparing results from these analyses with results from bivariate analyses may show how the (co)variances are partitioned when a genetic correlation of unity between performances in lactations 1 and 2 is implicitly assumed;

3) within lactation (for  $L_1$ ,  $L_2$  and  $L_3$ ) MV analyses for traits M, F and P. An algorithm proposed by Thompson and Hill (1990) was used to estimate (co)variances. Their algorithm was designed to reduce a multivariate estimation problem to a set of independent univariate estimations. Assuming equal design matrices for  $p$  traits, Thompson and Hill (1990) proposed performing  $q = p(p+1)/2$  univariate analyses, where the  $q$  “traits” are obtained from linear transformations of the  $p$  traits, and suggested finding a transformation matrix (iteratively) that would stabilise the backtransformed  $p \times p$  covariance matrices from one round to the next. Following Thompson and Hill’s suggestion, the initial transformation matrix was chosen so that  $p = 3$  traits and 3 sums of traits were analysed. Subsequently, after  $q = 6$  univariate analyses, a canonical transformation was calculated and 3 canonical variates were formed. The next “round” consisted of performing univariate analyses on these 3 canonical variates and on 3 pairwise sums of the canonical variates. The whole procedure was stopped after 5 complete rounds of iteration, since correlations on the original scale changed very little from round 4 to 5. Thompson and Hill (1990) proposed their algorithm for the case of equal design matrices and more than 2 random effects in the linear model. For the analyses described above, only 2 random effects (animal and residual) were fitted, so that a “standard” canonical transformation (see *eg* Meyer, 1985) could have been applied. Both methods, however, should give similar estimates, since the described algorithm was found to be highly efficient (Thompson and Hill, 1990);

4) bivariate (BV) analyses on all pairwise combinations of traits in different lactations. Unfortunately, analysing the data using a general MV model (for example with 3 traits in 3 lactations, *ie* for 9 traits) was computationally not feasible. Therefore, selection bias is likely to affect some of the parameter estimates. In particular, (co)variances estimated for lactation 2 and 3 will be biased if culling

was based on performance in the first lactation. For all BV analyses the fixed effect structure was different for both traits.

For most analyses data sets were too large to be handled in one single likelihood evaluation. Data sets were therefore randomly subdivided into subsets of herd groups. The estimates from each sample of herds were assumed to be independent of other estimates. This assumption is not strictly true, since some sires had progeny in different subsets. The correlation between estimates from different samples depends on the number of sires represented in different samples and their contribution to the parameter estimates in each sample. For analyses 1) and 2) data were split into 5 subsets of 20 herds each, for analyses 3) into 5, 4 and 2 subsets (for  $L_1$ ,  $L_2$ , and  $L_3$  respectively), and for 4) into 10 groups of 10 herds. For the univariate analyses and the analyses using a repeatability model, the standard errors (SE) of the estimates were calculated by approximating the likelihood surface at the maximum likelihood estimates by a quadratic function in the parameters of interest and using the matrix of second differentials to calculate asymptotic variances of the estimates (see Visscher *et al.*, 1991, for an application and discussion of this procedure). For the within lactation MV analyses and the BV analyses, the average (co)variance estimates are presented with the empirical standard error of the mean estimate. No weighting of estimates was applied because subsets were roughly of equal size and there was insufficient information about the sampling (co)variances of the variance components (a weighting according to the number of records in the analysis was tried and showed differences between weighted and unweighted means of the order of 1% of the mean).

It was not clear how to combine the different estimates efficiently into one overall ( $9 \times 9$ ) covariance matrix, since there was insufficient information about sampling variances and culling bias. Estimates of variances and covariances of M, F and P in lactation 3, for example, were available from bivariate analyses  $L_1L_3$  and  $L_2L_3$  and from MV analyses within  $L_3$ , all of which were probably subject to culling bias. The following method was chosen to create  $9 \times 9$  genetic, environmental and phenotypic covariance matrices which were consistent with each other: for  $L_1$  the (co)variances from analyses 3) were used. The variances (diagonals) in  $L_2$  and  $L_3$  were taken from BV analyses  $L_1L_2$  and  $L_1L_3$  using the same trait in each lactation. For example, the variance estimate for  $P_3$  was used from analysis  $P_1P_3$ . Within lactation genetic and environmental covariances between M, F and P for lactations 2 and 3 were calculated using the variances as described above and the estimates of the within lactation genetic and environmental correlations. The phenotypic covariances were calculated as the sum of the genetic and environmental covariances thus created and phenotypic correlations were calculated from these. The same method was used to calculate covariances between different traits in different lactations, now using the genetic and environmental correlations estimated from BV analyses. This somewhat arbitrary way of combining different estimates was found to give fewest problems of negative definite covariance matrices. It was thus assumed that variances from BV analyses  $L_1L_2$  and  $L_1L_3$ , and genetic and environmental correlations between traits within lactations, were least biased by selection.

To summarise the calculation of the  $9 \times 9$  covariance matrices:

- All genetic, environmental and phenotypic (co)variances within lactation one were from multivariate analyses on  $M_1F_1P_1$ ;
- Environmental and genetic correlations between milk, fat and protein yield within lactations 2 and 3 were from multivariate analyses on  $M_2F_2P_2$  and  $M_3F_3P_3$  respectively;
- Environmental, genetic and phenotypic variances for  $M_2, F_2, P_2, M_3, F_3,$  and  $P_3$  were calculated from bivariate analyses on  $M_1M_2, F_1F_2, P_1P_2, M_1M_3, F_1F_3$  and  $P_1P_3$  respectively;
- Environmental and genetic correlations between traits between lactations were taken from bivariate analyses for each pairwise comparison;
- All remaining phenotypic covariances and phenotypic correlations followed directly from combining the above calculated elements.

Parameters for fat and protein content were approximated using a first order Taylor series expansion. If  $x_i/y_i$  and  $x_j/y_j$  are ratio traits in lactations  $i$  and  $j$  respectively, then an approximation of the covariance between those 2 traits is:

$$\text{cov}(x_i/y_j, x_j/y_j) \approx \frac{\mu_{x_i} \mu_{x_j}}{\mu_{y_i} \mu_{y_j}} \left( \text{CV}(x_i) \text{CV}(x_j) r_{x_i, x_j} - \text{CV}(x_i) \text{CV}(y_j) r_{x_i, y_j} - \text{CV}(y_i) \text{CV}(x_j) r_{y_i, x_j} + \text{CV}(y_i) \text{CV}(y_j) r_{y_i, y_j} \right) \quad [1]$$

with CV the coefficient of variation ( $= \sigma/\mu$ ) and  $r_{x,y}$  the correlation between traits  $x$  and  $y$ . Formula [1] was applied using estimates of the coefficients of variation and estimates of the (co)variances for the yield traits in lactations 1-3.

## RESULTS

The main results of the different analyses are presented in tables II-X. Heritabilities for production traits for the first lactations (table II) were moderate to high. Although the genetic parameter estimates from the univariate analysis for lactation 2 may be biased by selection, increase in the environmental variance for lactation 2 (which is unlikely to be greatly affected by culling) was striking; the ratio of environmental variances in lactation 2 to that in 1 was approximately 1.6. Part of the increase in variances for the second lactation may be a scale effect associated with a larger mean (see also tables I and XIII for means and coefficients of variation), since the (biased) genetic variance for lactation 2 is also larger than the first lactation genetic variance. Results for lactation 3 also showed an increase in environmental and phenotypic variance. The estimate of the genetic variance in lactation 3 was smaller than estimates for either lactation 1 or 2, most likely explained by ignoring the effect of culling from lactation 1 to 2 and from lactation 2 to 3.

Results from analyses with a repeatability model are presented in table III. Heritabilities were slightly lower than those estimated from univariate analyses on first lactations only. The variance component estimates from the analyses using a

**Table II.** Univariate REML estimates for lactations 1-3 (variances in kg<sup>2</sup>).

	$M_1$	$F_1$	$P_1$	$M_2$	$F_2$	$P_2$	$M_3$	$F_3$	$P_3$
$\hat{\sigma}_a^2$	238 564	330.7	193.2	246 425	349.9	218.9	207 491	300.3	199.0
$\hat{\sigma}_e^2$	371 956	584.5	351.0	608 266	950.2	571.3	693 720	1 121.6	647.1
$\hat{\sigma}_p^2$	610 520	915.2	544.2	854 691	1300.1	790.2	901 211	1 421.9	846.1
$h^2$	0.39	0.36	0.36	0.29	0.27	0.28	0.23	0.21	0.24
SE( $h^2$ )	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.03

repeatability model (table III) may be explained using the general bivariate model results from tables V to VII; it seems that both the genetic and phenotypic variances from the repeatability model were roughly the average of the bivariate first and second lactation parameters, and the average environmental variance in lactation 1 and 2 was partitioned into a permanent environmental and residual variance. If selection were on first lactation performance, a repeatability model should account for this selection effect, conditional on a genetic correlation of unity between first and second lactation performance.

**Table III.** Univariate REML estimates from first and second lactations using a repeatability model.

<i>Variance components</i>	<i>M</i>	<i>F</i>	<i>P</i>
$\hat{\sigma}_a^2$	255 855	355.1	214.5
$\hat{\sigma}_{pe}^2$ (permanent environment)	149 167	224.9	152.6
$\hat{\sigma}_e^2$	313 138	509.6	295.8
$\hat{\sigma}_p^2$	718 160	1 089.6	662.9
$h^2$	0.36	0.33	0.32
SE ( $h^2$ )	0.01	0.01	0.01
Repeatability	0.56	0.53	0.55
SE (repeatability)	0.01	0.01	0.01

Table IV shows the heritability and correlation estimates from the within lactation MV analyses. Heritabilities were similar to univariate (unitrait) estimates from table II, as expected, and again heritability estimates from  $L_2$  and  $L_3$  are expected to be biased downwards. Phenotypic correlations between yield traits were very similar for different lactations, and genetic correlations were slightly lower in  $L_2$  in comparison with  $L_1$ , but similar for  $L_2$  and  $L_3$ . Genetic and phenotypic correlations between milk and protein yield were very high, and environmental correlations for these traits calculated using the estimates from table IV were close to unity.

**Table IV.** Within lactation correlation matrices ( $\times 100$ ) from within lactation MV analyses on M, F, and P<sup>1,2,3</sup>.

	$M_1$	$F_1$	$P_1$		$M_2$	$F_2$	$P_2$		$M_3$	$F_3$	$P_3$
$M_1$	39 (1)	75 (2)	91 (1)	$M_2$	28 (2)	62 (6)	88 (1)	$M_3$	24 (1)	58 (1)	86 (1)
$F_1$	84 (1)	36 (1)	82 (1)	$F_2$	82 (1)	25 (1)	75 (4)	$F_3$	82 (1)	21 (1)	73 (1)
$P_1$	95 (1)	87 (1)	36 (1)	$P_2$	94 (1)	87 (1)	26 (3)	$P_3$	94 (1)	86 (1)	24 (2)

<sup>1</sup> For each  $3 \times 3$  matrix: heritabilities ( $\times 100$ ) on diagonals, genetic correlations above and phenotypic correlations below diagonals. <sup>2</sup> Empirical standard errors ( $\times 100$ ) below each estimate. <sup>3</sup> Mean and empirical SE of parameter estimates were based on 5, 4 and 2 samples for  $L_1$ ,  $L_2$   $L_3$  respectively.

In tables V to VII the combined  $9 \times 9$  covariance matrices are presented. The similarity between the various  $3 \times 3$  lactation by lactation covariance blocks is striking. In a subsequent study the consequences of these results for prediction of breeding values are investigated further. From table V it seems that genetically  $L_2$  and  $L_3$  are essentially the same for the yield traits, with genetic correlations between performances in second and third lactations in excess of 0.97. Comparing pairs of covariances or correlations such as  $M_1F_2$  and  $F_1M_2$  shows that their values are similar, which indicates that the ratio of variances for traits in different lactations are similar for M, F, and P. Similar proportionalities seem to exist for

**Table V.** Additive genetic covariance matrix (upper triangle) and genetic correlations ( $\times 100$ ; below diagonals) for M, F and P in lactations 1-3<sup>1,2,3,4</sup>.

	$M_1$	$F_1$	$P_1$	$M_2$	$F_2$	$P_2$	$M_3$	$F_3$	$P_3$
$M_1$	241 594	6 699	6 276	227 270	5 893	6 019	217 083	5 523	5 808
$F_1$	75	329.3	208.3	5 931	316.8	198.4	5 695	314.9	195.9
$P_1$	91	82	196.0	5 871	185.1	188.7	5 736	182.1	189.4
$M_2$	87	62	79	282 462	6 651	7 289	272 594	7 174	7 381
$F_2$	59	86	65	62	410.1	236.0	7 384	401.8	256.9
$P_2$	78	70	86	88	75	244.0	7 464	254.8	251.1
$M_3$	84	60	78	98	70	91	274 479	6 164	7 439
$F_3$	55	85	64	66	97	80	58	415.8	243.3
$P_3$	72	66	82	85	77	98	86	73	270.1

<sup>1</sup> Parameters for  $L_1$  are from within first lactation MV analyses (consistent with first block from table IV). <sup>2</sup> Within lactation off-diagonals for  $L_2$  and  $L_3$  are calculated using variance components from BV analyses (diagonals) and correlations from within lactation MV analyses. <sup>3</sup> All other estimates are averages from BV analyses on 10 samples. <sup>4</sup> Range empirical SE ( $\times 100$ ) of correlations:  $L_1L_2$ : 2-4;  $L_1L_3$ : 2-6;  $L_2L_3$ : 1-5.



environmental components (table VI). Environmental correlations between traits within lactations were similar for lactations 1-3. Phenotypic correlations between traits within lactations (table VII) are not necessarily the same as those from table IV, because of the way this table was constructed. Little change, however, is observed. Phenotypic correlations for  $M_1M_2$ ,  $F_1F_2$  and  $P_1P_2$  were slightly higher than repeatability estimates from table III. Again the proportionality of the various  $3 \times 3$  covariance blocks is striking.

**Table VI.** Environmental covariance matrix (upper triangle) and environmental correlations ( $\times 100$ ) below diagonals) for M, F and P in lactations 1-3<sup>1</sup>.

	$M_1$	$F_1$	$P_1$	$M_2$	$F_2$	$P_2$	$M_3$	$F_3$	$P_3$
$M_1$	373 134 13 185	11 098	193 774	6 446	6 053	188 310	5 932	5 379	
$F_1$	89	589.5	409.6	6 589	298.8	228.0	6 198	274.1	199.6
$P_1$	97	90	353.7	5 635	207.6	193.5	5 300	186.9	168.4
$M_2$	42	36	39	584 290	20 685	17 481	263 259	8 532	7 866
$F_2$	35	41	37	90	914.2	648.6	8 471	391.1	285.4
$P_2$	42	40	44	97	91	559.3	7 860	286.1	259.5
$M_3$	38	32	35	43	35	41	644 499	23 051	18 846
$F_3$	30	35	31	35	40	38	89	1 040.8	710.3
$P_3$	36	34	37	42	39	45	97	91	591.8

<sup>1</sup> Range SE ( $\times 100$ ) of correlations:  $L_1L_2$ : 1-2;  $L_1L_3$ : 2-2;  $L_2L_3$ : 1-2.

**Table VII.** Phenotypic covariance matrix (upper triangle) and phenotypic correlations ( $\times 100$ ; below diagonals) for M, F and P in lactations 1-3.

	$M_1$	$F_1$	$P_1$	$M_2$	$F_2$	$P_2$	$M_3$	$F_3$	$P_3$
$M_1$	614 728 19 883	17 373	421 044	12 339	12 072	405 393	11 454	11 187	
$F_1$	84	918.8	617.9	12 520	615.6	426.4	11 893	589.0	395.6
$P_1$	95	87	549.7	11 506	392.7	382.2	11 035	369.0	357.7
$M_2$	58	44	53	866 752	27 337	24 770	535 853	15 707	15 247
$F_2$	43	56	46	81	1 324.3	884.5	15 853	792.9	542.3
$P_2$	54	50	58	94	86	803.3	15 324	540.9	510.5
$M_3$	54	41	49	60	45	56	918 978	29 215	26 286
$F_3$	38	51	41	44	57	50	80	1 456.6	953.6
$P_3$	49	44	52	56	51	61	93	85	861.9

In table VIII, heritability estimates for the 9 "traits" are given which are expected to be least biased through selection, with coefficients of variation for genetic, environmental and phenotypic effects. As before, lactations 2 and 3 seem very similar. For all yield traits the additive genetic CV slightly decreased from  $L_1$  to  $L_2$ , and the environmental CV increased from  $L_1$  to  $L_2$ . Scale effects therefore act differently for genetic and environmental effects, and there seems to be no single

**Table VIII.** Heritabilities ( $\times 100$ ), their empirical standard errors ( $\times 100$ ) and coefficients of variation (CV; in %) from bivariate analyses<sup>1,2</sup>.

	$M_1$	$F_1$	$P_1$	$M_2$	$F_2$	$P_2$	$M_3$	$F_3$	$P_3$
$h^2$	40	37	36	33	31	30	30	29	31
SE ( $h^2$ )	2	1	1	2	2	2	2	2	2
CV <sub>a</sub>	9.2	8.7	8.0	8.7	8.4	7.8	7.9	7.9	7.6
CV <sub>e</sub>	11.5	11.6	10.8	12.4	12.6	11.7	12.1	12.5	11.3
CV <sub>p</sub>	14.8	14.5	13.5	15.2	15.2	14.1	14.4	14.8	13.6

<sup>1</sup> First and second lactation estimates are from  $M_1M_2$ ,  $F_1F_2$  and  $P_1P_2$ . Third lactation estimates are from  $M_1M_3$ ,  $F_1F_3$  and  $P_1P_2$ . <sup>2</sup> CVs are  $\sigma_a/\bar{x}$ ,  $\sigma_e/\bar{x}$  and  $\sigma_p/\bar{x}$  respectively, using the means from table I.

scale transformation which would standardise both genetic and residual variances across lactations.

Many analyses that were carried out yielded different estimates for the same variance component. For example, an estimate for  $M_1$  was available from a univariate analysis, from a MV analysis with  $F_1$  and  $P_1$ , and from 6 different BV analyses. All those different estimates for the same component are shown in table IX. For each row the 2 identical values were from within lactation MV analyses, since, for example,  $M_1$ ,  $F_1$  and  $P_1$  were analysed multivariately but pairwise combinations  $M_1F_1$ ,  $M_1P_1$  and  $F_1P_1$  were not analysed bivariately. Diagonals in table IX were from univariate analyses (see table II). As expected, the various estimates for first lactation variances are very similar, since these estimates are free from selection bias. Ignoring first lactation information to estimate variances in later lactations reduces the additive genetic variances by approximately 10%, most likely due to culling bias. It is not clear why the highest estimate for any trait in  $L_2$  was from a combined analysis with the same trait in  $L_1$ , *ie*  $M_1M_2$  gave the highest estimate for  $M_2$ , and  $F_1F_2$  and  $P_1P_2$  showed the highest estimates for  $F_2$  and  $P_2$  respectively. Using prediction equations for selection biases from Meyer and Thompson (1984), no selection strategy for first lactation production traits was found that would produce these results.

A summary of the parameters calculated for fat and protein content (F% and P% respectively), from using equation (1), is presented in tables X and XI. Heritabilities for F% and P% were high and were fairly constant across lactations. Genetic correlations for  $F_2\%F_3\%$  and  $P_2\%P_3\%$  were substantially lower than the genetic correlations between yield traits in second and third lactations. Parameters for first lactation traits ( $M_1$ ,  $F_1$ ,  $P_1$ ,  $F_1\%$  and  $P_1\%$ ) were similar to estimates from a  $5 \times 5$  MV analysis on all traits in lactation one (results not presented). Genetic correlations between protein yield and protein percentage were negative in first and positive in later lactations, although small in all cases.

**Table IX.** Comparison of variance components estimates (kg<sup>2</sup>) from different pairwise analyses.

Trait 2 :	M <sub>1</sub>	F <sub>1</sub>	P <sub>1</sub>	M <sub>2</sub>	F <sub>2</sub>	P <sub>2</sub>	M <sub>3</sub>	F <sub>3</sub>	P <sub>3</sub>
<i>Genetic components for trait 1, from combined analysis with trait 2</i>									
<i>Trait 1</i>									
M <sub>1</sub>	238 564	241 594	241 594	242 903	239 838	240 284	239 325	235 846	237 094
F <sub>1</sub>	329.3	330.7	329.3	331.1	334.5	331.5	325.1	326.3	325.2
P <sub>1</sub>	196.0	196.0	193.2	195.1	194.7	196.0	194.4	192.5	192.8
M <sub>2</sub>	282 462	266 699	271 265	246 425	238 205	238 205	250 612	241 764	248 120
F <sub>2</sub>	390.1	410.1	392.8	322.7	349.9	322.7	360.3	354.4	361.8
P <sub>2</sub>	242.3	238.5	244.0	207.6	207.6	218.9	221.7	215.3	224.4
M <sub>3</sub>	274 479	244 164	260 351	255 960	235 948	246 660	207 491	216 977	216 977
F <sub>3</sub>	381.9	415.8	387.9	333.9	374.5	344.5	300.0	300.3	300.0
P <sub>3</sub>	262.2	248.3	270.1	241.6	235.7	249.2	204.2	204.2	199.0
<i>Environmental components for trait 1, from analysis with trait 2</i>									
M <sub>1</sub>	371 956	373 134	373 143	367 511	369 236	369 113	368 209	370 676	369 857
F <sub>1</sub>	589.4	584.5	589.4	582.7	579.4	582.8	585.7	584.8	585.0
P <sub>1</sub>	353.7	353.7	351.0	349.7	350.1	349.4	349.1	349.9	350.4
M <sub>2</sub>	584 291	588 395	588 537	608 266	603 973	603 973	587 910	594 193	589 094
F <sub>2</sub>	916.2	914.2	917.3	961.3	950.2	961.3	919.6	925.0	918.4
P <sub>2</sub>	558.3	559.8	559.2	574.5	574.5	571.3	556.8	561.3	554.6
M <sub>3</sub>	644 499	659 592	652 339	649 053	658 737	654 627	693 720	673 182	673 182
F <sub>3</sub>	1 044.7	1 040.7	1 044.0	1 074.8	1 054.9	1 072.1	1 105.5	1 121.6	1 105.5
P <sub>3</sub>	594.7	605.4	591.8	603.7	606.7	601.4	633.4	633.4	647.1

**Table X.** Parameters for fat content (F%) and protein content (P%) in lactations 1–3<sup>1</sup>.

	F <sub>1</sub> %	P <sub>1</sub> %	F <sub>2</sub> %	P <sub>2</sub> %	F <sub>3</sub> %	P <sub>3</sub> %
F <sub>1</sub> %	58	60	85	54	84	52
P <sub>1</sub> %	37	62	45	78	44	76
F <sub>2</sub> %	55	34	64	64	74	48
P <sub>2</sub> %	33	56	40	63	49	78
F <sub>3</sub> %	52	34	62	38	61	62
P <sub>3</sub> %	35	57	39	60	42	62

<sup>1</sup> Heritabilities ( $\times 100$ ) on diagonals, genetic correlations ( $\times 100$ ) above diagonals and environmental correlations ( $\times 100$ ) below diagonals.

**Table XI.** Genetic and environmental correlations ( $\times 100$ ) between yield and content traits within lactations 1–3.

Trait combination	L <sub>1</sub>		L <sub>2</sub>		L <sub>3</sub>	
	$r_g$	$r_e$	$r_g$	$r_e$	$r_g$	$r_e$
F%M	-42	-22	-47	-20	-46	-17
F%F	28	25	40	26	46	30
F%P	-20	-13	-18	-10	-15	-6
P%M	-50	-36	-44	-34	-33	-37
P%F	-9	-18	11	-15	24	-17
P%P	-10	-10	4	-9	19	-12

## DISCUSSION

Univariate first lactation heritabilities were similar to the most recent UK estimates using a sire model (Meyer, 1987), but higher than estimates of Hill *et al* (1983) and Meyer (1983 and 1984). Heritability estimates from pedigree populations are often higher than from non-pedigree populations (Meyer, 1987; Carabaño *et al*, 1990). In dairy cattle, heritability estimates from daughter–dam regression are usually higher than estimates from paternal half-sib comparisons (Maijala and Hanna, 1974; Van Vleck 1986), and since the AM–REML estimates are a combination of both, this may “explain” why the AM estimates are higher than previous estimates from sire models. Swalve and Van Vleck (1987) found AM–REML heritability estimates of approximately 0.33 for milk yield in the first 3 lactations, using a trivariate model and ignoring relationships between animals across herds. Information contributing to their heritability estimates was therefore mainly from daughter–dam comparisons. Van Vleck and Dong (1988) reported AM heritability estimates of 0.36, 0.35 and 0.33 for milk, fat and protein yield in first lactations. The increase of the phenotypic variance over time, additional to an increase associated with a higher mean production, is striking; a regression of the coefficients of variation (CV) of milk production in the UK on time, using literature estimates from Hill *et al* (1983), Meyer (1984, 1987) and Visscher *et al* (1991),

shows a slight increase in the phenotypic CV from 1976–1987 and an increase in the genetic CV from 7 to 9%. The explanation for this observation is not clear, although perhaps better estimation procedures, in particular those accounting for selection on the data, may account for some increase in the estimate of the genetic variance in addition to a scale effect.

Genetic and phenotypic correlations between  $M_1$ ,  $F_1$  and  $P_1$  were slightly higher than the correlations found by Van Vleck and Dong (1988). Genetic correlations between  $M_1$ ,  $M_2$  and  $M_3$  were almost identical to the estimates of Swalve and Van Vleck (1987) and slightly lower than the sire model estimates of Meyer (1987). A small negative genetic correlation between protein yield and protein content in lactation 1 was also reported by Swanson and Gnanasakthy (1991). Genetic correlations between protein percentage and yield traits indicate that response to selection for fat and protein yield can be achieved without a reduction in the level of protein percentage, which accords with the wishes of many European dairy breeders. The explanation for the substantially lower genetic correlation between content traits in lactation 2 and 3, *ie* for  $F_2\%F_3\%$  and  $P_2\%P_3\%$ , compared with near unity correlations for the yield traits is not clear. Applying equation [1] to  $F_2\%$  and  $F_3\%$ , assuming all CVs are equal and genetic correlations for  $F_2F_3$  and  $M_2M_3$  are unity, gives:

$$r_{F_2\%F_3\%} = \left[ 1 - 1/2(r_{F_2M_3} + r_{M_2F_3}) \right] / \left[ (1 - r_{M_2F_2})(1 - r_{M_3F_3}) \right]^{1/2}$$

Therefore one explanation may be that the within lactation correlations, calculated from within lactation MV analyses were biased downwards relatively more than the between lactation between trait correlations which were calculated from BV analyses.

If culling of first lactation cows were on some linear combination of their milk, fat and protein production in the first lactation or on any "culling variate" correlated with the traits being analysed, this form of selection would only partially be accounted for when using a bivariate REML estimation (see Robertston, 1966, for a detailed theoretical framework of a culling process). Therefore the BV second lactation parameter estimates may be slightly biased. The 3 traits considered were highly correlated, however, and the ratio of bivariate to univariate variance components was similar for all traits, which suggests that the bias may be small. Meyer and Thompson (1984) presented prediction equations of selection biases for a one-way sire classification, when culling is on a trait correlated with yield in the first lactation and maximum likelihood is used to estimate the parameters. Using their prediction formulas the selection bias was investigated for various combinations of genetic and environmental correlations between the culling variate and the traits in the BV analyses. Selection intensity was calculated from the relative number of cows that had second lactations. It was found that for a range of parameter values likely to correspond with the true population values for milk, fat and protein yield, small biases were predicted for the estimates of the genetic parameters, but substantial biases (up to 40% of the true values) could occur for the environmental correlations between the 2 traits in the analyses. For example, if the culling variate was fat yield in lactation 1, the percentage biases in the estimate of the heritability for the trait in lactation 2 and for the genetic and environmental correlation would be 0,

0.4 and  $-4.4$  respectively for  $M_1M_2$ , and 0.2, 0.3 and 4.4 for  $P_1P_2$ , using the BV parameter estimates as true population values. Although most of the information used in AM-REML is a combination of comparisons between (paternal) half-sibs and daughter-dam pairs, the effect of selection on a correlated trait is unlikely to be large for the range of parameters investigated.

The parameter estimates from the bivariate model clearly showed that production traits in the second lactation are not repeated observations of first lactation records. Still, most countries use a repeatability model in their national AM evaluation, albeit with a lower weighting given to second and later lactation records. The weighting of later lactations seems the only instrument within the present day national AM evaluations to approximate the more appropriate multivariate model, for which heritabilities are lower and variances are much higher in later lactations. Additional to the implicit assumption of a genetic correlation of unity between first and later lactation yields, an improper weighting of later lactations when using a repeatability model will reduce genetic progress. Some calculations thereof are given in a subsequent study.

As described previously, the method used to create  $9 \times 9$  covariance matrices from various available estimates was somewhat arbitrary. Any combination of estimates is expected to give sampling problems, since the traits are so highly correlated. For example, using heritability estimates from table VIII with genetic and phenotypic correlations from table IV gives 3 within lactation environmental covariance matrices which all are negative definite. Using estimates of environmental correlations between  $M_1$ ,  $F_1$  and  $P_1$  from Majjala and Hanna (1974), Meyer (1985) and Van Vleck and Dong (1988), determinants of the environmental correlation matrix were found to be  $-0.003$ ,  $0.012$  and  $0.03$  respectively, indicating that sampling problems may be expected with these traits. Still, when using the method described to calculate full  $9 \times 9$  covariance matrices, sampling problems were not eliminated: the  $9 \times 9$  genetic covariance matrix presented in table V is not positive definite. However, the only negative eigenvalue in this matrix was relatively close to zero ( $-0.04$  after standardising all phenotypic variances to unity for  $M_1$ ,  $F_1$  and  $P_1$ ). Setting this eigenvalue to a small positive number (eg  $10^{-6}$ ) and recalculating all matrices showed very little difference for all variance components.

## ACKNOWLEDGMENTS

PM Visscher acknowledges financial support from the Milk Marketing Boards in the UK. We thank WG Hill for many helpful comments.

## REFERENCES

- Bonaiti B, Boichard D (1990) Benefits from animal model evaluation of dairy cattle in France. *In: Proc 4th World Congr Genet Appl Livest Prod, Edinburgh, Scotland, July 23-27*, 13, 364-373
- Carabaño MJ, Wade KM, Van Vleck LD (1990) Genotype by environment interaction for milk and fat production across regions of the United States. *J Dairy Sci* 73, 173-180

- Ducrocq V, Boichard D, Bonaiti B, Barbat A, Briend M (1990) A pseudo-absorption strategy for solving animal model equations for large data files. *J Dairy Sci* 73, 1945-1955
- Hill WG, Swanson GJT (1983) A selection index for dairy cows. *Anim Prod* 37, 313-319
- Hill WG, Edwards MR, Ahmed MKA, Thompson R (1983) Heritability of milk yield and composition at different levels and variability of production. *Anim Prod* 36, 59-68
- Interbull (1988) Sire evaluation procedures for dairy production traits practised in various countries. *Bull No 3*, Uppsala, Sweden
- Jones LP, Goddard ME (1990) Five years experience with the animal model for dairy cattle evaluations in Australia. In: *Proc 4th World Congr Genet Appl Livest Prod, Edinburgh, Scotland, July 23-27*, 13, 382-385
- Maijala K, Hanna M (1974) Reliable phenotypic and genetic parameters in dairy cattle. In: *Proc 1st World Congr Genetics Applied to Livestock Production, Madrid*, vol 1, 541-563
- Meyer K (1983) Scope for evaluating dairy sires using first and second lactation records. *Livest Prod Sci* 10, 531-553
- Meyer K (1984) Estimates of genetic parameters for milk and fat yield for the first three lactations in British Friesian cows. *Anim Prod* 38, 313-322
- Meyer K (1985) Maximum Likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* 41, 153-165
- Meyer K (1987) Estimates of variance due to sire-herd interaction and environmental covariances between paternal half-sibs for first lactation dairy production. *Livest Prod Sci* 17, 95-115
- Meyer K (1988) DFREML - A set of programs to estimate variance components under an individual animal model. *J Dairy Sci* (suppl 2) 71, 33
- Meyer K (1989) Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free logarithm. *Genet Sel Evol* 21, 317-340
- Meyer K, Thompson R (1984) Bias in variance and covariance component estimators due to selection on a correlated trait. *J Anim Breed Genet* 101, 33-50
- Patterson HD, Thompson R (1971) Recovery of inter block-information when block sizes are unequal. *Biometrika* 52, 545-554
- Robertson A (1966) A mathematical model of the culling process in dairy cattle. *Anim Prod* 8, 95-108
- Swalve H, Van Vleck LD (1987) Estimation of genetic (co)variances for milk yield in first three lactations using an animal model and restricted maximum likelihood. *J Dairy Sci* 70, 842-849
- Swanson GJT, Gnanasakthy A (1991) Estimation of genetic parameters for milk cell count and correlations with production traits in Friesian Holstein heifers. In: *Proc Winter Meet Br Soc Anim Prod* (abstr), pap 168
- Thompson R, Hill WG (1990) Univariate REML analyses for multivariate data with the animal model. In: *Proc 4th World Congr Genet Appl Livest Prod, Edinburgh, Scotland, July 23-27*, 13, 484-487

- Van Vleck LD (1986) Evaluation of dairy cattle breeding programs: specialised milk production. *In: Proc 3rd World Congr Genet Appl Livest Prod, Lincoln, Nebraska*, 9, 141-152
- Van Vleck LD, Dong MC (1988) Genetic (co)variances for milk, fat, and protein yield in Holsteins using an animal model. *J Dairy Sci* 71, 3040-3046
- Visscher PM, Thompson R, Hill WG (1991) Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds. *Livest Prod Sci* 28, 273-290
- Visscher PM, Thompson R (1992) Comparison between genetic variances estimated from different types of relatives in dairy cattle. *Anim Prod* (in press)
- Visscher PM, Hill WG, Thompson R (1992) Univariate and multivariate parameter estimates for milk production traits using an animal model. II: Efficiency of selection when using simplified covariance structures. *Genet Sel Evol* 24, 431-447
- Wiggans GR, Misztal I, Van Vleck LD (1988a) Implementation of an Animal Model for genetic evaluation of dairy cattle in the United States. *J Dairy Sci* (suppl 2) 71, 54-69
- Wiggans GR, Misztal I, Van Vleck LD (1988b) Animal model evaluation of Ayrshire milk yield with all lactations, herd-sire interaction, and groups based on unknown parents. *J Dairy Sci* 71, 1319-1329