



**HAL**  
open science

**Application de la méthode du maximum de vraisemblance restreint (REML) à l'estimation des paramètres génétiques des trois premières lactations en race montbéliarde.**

Catherine Beaumont

► **To cite this version:**

Catherine Beaumont. Application de la méthode du maximum de vraisemblance restreint (REML) à l'estimation des paramètres génétiques des trois premières lactations en race montbéliarde.. *Genetics Selection Evolution, BioMed Central*, 1989, 21 (4), pp.493-506. hal-02718531

**HAL Id: hal-02718531**

**<https://hal.inrae.fr/hal-02718531>**

Submitted on 1 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Restricted maximum likelihood estimation of genetic parameters for the first three lactations in the *Montbéliarde* dairy cattle breed

✉ DBO  
JFC  
C. Beaumont

*Institut National de la Recherche Agronomique, Station de Recherches Avicoles, Nouzilly, 37380 Monnaie, France*

(received 12 January 1989, accepted 24 August 1989)

**Summary** – Genetic parameters for the first three lactations have been estimated for the main dairy traits (milk, fat, protein and useful yields adjusted for lactation length, fat and protein contents). Two data sets were analysed, including records on 30 751 cows born from 128 young sires and 52 proven sires. Daughters' performances from the most widely used proven sires were incorporated in order to improve the degree of connectedness among herds. The model fitted young sires as random and proven sires, herd-year, season-year of calving, age at first calving and length of the previous lactation as fixed effects. Relationships among bulls were included. Analysis was by restricted maximum likelihood using an EM-related algorithm and a Cholesky transformation. All genetic correlations were larger than 0.89. Correlations between the first and third lactations were slightly lower than the others. Heritabilities of milk, fat, protein and useful yields ranged from 0.17 to 0.27. Phenotypic correlations between successive lactations were higher than 0.6 and those between lactations 1 and 3 lower than 0.55. Heritabilities of fat and protein contents were higher than 0.44 with phenotypic correlations being stable at about 0.70. The "repeatability model" which considers all lactation records as a single trait can be considered in genetic evaluation procedures for dairy traits without significant losses in efficiency.

**dairy cattle – milk yield – fat and protein contents – genetic parameters – maximum likelihood**

**Résumé** – Application de la méthode du maximum de vraisemblance restreint (REML) à l'estimation des paramètres génétiques des trois premières lactations en race montbéliarde. Ce travail a pour but l'estimation des paramètres génétiques des 3 premières lactations des femelles Montbéliardes et porte sur les principales caractéristiques laitières (productions, ajustées pour la durée de lactation, de lait et de matières utiles, grasses et protéiques, taux butyreux et protéique). Deux fichiers sont étudiés. Ils rassemblent les performances de 30 751 femelles issues de 128 taureaux de testage et de 52 taureaux de service. Ceux-ci sont introduits dans l'analyse pour améliorer les connexions entre troupeaux. Le modèle comporte l'effet aléatoire "père de testage" et les effets fixes "père de service", "troupeau-année", "âge au premier vêlage", "année-saison de vêlage" et "durée de la lactation précédente". L'apparement des reproducteurs mâles est considéré. Les données transformées par la décomposition de Cholesky sont analysées par le maximum de vraisemblance restreint avec un algorithme apparenté à l'E.M.

Les corrélations génétiques des 6 caractères, toujours supérieures à 0,89, sont légèrement plus faibles pour les lactations 1 et 3. Pour les caractères de production, l'héritabilité varie

de 0,17 à 0,27. Les corrélations phénotypiques sont supérieures à 0,60 pour les lactations successives et inférieures à 0,55 pour les lactations 1 et 3. Les taux présentent une hérédité supérieure à 0,44 et des corrélations phénotypiques voisines de 0,7 et pratiquement indépendantes du couple de lactations considéré. Ces résultats indiquent que les différentes lactations peuvent être traitées comme des répétitions d'un même caractère. Ce modèle, dit de "répétabilité" permet d'alléger les calculs sans diminuer l'efficacité de la sélection.

**bovins laitiers - production laitière - composition du lait - paramètres génétiques - maximum de vraisemblance**

## INTRODUCTION

The goal of dairy selection is to improve lifetime production of cows, which implies taking into account the different lactations. Until now, genetic evaluation of the animals has in most cases been made under the assumption that these lactations are influenced by the same genes. In some countries only the first lactations are considered; in others the so-called "repeatability model" (Henderson, 1987) in which all lactations are treated as repetitions of one trait is fitted. But the lactations are made at various ages and physiological status of the animals and may therefore be determined somewhat by different genes. The accuracy of the genetic evaluation and thus the efficiency of dairy selection might be improved by fitting a multi-trait model to the lactations. Reliable estimates of the genetic parameters for the different lactations are needed to appreciate this possible gain in accuracy.

Data usually available for such estimations are selected as breeders cull about one quarter of the animals by the end of each lactation. Their decision is mostly based on dairy performance. Useful methods of estimation of these parameters have been available only recently. Henderson's methods (1953) assume animals are measured for all lactations, thus leading to results biased by the selection. However, the maximum likelihood (ML) (Hartley and Rao, 1967) and restricted maximum likelihood (REML) (Patterson and Thompson, 1971) estimators can take into account this selection (Im *et al.*, 1987), a necessary condition being that the selection process is based only on the observed data or on observed data and independent variables. REML was preferred to ML as it accounts for the loss of degrees of freedom in simultaneous estimation of the fixed effects. Moreover, theoretical studies have shown that the optimum statistical procedure maximising the genetic merit of selected animals consists of estimating variance and covariance components by REML and thereafter applying these estimates in the mixed model equations (Gianola *et al.*, 1986).

## MATERIALS AND METHODS

### Data

Records for the first 3 lactations of *Montbéliarde* cows whose first calving occurred between 1/09/1979 and 30/08/1982 were extracted from the National Milk Recording files. The conditions of editing are presented in Table I. Records made after cows changed herds were disregarded (Meyer, 1984). They represented 1.5% of the records for second lactation and 1.3% of the records for third lactation. Cows were nested

**Table I.** Selection criteria of records.

| <i>Criteria</i>                                       | <i>Norm</i>        |
|---|--------------------|
| <i>Selection criteria of animals</i>                  |                    |
| <i>Breed</i>  |                    |
| First calving date (day, month, year)                 | 1/9/79 – 30/8/82   |
| Age at first calving (month)                          | 22-42              |
| Dry period length (days)                              | <213               |
| Interval between calvings (days)                      | 250-603            |
| Milk yield (kg)                                       | 450-13000          |
| Fat yield (kg)  | 10-450             |
| Protein yield (kg)                                    | 10-450             |
| <i>Selection criteria of bulls</i>                    |                    |
| Total number of daughters                             |                    |
| – of the sampling bulls                               | >40                |
| – of the proven bulls                                 | >200               |
| Number, in the data set, of 1st lactation performance |                    |
| – of the sampling bulls                               | >0                 |
| – of the proven bulls                                 | >75                |
| Number, in the data set, of 3rd lactation performance |                    |
| – of the sampling bulls                               | >0                 |
| – of the proven bulls                                 | >0                 |
| Birth year of the sampling bulls                      |                    |
|   | 1975 (data set I)  |
|   | 1976 (data set II) |
| <i>Selection criteria of herds</i>                    |                    |
| Number of cows  | >1                 |
| Number of sampling cows                               | >0                 |

within herds and the absorption matrix of the herd effects was block diagonal for herds.

Two populations of females were considered. The first was made of daughters of test bulls. It was used to estimate sire components of variance and covariance. The second consisted of daughters of the most widely used proven sires. As these bulls had been selected, they were treated as fixed effects and were not considered for the estimation of sire components. The performances of their daughters were introduced in the analysis in order to improve the accuracy of the estimation through additional information, increased herd size and degree of connectedness between herds.

A total of 180 bulls of which 128 were random test bulls was considered. To simplify computation, records were split into 2 data sets, as did Meyer (1984, 1985a) and Swalve and Van Vleck (1987). The first data set consisted of the daughters of sampling bulls born in 1975 and the second of the daughters of sampling bulls born in 1976. For each of the 2 data sets, the most widely used

proven sires were determined and their daughters added. Table II summarizes their main characteristics.

The main dairy variables were considered: milk, fat, protein and useful yields, and fat and protein contents. Useful yield (UY) is defined as:

$$UY = (\text{fat yield} + (1.2 \times \text{protein yield}))/2$$

Yield traits were corrected multiplicatively for lactation length prior to analysis, according to Poutous and Mocquot (1975) as:

$$\text{Corrected yield} = (\text{total yield} \times 385)/(\text{lactation length} + 80)$$

Data were scaled to reduce rounding errors.

### Model

The following model was used for each of the 6 variables:

$$y = Xh + Wb + Zu + e$$

where

$y$  is the vector of the observations;

$h$  is the vector of fixed herd effects (the number of levels of which is shown in Table II);

$b$  is the vector of fixed year-season of calving (15 levels for each lactation), age at 1st calving (10 levels for each lactation) and length of the preceding lactation effects (8 levels for each lactation);

$u$  is the vector of the sire effects (this effect was treated as fixed when the sire was a proven bull, and as random and normally distributed when the sire was a young bull); and

$e$  is the vector of residual effects, assumed normally distributed;

$X$ ,  $W$  and  $Z$  are known incidence matrices for the herd effects, the other fixed effects and the sire effects.

Expectations and variances are defined as:

$$E \begin{bmatrix} u_1 \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{Var} \begin{bmatrix} u_1 \\ e \end{bmatrix} = \begin{bmatrix} G & O \\ O & R \end{bmatrix}$$

where  $u_1$  is the subvector of  $u$  corresponding to the effects of the young bulls and:

$$G = A * T$$

where  $A$  is the relationship matrix of the young bulls,  $T$  the matrix of the sire components and  $*$  the right direct product (Graybill, 1983). Let  $n$  denote the number of animals; with data ordered by lactations within animals,  $R$  is block diagonal having  $n$  blocks  $R_k$  ( $k = 1, \dots, n$ ). If the  $k$ th cow has made the first 3 lactations,  $R_k = E$  where  $E$  is the matrix of residual components; if it has been culled before, the rows and columns corresponding in  $E$  to the missing records are deleted.

### Method

Data were Cholesky transformed (Schaeffer, 1986) but, because the incidence matrix  $W$  varied for an animal from one lactation to the next, the vector  $b$  could

**Table II.** Characteristics of the data sets.

|  | <i>Data set</i> |           |
|--|-----------------|-----------|
|  | <i>I</i>        | <i>II</i> |
| Number of records                                |                 |           |
| 1st lactation                                    | 16343           | 14408     |
| 2nd lactation                                    | 12216           | 10978     |
| 3rd lactation                                    | 8970            | 8140      |
| Number of records of daughters of sampling bulls |                 |           |
| 1st lactation                                    | 4374            | 3770      |
| 2nd lactation                                    | 2991            | 2665      |
| 3rd lactation                                    | 2064            | 1885      |
| Birth year of sampling bulls                     | 1975            | 1976      |
| Number of sampling bulls                         | 66              | 62        |
| Number of proven bulls                           | 28              | 24        |
| Number of herds                                  |                 |           |
| 1st lactation                                    | 3214            | 2725      |
| 2nd lactation                                    | 2823            | 2439      |
| 3rd lactation                                    | 2395            | 2089      |
| Mean of milk yield (kg)                          |                 |           |
| 1st lactation                                    | 4301            | 4346      |
| 2nd lactation                                    | 5140            | 5178      |
| 3rd lactation                                    | 5543            | 5575      |
| Mean of fat yield (kg)                           |                 |           |
| 1st lactation                                    | 160             | 162       |
| 2nd lactation                                    | 191             | 193       |
| 3rd lactation                                    | 206             | 208       |
| Mean of protein yield (kg)                       |                 |           |
| 1st lactation                                    | 137             | 139       |
| 2nd lactation                                    | 165             | 166       |
| 3rd lactation                                    | 176             | 178       |
| Mean of useful yield (kg)                        |                 |           |
| 1st lactation                                    | 162             | 164       |
| 2nd lactation                                    | 194             | 196       |
| 3rd lactation                                    | 209             | 211       |
| Mean of fat content (g/kg)                       |                 |           |
| 1st lactation                                    | 37.2            | 37.3      |
| 2nd lactation                                    | 37.1            | 37.3      |
| 3rd lactation                                    | 37.2            | 37.3      |
| Mean of protein content (g/kg)                   |                 |           |
| 1st lactation                                    | 31.9            | 32.1      |
| 2nd lactation                                    | 32.0            | 32.1      |
| 3rd lactation                                    | 31.8            | 31.9      |

not be transformed and the mixed model equations for this parameter remained unchanged.

A combined REML/ML procedure (Meyer, 1983a, 1984, 1985a) was used. From a Bayesian viewpoint, only herd effects were integrated in the posterior density. From a classical viewpoint, the inference was based on the likelihood of  $(n - r(\mathbf{X}))$  error contrasts  $\mathbf{K}'\mathbf{y}$  (where  $\mathbf{K}$  is an  $n \times (n - r(\mathbf{X}))$  matrix such that  $\mathbf{K}'\mathbf{X} = 0$  (Harville, 1977)). An algorithm "related to the E.M." (Henderson, 1985) was used. First derivatives of the restricted likelihood function are set to zero (eqn. (7) of Meyer (1986)). For the computation of residual components, use is made of eqn. (8) of Meyer (1986). For computations, the iterations were stopped when the relative difference between the estimates of the components from one round to the following fell below 1%.

The asymptotic standard errors of the estimates were calculated. This required the computation of the information matrix, which is very extensive. Therefore, in this part of the analysis, the fixed effects of the year-season of calving, of the age at 1st calving and of the length of the preceding lactation were ignored, so that the Cholesky transformation was fully efficient. The relationships among bulls were also ignored in order to reduce the computational requirements. The information matrix  $\mathbf{I}_c$  of the transformed data was first calculated and, after back transformation, the information matrix  $\mathbf{I}$  of the original data was obtained as it can be showed that:

$$\mathbf{I} = \mathbf{D}\mathbf{I}_c\mathbf{D}'$$

where  $\mathbf{D}$  is the  $(6 \times 6)$  matrix whose element  $(i, j)$  is

$$\frac{\delta\theta_{cj}}{\delta\theta_i}$$

(where  $\theta_c$  is the vector of (transformed) sire and residual variance and covariance components). Computations of the  $\mathbf{I}_c$  matrix was made using Meyer's algorithm (1983a) and taking advantage of the simplifications the Cholesky transformation made possible.

Because of the computational costs, the two data sets were analyzed separately and the mean of the estimates calculated although the two data sets were not totally independent. Similarly, the asymptotic standard errors of the means of the estimates were obtained as if the asymptotic standard errors of the estimates in the 2 data sets were independent.

## RESULTS

The estimates from the 2 data sets differed by less than 1 standard error, except for the variance of the protein content in the third lactation which differed by a little less than 2 standard errors. The asymptotic standard errors in the two data sets differed by less than 0.01 (Table III).

The estimates of the genetic parameters for the yields were similar. The phenotypic variances increased with lactation number. The change of the phenotypic standard deviations was proportional to the increase in the corresponding means and may be considered to be, at least partly, due to a scale effect. By contrast, the genetic components remained nearly constant from the first to the second lactation

Table III. Means of the estimates of genotypic and phenotypic variances.

|                               | Milk yield |                   | Fat yield |                   | Protein yield |                   | Useful yield |                   | Fat content |                   | Protein content |                   |
|-------------------------------|------------|-------------------|-----------|-------------------|---------------|-------------------|--------------|-------------------|-------------|-------------------|-----------------|-------------------|
|                               | Mean       | S.E. <sup>a</sup> | Mean      | S.E. <sup>a</sup> | Mean          | S.E. <sup>a</sup> | Mean         | S.E. <sup>a</sup> | Mean        | S.E. <sup>a</sup> | Mean            | S.E. <sup>a</sup> |
| <i>Genetic variances</i>      |            |                   |           |                   |               |                   |              |                   |             |                   |                 |                   |
| Lactation 1                   | 173000     | 29800             | 244       | 42                | 112           | 22                | 186          | 34                | 5.5         | 0.8               | 2.2             | 0.3               |
| Lactation 2                   | 185000     | 35200             | 235       | 48                | 141           | 30                | 185          | 40                | 4.8         | 0.8               | 2.1             | 0.3               |
| Lactation 3                   | 221000     | 44400             | 316       | 67                | 174           | 39                | 264          | 57                | 5.0         | 0.8               | 1.9             | 0.3               |
| <i>Genetic covariances</i>    |            |                   |           |                   |               |                   |              |                   |             |                   |                 |                   |
| Lactations 1 and 2            | 169000     | 29900             | 224       | 41                | 118           | 23                | 167          | 33                | 4.8         | 0.8               | 2.1             | 0.3               |
| Lactations 1 and 3            | 176000     | 32300             | 247       | 47                | 125           | 25                | 200          | 39                | 4.7         | 0.8               | 1.9             | 0.3               |
| Lactations 2 and 3            | 195000     | 36300             | 261       | 51                | 151           | 31                | 208          | 43                | 4.6         | 0.8               | 1.9             | 0.3               |
| <i>Phenotypic variances</i>   |            |                   |           |                   |               |                   |              |                   |             |                   |                 |                   |
| Lactation 1                   | 635000     | 9080              | 933       | 13                | 629           | 7                 | 882          | 11                | 10.1        | 0.2               | 3.9             | 0.1               |
| Lactation 2                   | 774000     | 11490             | 1179      | 17                | 787           | 11                | 1094         | 15                | 10.9        | 0.2               | 4.1             | 0.1               |
| Lactation 3                   | 801000     | 14180             | 1294      | 22                | 803           | 13                | 1154         | 19                | 11.5        | 0.2               | 3.9             | 0.1               |
| <i>Phenotypic covariances</i> |            |                   |           |                   |               |                   |              |                   |             |                   |                 |                   |
| Lactations 1 and 2            | 438000     | 9110              | 633       | 13                | 443           | 8                 | 603          | 11                | 7.4         | 0.2               | 2.8             | 0.1               |
| Lactations 1 and 3            | 393000     | 9910              | 578       | 15                | 380           | 8                 | 533          | 13                | 7.2         | 0.2               | 2.7             | 0.1               |
| Lactations 2 and 3            | 502000     | 11310             | 753       | 17                | 515           | 11                | 706          | 15                | 7.6         | 0.2               | 2.8             | 0.1               |

<sup>a</sup> Asymptotic standard error of the means of the estimates.



(except for protein yield where it differed by 26%, but this difference was not significant). Except for fat yield, the genetic component of the third lactation was the highest but the difference was not significant.

The heritabilities for the 3 lactations were therefore slightly but not significantly different (Table IV): the heritabilities for the first and third lactations were similar and higher than for the second lactation. The only exception was protein yield, where the heritabilities for the first and second lactations were equal to 0.18 and slightly smaller than for the third lactation (0.22). All genetic correlations were higher than 0.89. The correlation between the first and third lactations was smaller than the others, which except for useful yield were very similar. The same trend was observed for the phenotypic correlations: the correlations between adjacent lactations (first and second lactations or second and third lactations) was higher than between first and third lactations. All phenotypic correlations were between 0.53 and 0.65.

Genetic parameters for the contents measured showed different trends. First the means for the different lactations were very similar (Table II) so there was no scale effect. The genetic components decreased with lactation number, whereas the phenotypic components remained constant. Thus the heritabilities decreased with lactation number, but the differences were not significant. The genetic correlations showed the same trend as for yields and were between 0.90 and 0.96. By contrast, the phenotypic correlations were higher than for yields (between 0.67 and 0.71) and did not vary much.

## DISCUSSION

### *Choice of the method*

Different iterative algorithms may be used for REML estimation. They differ in convergence speed and computational requirements per round of iteration. Primarily, 3 algorithms have been advocated for analysis on selected data: Fisher's method (Meyer (1983a)), algorithms related to the E.M. algorithm of Dempster *et al.* (1977) and Meyer's algorithm (Meyer, 1986). Although Fisher's method has the highest convergence speed, it appears to be the most expensive (Meyer, 1986) and was therefore disregarded. Algorithms called "related to the E.M. algorithm" by Henderson (1985) converge very slowly but have the property of forcing the estimate within the parameter space. Meyer's "short cut" algorithm estimates the residual components via an algorithm related to the E.M. and the genetic components via Fisher's method. Thus the convergence speed is quicker than for algorithms related to the E.M., but the estimates may lie outside the parameter space. The first data set was analysed using Meyer's algorithm, and the estimate of the genetic correlation between first and third lactations was equal to 1.05 and between second and third to 1.09. Such estimates cannot be considered as maximum likelihood estimators (Harville, 1977). They cannot be used in the mixed model equations without using a transformation of the results such as the "bending" of Hayes and Hill (1981). In contrast, the E.M. type algorithm gave estimates that were within the parameter space. It required on our data set slightly less computations than Meyer's algorithm, although 8 iterations were needed before convergence was

Table IV. Means of the estimates of genetic parameters in the two data sets.

| Parameters                     | Milk yield |                   | Fat yield |                   | Protein yield |                   | Useful yield |                   | Fat content |                   | Protein content |                   |
|--------------------------------|------------|-------------------|-----------|-------------------|---------------|-------------------|--------------|-------------------|-------------|-------------------|-----------------|-------------------|
|                                | Mean       | S.E. <sup>a</sup> | Mean      | S.E. <sup>a</sup> | Mean          | S.E. <sup>a</sup> | Mean         | S.E. <sup>a</sup> | Mean        | S.E. <sup>a</sup> | Mean            | S.E. <sup>a</sup> |
| <i>Heritabilities</i>          |            |                   |           |                   |               |                   |              |                   |             |                   |                 |                   |
| Lactation 1                    | 0.27       | 0.04              | 0.26      | 0.04              | 0.18          | 0.03              | 0.21         | 0.04              | 0.55        | 0.07              | 0.57            | 0.1               |
| Lactation 2                    | 0.24       | 0.04              | 0.20      | 0.04              | 0.18          | 0.03              | 0.17         | 0.04              | 0.44        | 0.06              | 0.52            | 0.1               |
| Lactation 3                    | 0.27       | 0.04              | 0.24      | 0.05              | 0.22          | 0.05              | 0.23         | 0.05              | 0.44        | 0.07              | 0.50            | 0.1               |
| <i>Genetic correlations</i>    |            |                   |           |                   |               |                   |              |                   |             |                   |                 |                   |
| Lactations 1 and 2             | 0.95       | 0.03              | 0.94      | 0.03              | 0.94          | 0.04              | 0.90         | 0.05              | 0.94        | 0.02              | 0.95            | 0.01              |
| Lactations 1 and 3             | 0.90       | 0.05              | 0.89      | 0.05              | 0.90          | 0.06              | 0.90         | 0.05              | 0.90        | 0.03              | 0.92            | 0.02              |
| Lactations 2 and 3             | 0.97       | 0.04              | 0.96      | 0.04              | 0.97          | 0.04              | 0.94         | 0.05              | 0.94        | 0.02              | 0.96            | 0.02              |
| <i>Phenotypic correlations</i> |            |                   |           |                   |               |                   |              |                   |             |                   |                 |                   |
| Lactations 1 and 2             | 0.62       | 0.01              | 0.60      | 0.01              | 0.63          | 0.01              | 0.61         | 0.01              | 0.71        | 0.01              | 0.69            | 0.01              |
| Lactations 1 and 3             | 0.55       | 0.01              | 0.53      | 0.01              | 0.54          | 0.01              | 0.53         | 0.01              | 0.67        | 0.01              | 0.69            | 0.01              |
| Lactations 2 and 3             | 0.64       | 0.01              | 0.61      | 0.01              | 0.65          | 0.01              | 0.63         | 0.01              | 0.68        | 0.01              | 0.71            | 0.01              |

<sup>a</sup> Asymptotic standard error of the means of the estimates.

achieved (instead of 6 with Meyer's algorithm), as each iteration took 25% more time with Meyer's algorithm than with the E.M. type algorithm.

The Cholesky transformation makes the absorption of the herd effects quicker. In reference to the time necessary for the absorption of untransformed data, the same process took after Cholesky transformation 43% more time on the first round and 48% less time on the following. This transformation also spares a lot of computations for the estimation of the asymptotic standard errors.

### ***Results of different methods***

Maximum likelihood estimators can only take into account selection if it is based on observed data or on observed data and other independent variables. In this analysis, selection occurred both between generations (for the choice of the parents) and within a generation (by the end of each lactation). As the performance of the parents of the animals could not be analysed because of the computational requirements, only the later selection was considered. This is the case for most REML estimates of genetic parameters for lactations. The results are in accordance with studies using maximum likelihood related estimators (Tables V and VI). Except for Rothschild and Henderson (1979), the authors used restricted maximum likelihood estimates but the algorithms varied: Fisher's method for Meyer (1983a) and Hagger *et al.* (1982); "short cut" for Meyer (1985a); E.M. type for Colaco *et al.* (1987), Rothschild and Henderson (1979), Simianer (1986b), Swalve and Van Vleck (1987) and Tong *et al.* (1979). All considered a sire model except for Swalve and Van Vleck (1987), who used an animal model which may better take into account selection, since all relationships are included in the model (Sorensen and Kennedy, 1984) and thus did not observe any decrease in heritability for the second lactation. This decrease that most authors observe might be due to a selection bias. However Swalve and Van Vleck (1987) neglected relationships among herds and thus ignored selection across herds.

Henderson's methods lead to different results as they are affected by selection of the data (Rothschild *et al.*, 1979; Meyer and Thompson, 1984). Because of selection at the end of first lactation, they underestimate heritabilities for later lactations. For milk yield, the weighted means of the estimates in the literature are 0.26 for first lactation, 0.20 for second lactation and 0.17 for third lactation (Maijala and Hannah, 1974). In the first data set, Henderson's method III estimates for useful yield were respectively 0.21, 0.08 and 0.19. The first lactation estimate is in accordance with REML estimates because selection has not yet occurred. The third lactation estimate does not differ much either. As the criteria of selection depend less on milk production at the end of the second lactation than of first lactation, the selection bias may be less important. This result may also be, at least partly, due to sampling errors. Similarly, the decrease in the heritabilities for the later lactations for content measures may partly be due to the fact that the selection bias is not well removed for these variables, because the selection criteria are usually based more on milk yield than on content.

The parameters of the first 3 lactations are very similar. The heritabilities for the first and third lactations may, at least for yields, be treated as equal. The slight decrease of the heritability for the second lactation is not significant. It may, at least partly, be due to a selection bias which cannot be totally removed when using

Table V. Maximum likelihood estimates of heritabilities.

| Authors                        | Country | Milk yield |         |         | Fat yield |         |         | Protein yield |         |         | Fat content |         |         | Protein content |         |         |
|--------------------------------|---------|------------|---------|---------|-----------|---------|---------|---------------|---------|---------|-------------|---------|---------|-----------------|---------|---------|
|                                |         | $h_1^2$    | $h_2^2$ | $h_3^2$ | $h_1^2$   | $h_2^2$ | $h_3^2$ | $h_1^2$       | $h_2^2$ | $h_3^2$ | $h_1^2$     | $h_2^2$ | $h_3^2$ | $h_1^2$         | $h_2^2$ | $h_3^2$ |
| Colago <i>et al.</i> , 1987    | POR     | 0.30       |         |         |           |         |         |               |         |         |             |         |         |                 |         | 0.63    |
| Hagger <i>et al.</i> , 1982    | S       | 0.29       |         |         |           |         |         |               |         |         |             |         |         |                 |         |         |
| Lawlor <i>et al.</i> , 1984    | USA     | 0.19       |         |         | 0.26      |         |         |               |         |         |             |         |         |                 |         |         |
| Meyer, 1983a                   | GB      | 0.33       | 0.35    | 0.28    | 0.32      | 0.32    | 0.23    | 0.24          | 0.29    | 0.12    | 0.36        | 0.44    | 0.43    | 0.18            | 0.53    | 0.61    |
| Meyer, 1984                    | GB      | 0.28       | 0.19    | 0.24    | 0.27      | 0.21    | 0.25    |               |         |         |             |         |         |                 |         |         |
| Meyer, 1985a                   | AUS     | 0.17       | 0.15    | 0.18    | 0.13      | 0.17    |         |               |         |         | 0.49        | 0.45    |         | 0.16            | 0.37    |         |
| Meyer, 1985b                   | GB      | 0.21       |         |         | 0.16      |         |         | 0.15          |         |         |             |         |         | 0.15            |         |         |
| Rothschild and Henderson, 1979 | USA     | 0.41       | 0.35    |         |           |         |         |               |         |         |             |         |         |                 |         |         |
| Simianer, 1986b                | WG      | 0.29       | 0.21    | 0.22    |           |         |         |               |         |         |             |         |         |                 |         |         |
| Swalve and Van Vleck, 1987     | USA     | 0.33       | 0.33    | 0.34    |           |         |         |               |         |         |             |         |         |                 |         |         |
| Tong <i>et al.</i> , 1979      | CAN     | 0.26       | 0.19    | 0.17    | 0.26      | 0.17    | 0.15    | 0.19          | 0.15    |         | 0.35        | 0.43    | 0.51    | 0.52            | 0.78    |         |

$h_1^2$  = heritability for the first lactation.

$h_2^2$  = heritability for the second lactation.

$h_3^2$  = heritability for the third lactation.

<sup>a</sup> POR = Portugal, S = Switzerland, GB = Great Britain, AUS = Australia, WG = West Germany, CAN = Canada.

Table VI. Maximum likelihood estimates of genetic correlations among lactations.

| Authors                        | Country | Milk yield |           |           | Fat yield |           |           | Protein yield |           |           | Fat content |           |           | Protein content |           |           |
|--------------------------------|---------|------------|-----------|-----------|-----------|-----------|-----------|---------------|-----------|-----------|-------------|-----------|-----------|-----------------|-----------|-----------|
|                                |         | $r_{g12}$  | $r_{g13}$ | $r_{g23}$ | $r_{g12}$ | $r_{g13}$ | $r_{g23}$ | $r_{g12}$     | $r_{g13}$ | $r_{g23}$ | $r_{g12}$   | $r_{g13}$ | $r_{g23}$ | $r_{g12}$       | $r_{g13}$ | $r_{g23}$ |
| Meyer, 1983a                   | GB      | 1.02       | 0.83      | 0.98      | 0.97      | 0.84      | 1.03      | 0.99          | 0.89      | 1.15      | 0.96        | 0.81      | 0.84      | 0.84            | 0.69      | 1.04      |
| Meyer, 1984                    | GB      | 0.91       | 0.91      | 0.95      | 0.91      | 0.91      | 0.99      |               |           |           |             |           |           |                 |           |           |
| Meyer, 1985a                   | AUS     | 0.96       |           |           | 0.98      |           |           | 1.00          |           |           | 0.98        |           |           | 0.97            |           |           |
| Rothschild and Henderson, 1979 | USA     | 0.92       |           |           |           |           |           |               |           |           |             |           |           |                 |           |           |
| Simianer, 1986b                | WG      | 0.81       | 0.77      | 0.92      |           |           |           |               |           |           |             |           |           |                 |           |           |
| Swalve and Van Vleck, 1987     | USA     | 0.88       | 0.85      | 0.87      |           |           |           |               |           |           |             |           |           |                 |           |           |
| Tong <i>et al.</i> , 1979      | CAN     | 0.89       | 0.85      | 0.89      | 0.90      | 0.87      | 0.90      | 0.83          |           |           | 0.92        | 0.88      | 0.92      | 0.92            | 0.89      |           |

$r_{g12}$  = genetic correlation between first and second lactations.

$r_{g13}$  = genetic correlation between first and third lactations.

$r_{g23}$  = genetic correlation between second and third lactations.

<sup>a</sup> GB = Great Britain, AUS = Australia, WG = West Germany, CAN = Canada.

a sire model. The determinism of the second lactation may also be slightly different: this performance depends both on the dairy value of the animal and on its ability to recover from both growth and first lactation. The genetic correlations are very high but the correlation between first and third lactations is significantly different from 1. The older the cow is, the more disease it has had to resist and the more its ability to resist is important in the determinism of its lactations. However, the differences in the parameters of the lactations are very small.

It does not seem to be necessary to modify the current French genetic evaluation procedure which fits a repeatability model to the different lactations. All available lactations are taken into account because of the small mean herd size (34.5 cows per herd). Accuracy is increased using all records instead of first records only. This gain is due to both extra genetic information and increased degree of connectedness among herds (Meyer, 1983b). Ufford *et al.* (1979) reported such an increase even for young bulls whose daughters had only first lactations. Fitting a multi-trait model would imply a very large increase in computational requirements, as time needed for an iterative inversion of the coefficient matrix of the mixed model equations is proportional to the square of its size. But only a very small gain in accuracy could be expected. Simianer (1986a) and Schulte-Coerne (1983) estimate this increase to be less than 1% when 3 lactations are considered and all the genetic correlations are 0.80. The difference is expected to be even smaller in our case because the correlations are higher. However, they restricted their analysis to complete data (*i.e.* all animals were supposed to have made 3 lactations). In reality, some selection occurs. The selection bias can be totally removed only when the true genetic parameters are used, *i.e.* with the multi-trait model. But the difference between the 2 models is still expected to be small.

## ACKNOWLEDGMENTS

This research was conducted at INRA Département de Génétique Animale at the Station de Génétique Quantitative et Appliquée. The author wishes to thank F. Grosclaude and L. Ollivier for their support, B. Bonaiti and 2 referees for their useful comments on the manuscript and J.J. Colleau and J.L. Foulley for their helpful discussions and suggestions.

## REFERENCES

- Colaco J.A., Fernando R.L. & Gianola D. (1987) Variability in milk production among sires, herds and cows in Portuguese dairy cattle. *In: 38th Annual Meeting of the European Association for Animal Production, Lisbon, Portugal, Sept. 27-Oct. 1, 1987* (Abstr.)
- Dempster A.P., Laird N.M. & Rubin D.B. (1977) Maximum likelihood from incomplete data via the E.M. algorithm. *J.R. Stat. Soc. B.* 39, 1-22
- Gianola D., Foulley J.L., Fernando C.L. (1986) Prediction of breeding values when variances are not known. *Génét. Sél. Evol.* 18, 485-497
- Graybill F.A. (1983) *Matrices with applications in statistics*. 2nd Edn., Wadsworth, Belmont, USA

- Hagger C., Schneeberger M. & Dempfle L. (1982) ML, REML, MINQUE and Henderson 3 estimates of variance and covariance components for milk yield, fat and protein content of *Braunvieh* and *Brown Swiss* × *Braunvieh* sires. In: *Proceedings of the 2nd World Congress on Genetics Applied to Livestock Production, Madrid, Oct. 4-8 1982*, vol. 8, Editorial Garsi, Madrid, pp. 63-68
- Hartley H.O. & Rao J.N.K. (1967) Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54, 93-108
- Harville D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat., Ass.* 72, 320-338
- Hayes J.F., Hill W.G. (1981) Modification of estimates of parameters in the construction of genetic selection indices ("bending"). *Biometrics* 37, 483-493
- Henderson C.R. (1953) Estimation of variance and covariance components. *Biometrics* 9, 226-252
- Henderson C.R. (1985) MIVQUE and REML estimation of additive and non-additive genetic variances. *J. Anim. Sci.* 61, 113-121
- Henderson C.R. (1987) Theoretical aspects of genetic evaluation under the animal model. In: *38th Annual Meeting of the European Association for Animal Production, Lisbon, Portugal, Sept. 27-Oct. 1, 1987*
- Im S., Fernando R.L. & Gianola D. (1987) Likelihood inferences in animal breeding under selection: a missing data theory view point. Submitted to *Biometrics*
- Lawlor T.J., Pollack E.J. & Quaas R.L. (1984) Estimation of variance components with relationships included for a multiple trait model. *J. Dairy Sci.* 67 (Suppl. 1), 181 (Abstr. P. 204)
- Maijala K. & Hanna M. (1974) Reliable phenotypic and genetic parameters in dairy cattle. In: *Proceedings of the 1st World Congress on Genetics Applied to Livestock Production, Madrid, Oct. 7-11 1974*, vol. 1, Editorial Garsi, Madrid, pp. 541-563.
- Meyer K. (1983a) Maximum likelihood procedures for estimating genetic parameters for later lactations of dairy cattle. *J. Dairy Sci.* 66, 1988-1997
- Meyer K. (1983b) 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 3 lactations in British Friesian cows. *Anim. Prod.* 38, 313-322
- Meyer K. (1985a) Genetic parameters for dairy production of Australian Black and White cows. *Livest. Prod. Sci.* 12, 205-219
- Meyer K. (1985b) Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* 41, 153-165
- Meyer K. (1986) Between algorithms: a "short cut" restricted maximum likelihood procedure to estimate variance components. *J. Dairy Sci.* 69, 1904-1916
- Meyer K. & Thompson R. (1984) Bias in variance and covariance components estimators due to selection on a correlated trait. *J. Anim. Breed. Genet.* 101, 33-50
- Patterson H.D. & Thompson R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545-554

- Poutous M., Mocquot J.C. (1975) Etudes sur la production laitière des bovins. III. Relations entre critères de production, durée de lactation et intervalles entre le premier et le deuxième vêlages. *Ann. Génét. Sél. Anim.* 7, 181-189
- Rothschild M.F. & Henderson C.R. (1979) Maximum likelihood estimates of parameters of first and second lactation milk records. *J. Dairy Sci.* 62, 990-995
- Rothschild M.F., Henderson C.R. & Quaas R.L. (1979) Effects of selection on variances and covariances of simulated first and second lactations. *J. Dairy Sci.* 62, 996-1002
- Schaeffer L.R. (1986) Estimation of variances and covariances within the allowable parameter space. *J. Dairy Sci.* 69, 187-194
- Schulte-Coerne H. (1983) Comparison of selection indices using repeatability models and multiple trait models. In: *34th Annual Meeting of the European Association for Animal Production, Madrid, Oct. 3-6, 1983*
- Simianer H. (1986a) A general approach to the use of multiple traits with repeated measurements in estimation of breeding values. *Livest. Prod. Sci.* 15, 315-324
- Simianer H. (1986b) Restricted maximum likelihood estimation of variances and covariances from selected data. In: *Proceedings of the 3rd World Congress on Genetics Applied to Livestock Production, Lincoln, Nebraska, July 16-22, 1986* (Dickerson G.E. and Johnson R.K., éd.), University of Nebraska, USA, pp. 421-426
- Sorensen D.A. & Kennedy B.W. (1984) Estimation of genetic variances from unselected and selected populations. *J. Anim. Sci.* 59, 1213-1223
- Swalve H. & Van Vleck L.D. (1987) Estimation of genetic (co) variances for milk yield in first 3 lactations using an animal model and restricted maximum likelihood. *J. Dairy Sci.* 70, 842-849
- Tong A.K.W., Kennedy B.W. & Moxley J.E. (1979) Heritabilities and genetic correlations for the first 3 lactations from records subject to culling. *J. Dairy Sci.* 62, 1784-1790
- Ufford G.R., Henderson C.R., Keown J.F. & Van Vleck L.D. (1979) Accuracy of first lactation versus all lactation sire evaluations by best linear unbiased predictor. *J. Dairy Sci.* 62, 603-612