

Computing algorithm for dairy sire evaluation on several lactations considered as the same trait

B. BONAÏTI and Michèle BRIEND

*I.N.R.A., Station de Génétique quantitative et appliquée,
Centre de Recherches Zootechniques, F 78350 Jouy-en-Josas*

Summary

A computing algorithm is suggested for dairy sire evaluation on several lactations considered as the same trait when the model must include herd-year (HY), cow and sire as well as other environmental effects that HY (ENV). After description of equations leading to estimates of the different effects and of available computing methods, some improvements are proposed : 1) A method for cow equations absorption is described. 2) Instead of absorption of HY equations which is highly time consuming, computing of HY, ENV and sire effects by a block iterative procedure, is suggested. 3) Expressing all the former records as deviations from previous HY and ENV estimates, is proposed to combine former and recent data sets for sire evaluation without increasing too much the computing length.

Key words : Breeding value, BLUP, dairy cattle, computing algorithm.

Résumé

Algorithme de calcul pour l'évaluation de la valeur génétique des taureaux laitiers sur plusieurs lactations considérées comme un seul caractère

Une méthode de calcul est proposée pour l'indexation des taureaux laitiers sur plusieurs lactations, considérées comme un seul caractère, quand le modèle d'analyse doit tenir compte des effets troupeau-année (HY), d'environnement autres que HY (ENV), vache et père. Après une présentation des équations conduisant aux estimations des différents effets et des méthodes de résolution, quelques améliorations sont proposées : 1) Une méthode est décrite pour l'absorption des équations vache. 2) Au lieu de recourir à l'absorption des équations HY, qui serait trop longue, il est possible d'obtenir les solutions correspondantes aux effets HY, ENV et père par une procédure itérative. 3) En exprimant les performances antérieures en écart aux effets HY et ENV, à l'aide des solutions obtenues lors des calculs antérieurs, on propose de combiner les données anciennes et récentes pour l'estimation de la valeur génétique des taureaux sans trop augmenter la complexité des calculs.

Mots clés : Valeur génétique, BLUP, bovins laitiers, algorithme de calcul.

I. Introduction

The theoretical principles for estimation of breeding values were established by LUSH (1931) and later perfected by HENDERSON (1973) with the Best Linear Unbiased Predictor (BLUP). Computations providing BLUP estimates are very similar to those of least squares, and many applications have already been made in different species and for different characters. For large data sets, and especially for analysis with complicated models, computations can be very time consuming. In France, an algorithm such as proposed by UFFORD *et al.* (1978) for dairy sire evaluation with several lactations has not been used for 2 main reasons. First, the model had to include other environmental factors than those of herd effects. Second, including the complete data set in each analysis, as required for several lactations, would have led to excessive computing. For French dairy sire evaluation, POUTOUS *et al.* (1981) use an easier method based on data from the last three years only. This method enables to handle a large model because it does not require setting up of the coefficient matrix. Its 2 main features are an estimate of each effect obtained from a regressed mean deviation of data corrected for the other effects by estimates from previous analysis and a « selection » factor, at each level of which cows are ranked according to their first lactation deviation, to prevent cow effects in the model.

An alternative to this procedure currently applied in France, is proposed in this paper. The BLUP principles are maintained but some of the approximations of the French dairy sire evaluation method are adopted.

II. BLUP equations

Four main sources of variation are usually considered in the analysis of dairy field records :

- the sire, and in some cases, the maternal grandsire,
- the herd-year-season or herd-year effects (HY),
- the cow, if several lactations are considered for the same cow,
- and a set of other factors called ENV, related to the environment, but independant of HY. These factors can be month of calving, age and parity. Usually, they do not appear in the model of analysis as the data can be corrected for these factors prior to the analysis. However, in France, they have been included in the model from the onset of dairy sire evaluation.

The following linear model can then be chosen for the analysis of data and sire evaluation by the BLUP procedure :

$$Y = Sp + Tm + Rh + Zc + E$$

where p, m, h, c represent vectors of sire, ENV, HY and cow within sire effects, S, T, R and Z the related design matrices. Vector E represents random residual effects and is assumed to be multinormally distributed with covariance matrix $V \cdot \sigma_e^2$. Matrix V is assumed to be diagonal and the element corresponding to the 1th record of the kth cow is :

$$V [(kl) ; (kl)] = 1/w_{kl}$$

Thus, complete and incomplete records may be given different weights (w_{id}) according to lactation length as in French dairy sire evaluation (POUROUS *et al.*, 1981). Sire (p) and cow (c) effects are also assumed to be random effects with expected zero value. If σ_y^2 and r are variance and repeatability of records, if A is the numerator relationship of the sires and if the sire variance is $1/4$ additive genetic variance, we have :

$$\sigma_c^2 = (1 - r) \sigma_y^2$$

$$V(p) = 0.25 h^2 \sigma_y^2 A = (0.25 h^2 / (1 - r)) \sigma_c^2 A$$

$$V(c) = (r - 0.25 h^2) \sigma_y^2 = ((r - 0.25 h^2) / (1 - r)) \sigma_c^2$$

With these assumptions, the sire evaluation according to the BLUP methodology, is derived from :

$$\begin{bmatrix} S'V^{-1}S + kA^{-1} & S'V^{-1}T & S'V^{-1}R & S'V^{-1}Z \\ T'V^{-1}S & T'V^{-1}T & T'V^{-1}R & T'V^{-1}Z \\ R'V^{-1}S & R'V^{-1}T & R'V^{-1}R & R'V^{-1}Z \\ Z'V^{-1}S & Z'V^{-1}T & Z'V^{-1}R & Z'V^{-1}Z + \alpha I \end{bmatrix} \begin{bmatrix} \hat{p} \\ \hat{m} \\ \hat{h} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} S'V^{-1}Y \\ T'V^{-1}Y \\ R'V^{-1}Y \\ Z'V^{-1}Y \end{bmatrix} \quad (I)$$

where $k = (1 - r)/0.25 h^2$ and $\alpha = (1 - r)/(r - 0.25 h^2)$.

UFFORD *et al.* (1978) and SCHAEFFER (1975) described efficient methods to be used when the ENV effects are not in the model and when cow effects can be considered to be within herd nested. The 2 main steps are :

- absorption of cow and then HYS equations,
- solution of the resulting equations by an iterative procedure.

III. Adaptation to model with sire, herd-year and other environmental effects (ENV)

The 2 successive absorptions of cow and herd-year equations are more difficult when the ENV effects are considered in addition to sire effects. On the one hand, the resulting equations is too large to be set up within core storage. If each element must be stored on peripheral storage equipment and accumulated later, then the number of these elements is too large. In addition, cow effects are not nested within all the other effects. Some adaptations can then make the sire evaluation easier.

The set of equation (I) can be written :

$$\begin{bmatrix} U'V^{-1}U + \Delta_r & U'V^{-1}Z \\ Z'V^{-1}U & Z'V^{-1}Z + \alpha I \end{bmatrix} \begin{bmatrix} \hat{f} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} U'V^{-1}Y \\ Z'V^{-1}Y \end{bmatrix} \quad (II)$$

where : $\hat{f}' = (\hat{p}', \hat{m}', \hat{h}')$ $U = (S, T, R)$

and Δ_r is a block diagonal matrix, with the same dimensions as $U'V^{-1}U$, in which the upper block relative to sire effect (p) is kA^{-1} and the others are zero matrices.

Later on, f is split up into different factors f_φ so that each row of the related design matrices U_φ has only one non zero element, which is also equal to 1. For example, f_φ may represent months, age or sire effect. If a level of any factor f_φ is related to the u^{th} column of matrix U , sums of weights (w_{ki}) relative to this level (u) will be w_u and w_{ku} , respectively for the whole data set and the k^{th} cow. The related sums for a combination of 2 levels (u and u') will be $w_{uu'}$ and $w_{kuu'}$, respectively.

A. Cow equation absorption

With $\Gamma = Z(Z'V^{-1}Z + \alpha I)^{-1}Z'$

$$\mathbf{p} = U'V^{-1}U - U'V^{-1}\Gamma V^{-1}U$$

$$\mathbf{s} = U'V^{-1}Y - U'V^{-1}\Gamma V^{-1}Y$$

absorption of cow equations leads to :

$$(\mathbf{p} + \Delta_r) \hat{\mathbf{f}} = \mathbf{s} \quad (\text{III})$$

In order to get elements $\mathbf{p}[u;u']$ and $\mathbf{s}[u]$, it is possible to :

- compute the 2 quantities

$$(w_{kuu'} - w_{ku} w_{ku'}) / (w_k + \alpha) \quad \text{and} \quad (y_{ku} - w_{ku} m_k)$$

$$\text{where } w_k = \sum_l w_{kl} \quad m_k = (\sum_l w_{kl} y_{kl}) / (w_k + \alpha)$$

$$y_{ku} = \sum_l U[(kl); u] w_{kl} y_{kl}$$

- for each cow, cumulate them in $\mathbf{p}[u; u']$ and $\mathbf{s}[u]$.

But this method is efficient when values of w_{ku} are large as in the case of absorption of herd-year equations. For cow effects, there are mostly one record per cell defined by the combination of levels u and k . In this case, adding separately for each record a certain quantity to the related element of \mathbf{p} or \mathbf{s} without computing the figures, w_{ku} , $w_{kuu'}$, y_{ku} for each cow, is more efficient. This is possible with an algorithm (derived from results given in appendix) still reliable even if 2 or more records of the same cow appear within the same level of any factor.

Using the notation $(kl\varphi)$ to identify the row or the column in \mathbf{p} or \mathbf{s} related to the level of the factor φ for the l^{th} record of the k^{th} cow, the algorithm proposed here, consists in the following additions for each cow :

- for each record l , add :

$$w_{kl} (y_{kl} - m_k) \text{ to } \mathbf{s} [(kl\varphi)] \text{ for all factors } \varphi.$$

$$(w_{kl} - w_{kl}^2 / (w_k + \alpha)) \text{ to } \mathbf{p} [(kl\varphi); (kl\varphi')] \text{ for all ordered pairs of sub-factors } (\varphi, \varphi') \text{ with } \varphi \text{ equal or not to } \varphi'.$$

- for each ordered pairs of records (l, l') with $l \neq l'$, add $-(w_{kl} w_{kl'}) / (w_k + \alpha)$ to $\mathbf{p} [(kl\varphi); (kl'\varphi')]$ for all the ordered pairs of sub-factors (φ, φ') with φ equal or not to φ' .

B. Absorption of herd year equations or block iterative procedure

I. Absorption procedure

As the equation system (III) remains too large, another absorption of herd-year equations is usually suggested to reduce the size.

The principle of this operation is the following. Let :

$$g' = (p', m') \quad W = (S, T)$$

$$Q_1 = W' (V^{-1} - V^{-1}\Gamma V^{-1}) W, \quad Q_2 = W' (V^{-1} - V^{-1}\Gamma V^{-1}) R$$

$$Q_3 = R' (V^{-1} - V^{-1}\Gamma V^{-1}) R$$

Δ_g be block diagonal matrix, with the same dimensions as Q_1 in which the block relative to sire effect (p) is kA^{-1} and the others zero.

$$r_1 = W' (V^{-1} - V^{-1}\Gamma V^{-1}) Y, \quad r_2 = R' (V^{-1} - V^{-1}\Gamma V^{-1}) Y$$

Equation III can be written in another form :

$$\begin{bmatrix} Q_1 + \Delta_g & Q_2 \\ Q'_2 & Q_3 \end{bmatrix} \begin{bmatrix} \hat{g} \\ \hat{h} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \quad (IV)$$

Absorption of herd-year equations leads to :

$$(Q_1 + \Delta_g^{-1} - Q_2 Q_3^{-1} Q'_2) \hat{g} = r_1 - Q_2 Q_3^{-1} r_2 \quad (V)$$

The cows are assumed to be nested within herds and matrix Q_3 is block diagonal. If matrices Q_1, Q_2, Q_3, r_1, r_2 are split up, according to herd, into $Q_{1j}, Q_{2j}, Q_{3j}, r_{1j}$ and r_{2j} respectively in the following way :

$$Q_1 = \sum_j Q_{1j} \quad Q_2 = (Q_{21}, Q_{22}, \dots, Q_{2j}, \dots) \quad Q_3 = \sum_j^{\oplus} Q_{3j}$$

$$r_1 = \sum_j r_{1j} \quad r'_2 = (r'_{21}, r'_{22}, \dots, r'_{2j}, \dots)$$

the two members of the equation (V) can be derived from :

$$(Q_1 - Q_2 Q_3^{-1} Q'_2) = \sum_j (Q_{1j} - Q_{2j} Q_{3j}^{-1} Q'_{2j})$$

$$(r_1 - Q_2 Q_3^{-1} r_2) = \sum_j (r_{1j} - Q_{2j} Q_{3j}^{-1} r_{2j})$$

However, this absorption appears to be highly time consuming mainly because of the expression $Q_{2j} Q_{3j}^{-1} Q'_{2j}$ for each herd. For example, for a model including 10 year effects and a vector g of 150 levels, computing needs 0.5 seconds per herd and therefore about 7 hours, for the 50 000 herds in the French dairy recording data set. For that reason this method cannot be easily used when a large model is applied to a large data set.

2. Block iterative procedure

Instead of an absorption procedure, one may use a block iterative method in which the solution of the equation IV is derived at the n^{th} iteration from the solution of the previous iteration :

$$Q_3 h^{(n)} = r_2 - Q'_2 g^{(n-1)}$$

$$(Q_1 + \Delta_b) g^{(n)} = r_1 - Q_2 h^{(n)}$$

The following relationship exists between two consecutive solutions of g :

$$(Q_1 + \Delta_b) g^{(n)} = r_1 - Q_2 Q_3^{-1} r_2 + Q_2 Q_3^{-1} Q'_2 g^{(n-1)}$$

which is not very different from that usually used when equation (V) from the absorption procedure is solved by the Gauss-Seidel iterative method. But, for 3 reasons, computations of the solutions may be faster with the block iterative procedure :

a) Computing of $Q_2 Q_3^{-1} Q'_2$ is not necessary with this method.

b) Matrix Q_3 , which is block diagonal, may be inverted only once, at the first iteration, and then stored. This may also be the case of matrix $(Q_1 + \Delta_b)$ if the size of vector g is small or if the relationship matrix A is not considered.

c) The right hand side coefficients can be written in another form :

$$(r_2 - Q'_2 g^{(n-1)}) = R' (V^{-1} - V^{-1} \Gamma V^{-1}) (Y - W g^{(n-1)})$$

$$(r_1 - Q_2 h^{(n)}) = W' (V^{-1} - V^{-1} \Gamma V^{-1}) (Y - R h^{(n)})$$

This may be easily obtained from the previous algorithm relative to cow absorption on a variable corrected for $g^{(n-1)}$ or $h^{(n)}$. Therefore, after the first iteration, only the right hand side coefficients have to be recalculated.

Computing length of the block iterative procedure depends mainly on the number of iteration steps required to reach an acceptable solution. This is related to the convergence towards zero of :

$$\Delta^{(n)} = (\hat{g}^{(n)} - \hat{g}) = [(Q_1 + \Delta_b)^{-1} Q_2 Q_3^{-1} Q'_2]^n (\hat{g}^{(0)} - \hat{g})$$

for which no general method of evaluation is available.

3. Numerical comparison between absorption and block iterative procedure

According to the French sire evaluation, the speed of convergence of $\Delta^{(n)}$ might be very good. Thus the 2 procedures (absorption/block iterative) were compared on a rather large data set (300 000 records) prepared with the first 3 lactations of 3 French departments between 1976 and 1981. Data were analysed according to the 2 following models

$$Y_{ijklmnr} = HY_{ij} + YSP_{jkl} + YSM_{jkm} + V_{ln} + C_{ic} + E_{ijklmnr} \quad (\text{model I})$$

$$Y_{ijsr} = HY_{ij} + S_s + C_{isc} + E_{ijsr} \quad (\text{model II})$$

where $Y_{ijklmnr}$: milk production in kg.

- HY_{ij} : fixed effect of i^{th} herd and j^{th} year.
- YSP_{jkl} : fixed effect of l^{th} parity and of k^{th} calving season within j^{th} year ($3 \times 4 \times 5$ levels).
- YSM_{jkm} : fixed effect of m^{th} month of calving within k^{th} season and j^{th} year ($3 \times 4 \times 5$ levels).
- V_{in} : fixed effect of n^{th} class of age or calving interval (for lactation 2 and 3) within 1^{th} parity (10×3 levels).
- C_{ic} : random effect of c^{th} cow within i^{th} herd with expected value zero and variance σ_c^2 .
- S_s : random effect of s^{th} sire with expected value zero and variance σ_s^2 (4557 sires).
- C_{isc} : the same as C_{ic} but within i^{th} herd and s^{th} sire.

Solutions relative to least squares (model I) or BLUP (model II) equations were obtained with the 2 methods, absorption and block iterative procedures (tabl. 1 et 2). Block iterative estimates rapidly approximated those resulting from the absorption procedure. With model I, the root mean square of the error (difference between block iterative and absorption solutions) quickly decreases. At the fourth iteration the maximum error was less than 2 kg and the root mean square less than 1 kg. Convergence of sire solutions was not as quick with model II, probably because of some association between herds and bulls. After seven iterations, the root mean square of the error was 2.7 kg, and the maximum error 8.6 kg. Comparison of computing times for model I and II shows that the block iterative procedure was much more efficient (tabl. 4).

In practice, the values of most effects being well known before the first iteration, the number of iterations needed to get good solutions could be very small (2 or 3). This enhances the advantage of the block iterative procedure because its computing requirements mostly depend on the number of iterations whereas the computing time for the absorption procedure depends on the absorption itself. However, use of the absorption procedure should not be excluded with model II. The fact that matrix Q_1 is diagonal disappears if the relationship matrix (A) is used in the analysis. An association between herds and bulls might require such a large number of iterations that the absorption procedure may become more efficient in some practical conditions.

The results from models I and II prompted us to try a model III including all effects of both models I or II :

$$Y_{ijcklmnr} = HY_{ij} + YSP_{jkl} + YSM_{jkm} + V_{in} + S_s + C_{isc} + E_{ijcklmnr} \quad (\text{model III})$$

As the absorption procedure would have needed creation of too large a matrix, only the block iterative procedure was used. At each iteration, the herd-year effect (HY_{ij}), ENV effect (YSP_{jkl} , YSM_{jkm} , V_{in}) and sire effect were successively computed. Differences between successive solutions give some information about the speed of convergence towards exact solutions. Statistical parameters of these differences were computed separately for each of the effects : YSP_{jkl} , YSM_{jkm} , V_{in} and S_s . At the 8th iteration the root mean squares of the difference were smaller than 2 kg of milk for all the effects. Particularly the root mean square of difference between the 7th and 8th sire effect solutions was only 1 kg. The maximum differences were less than 4 kg for the effect YSP_{jkl} , YSM_{jkm} and V_{in} and 8 kg for the sire effect (tabl. 3).

TABLE 1
Comparison of block iterative and absorption procedure (Model I).

Number of iteration (n)		1	2	3	4
d	YSP effect				
	— Max		167	53	21
	— Rms er		70	15	5
	YSM effect				
	— Max		34	7	2
	— Rms er		13	3	0.6
V effect	— Max		52	12	4
	— Rms er		26	6	1.4
	YSP effect				
	— Max	243	75	23	2
D	— Rms er	89	19	5	0.3
	YSM effect				
	— Max	43	9	2	0.3
	— Rms er	16	3	0.6	0.1
	V effect				
	— Max	69	17	5	1.4
— Rms er	33	7	1.4	0.6	

Notation
d : difference between 2 successive solutions, n^{th} and $(n - 1)^{th}$, of the block iterative procedure.
D : difference between block iterative (n^{th} iteration) and absorption procedures.
Max and Rms er : the differences related to each level of the same effect are considered as a population for which is given the maximum in absolute value (Max) and the root mean square error (Rms er.).

TABLE 2
Comparison of block iterative and absorption procedure (Model II).

Number of iteration (n)		1	2	3	4	5	6	7	8	9	10
d	Mean		1.0	1.3	1.0	0.9	0.7	0.6	0.5	0.4	0.3
	Max		145	63	30	15	9	6	4	3	3
	Rms er ...		20	8	3.4	1.7	1.1	0.9	0.8	0.6	0.5
D	Mean	- 9	- 6	- 4	- 3	- 2	- 1.4	- 0.9	- 0.6	- 0.3	- 0.3
	Max	291	137	69	37	21	13	9	6	4	2.6
	Rms er ...	35	16	9	5	3.4	2.6	1.7	1.1	0.9	0.6

Notation
d : difference between 2 successive solutions, n^{th} and $(n - 1)^{th}$, of the block iterative procedure.
D : difference between block iterative (n^{th} iteration) and absorption procedures. The set of equations, reduced by absorption of herd year equations, is solved by an iterative method : iterations were stopped at the 24th when the deviation between 2 successive iterations was less than 0.03 kg.
Mean, Max and Rms er : the differences for the 4557 sire effects are considered as a population for which is given the mean (Mean), the maximum in absolute value (Max) and the root mean square (Rms er.).

TABLE 3
Block iterative procedure (Model III).

Number of iteration (n)		1	2	3	4	5	6	7	8
d	YSP effect								
	— Max		141	45	21	14	9	6	4
	— Rms er		50	21	12	7	5	3	2
	YSM effect								
	— Max		34	7	2	0.9	0.9	0.6	0.3
	— Rms er		12	3	0.9	0.6	0.3	0.3	0.0
	V effect								
	— Max		53	14	5	3	2	1.1	0.6
	— Rms er		22	7	3	1	0.9	0.6	0.3
	Sire effect								
	— Mean		- 5	0.9	0.9	0.6	0.3	0.3	0.3
	— Max		158	73	45	29	19	12	8
— Rms er		25	11	7	4	3	2	1	
d*	YSP effect								
	— Max	215	85	54	33	19	10	4	
	— Rms er	92	49	29	17	10	5	2	
	YSM effect								
	— Max	40	8	4	3	1	0.9	0.3	
	— Rms er	15	4	2	1	0.6	0.3	0.0	
	V effect								
	— Max	75	22	8	5	3	1	0.6	
	— Rms er	33	12	6	3	2	0.9	0.3	
	Sire effect								
	— Mean	1	- 3	- 3	- 1	- 1	- 0.3	0.0	
	— Max	287	181	113	68	39	20	8	
— Rms er	49	27	17	10	6	3	1		

Notation

d : difference between 2 successive solutions, n^{th} and $(n - 1)^{\text{th}}$, of the block iterative procedure.

*d** : difference between solutions at n^{th} and 8^{th} iterations.

Mean, Max and Rms er : the differences related to each level of the same effect are considered as a population for which is given the mean (*Mean*) the maximum in absolute value (*Max*) and the root mean square error (*Rms er*).

As no comparison with exact solutions (obtained from the absorption procedure) is possible, the value of the block iterative procedure cannot be accurately established. However, the only other solution would be :

- preparing a set of ENV effect estimates from an analysis of data without sire effect,
- correcting data according to these ENV effect estimates,
- analysing data according to model II.

In comparison, our method described above provides after a few iterations a solution of the ENV effects independent of the sire effect. This is an advantage when there is a relationship between some of the ENV effects (months or age at calving) and

the sire effects. Therefore, because of the short computing time (tabl. 4), this method may be used to analyse data simultaneously for the 3 types of effects (ENV, herd-year and sire effects).

TABLE 4

Comparison of computing times between block iterative and absorption methods (in seconds).

Method	Model	I	II	III
Absorption		4 620	3 840	
Block iterative :				
First iteration		360	143	609
Other iteration*		106	90	209

* By iteration.

IV. Proposal to simplify computing on former data

Another problem is related to the size of the data set studied. Although an analysis is made every year, it is necessary to analyse data over many years so as to obtain :

- an accurate evaluation of former bulls through a combination of former and recent information,
- an estimation of HY effects independent of genetic differences between herds,
- an accurate evaluation also for bulls in progeny testing,
- an estimation of genetics trends.

With first lactations only, information from different years can be accumulated by the addition of different sets of equation, because herd-year equations absorption may be done within a year. With several lactations, absorption of cow and herd-year equations cannot be done within a year. Analysis of data from many years therefore involves processing of all data without using previous computing. This quickly becomes impossible when many years of data are available.

Another method based on an approximation already used in the French dairy sire evaluation system (POUTOUS *et al.*, 1981), splits the data into 3 groups according to 2 criteria :

- Active record : a record initiated no longer than p years in the past.
- Active cow : a cow which has at least one active record.

The 3 groups are defined as followed :

- Group 1 : cow and record both inactive,
- Group 2 : cow active and record inactive,
- Group 3 : cow and record both active.

Matrices Y , U , R , T , S , Z , E are split up, each in the same way, into 3 sub-matrices according to the 3 groups. For example :

$$R' = (R'_1, R'_2, R'_3)$$

$$\text{Let } B = V^{-1} - V^{-1}TV^{-1}$$

equation III, obtained from cow equations absorption, can be written :

$$(U'BU + \Delta) \hat{f} = U'BY$$

The off-diagonal elements in B , which correspond to one record from group 1, on the one hand, and another from group 2 or 3, on the other hand, are always zero because they are not from the same cow. Matrix B can then be written as a direct sum :

$$B = B_1 \oplus B_{23} = \begin{bmatrix} B_1 & 0 \\ 0 & B_{23} \end{bmatrix}$$

where B_1 and B_{23} are cow equation absorption matrices defined as before for B and relative to group 1 and groups 2 and 3, respectively.

We have therefore :

$$U'BU = U'_1B_1U_1 + (U'_2, U'_3) B_{23} (U'_2, U'_3)'$$

$$U'BY = U'_1B_1Y_1 + (U'_2, U'_3) B_{23} (Y'_2, Y'_3)'$$

Matrices $U'_1B_1U_1$ and $U'_1B_1Y_1$, which give all information from group 1 remain very large as they correspond to ENV (m), sire (p) and herd-year (h) effects. A reduction in size cannot be obtained by absorption of herd-year equations, as this operation should be performed on the whole set of equations.

To facilitate the computing, values of herd-year and ENV effects (h and m) relative to inactive records (group 1 and 2) are assumed to be known from previous sire evaluation runs (estimates h^* and m^*). This requires that ENV effects (m) are defined on a within year basis.

Therefore analysis can be made on the following variables :

$$\hat{Y}_1 = Y_1 - R_1 h^* - T_1 m^*$$

$$\hat{Y}_2 = Y_2 - R_2 h^* - T_2 m^*$$

$$\hat{Y}_3 = Y_3$$

with the following analysis model :

$$\hat{Y}_1 = U_1 f + Z_1 c + E_1 = S_1 p + Z_1 c + E_1$$

$$\hat{Y}_2 = U_2 f + Z_2 c + E_2 = S_2 p + Z_2 c + E_2$$

$$\hat{Y}_3 = U_3 f + Z_3 c + E_3 = R_3 h + T_3 m + S_3 p + Z_3 c + E_3$$

Information from the first group of data can then be summarized with the smaller matrix :

$$U'_1 B_1 U_1 = S'_1 B_1 S_1 \text{ which is a diagonal matrix.}$$

$$U'_1 B_1 Y_1 = S'_1 B_1 Y_1$$

Moreover, data from the second group (\bar{Y}_2) (inactive performance of active cows) only contribute to computing m_k and the sum w_k of weights w_{ki} because the model only includes the sire factor. This may also allow to reduce the size of the data set actually used for sire evaluation.

Our approximation might bias breeding values if differences between previous estimates (h^* and m^*) and exhaustive estimates (from the processing of all data) are large and not distributed at random. Bias corresponding to cow culling might not be fully prevented like after accurate application of BLUP procedures. The main risk is a bad estimation of genetic trends and therefore of differences between bulls used in different years. Extent of bias depends on the choice of the p value, i.e. the number of years during which records are considered as active. Indeed, the best is the highest possible p value according to computing facilities. This requires further studies which could involve checking of results, mainly estimated genetic trends, according to the choice of the p value or in comparison to the first lactation sire evaluation which can easily be applied to all data.

V. Conclusion

Dairy sire evaluation according to Henderson's BLUP methodology is difficult when several lactations of one and the same cow are analysed by means of a model involving not only the usual effects (sire, year, herd and cow), but also other environmental effects such as month and age of calving or parity. A rigorous application on the large dairy recording files seems to be impossible with the present computation possibilities. Thus, rather than simplifying the model of analysis which would be the only way of exactly applying the BLUP principles, this paper describes a computing algorithm allowing to partly solve the difficulties owing to 2 approximations. However, the validity of the second approximation was only partly shown. The main risk is a bad estimation of genetic trends and therefore of differences between bulls used in different years. Thus, further research is needed to evaluate the extend of this bias and to allow the computing algorithm proposed in this paper to be used.

Received October 26, 1982.

Accepted July 3, 1985.

References

- HENDERSON C.R., 1973. Sire evaluation and genetic trends. *Proceeding of the Animal Breeding and Genetics Symposium in honor of Dr. Jay L. LUSH, Blacksburg, Virginia, July 29, 1972*, 10-41, ASA-ADSA, Champaign, Illinois.
- LUSH J.L., 1931. The number of daughters necessary to prove a sire. *J. Dairy Sci.*, **14**, 209.
- POUTOUS M., BRIEND Michèle, CALOMITI S., DOAN D., FELGINES C., STEIER G., 1981. Méthode de calcul des index laitiers. Bases générales, *Bul. Tech. Ing.*, **361**, 433-446.
- SCHAEFFER L.R., 1975. *Dairy sire evaluation for milk and fat production*. 46 pp., University of Guelph. Mimeograph.
- SEARLE S.R., 1965. *Matrix Algebra for the biological sciences*. 296 pp., Wiley, New York.
- UFFORD G.R., HENDERSON C.R., VAN VLECK L.P., 1978. Derivation of computing algorithms for sire evaluation, using all lactation records and natural service sires. 46 pp., Animal Science Mineograph Series n° 39. Cornell University, Ithaca, NY.

Appendix

Data must be analysed according to the model :

$$Y = Uf + Zc + E$$

Vector f represents a group of several effects :

$$f' = (f'_1, f'_2, \dots, f'_\varphi, \dots)$$

with $U = (U_1, U_2, \dots, U_\varphi, \dots)$ the related design matrices. The definition of f'_φ is such that each row of the related design matrix U_φ has only one non zero element, also equal to 1.

$U [(kl) ;]$ represents the row of U corresponding to the 1th record of the kth cow. Vectors E and c represent random effects with :

$$\text{Var} (E) = V \sigma_e^2$$

$$\text{Var} (c) = I (\sigma_c^2/\alpha)$$

Matrix V is a diagonal matrix with

$$V [(kl) ; (kl)] = 1/w_{kl}$$

The mixed model equation is :

$$\begin{bmatrix} U'V^{-1}U & U'V^{-1}Z \\ Z'V^{-1}U & Z'V^{-1}Z + \alpha I \end{bmatrix} \begin{bmatrix} \hat{f} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} U'V^{-1}Y \\ Z'V^{-1}Y \end{bmatrix}$$

With $\Gamma = Z (Z'V^{-1}Z + \alpha I)^{-1}Z'$

$$p = U'V^{-1}U - U'V^{-1}\Gamma V^{-1}U$$

$$s = U'V^{-1}Y - U'V^{-1}\Gamma V^{-1}Y$$

absorption of cow equations gives :

$$\mathbf{p}\hat{\mathbf{f}} = \mathbf{s}.$$

\mathbf{p} and \mathbf{s} may be computed using the following results :

Let n_k the number of records of the k^{th} cow and $w_k = \sum_l w_{kl}$

J_{n_k} a matrix $(n_k \times n_k)$ whose elements are all equal to one

$$\Gamma_k = (1/(\alpha + w_k)) \bar{J}_{n_k}$$

U_k and V_k the submatrices of U and V corresponding to the k^{th} cow

We have :

$$U' = (U'_{1,}, U'_{2,}, \dots, U'_{k,}, \dots)$$

$$V = \sum_k^{\oplus} V_k$$

$$\Gamma = \sum_k^{\oplus} \Gamma_k$$

$$\mathbf{p} = \sum_k (U'_k V_k^{-1} U_k - U'_k V_k^{-1} \Gamma_k V_k^{-1} U_k)$$

Each element of \mathbf{p} can then be computed from the expression :

$$\begin{aligned} \mathbf{p} [u ; u'] &= \sum_k \{ \sum_l U [(kl) ; u] U [(kl) ; u'] (w_{kl} - w_{kl}^2 / (\alpha + w_k)) \\ &\quad - \sum_{l,l' \neq 1} U [(kl) ; u] U [(kl)'; u'] w_{kl} w_{kl'} / (\alpha + w_k) \} \end{aligned}$$

In the same way for \mathbf{s}

$$\mathbf{s} [u] = \sum_{kl} U [(kl) ; u] w_{kl} (y_{kl} - m_k)$$

where $m_k = ((Z'V^{-1}Z + \alpha I)^{-1} Z'V^{-1}Y) [k] = (\sum_l w_{kl} y_{kl}) / (\alpha + w_k)$.

Notation : \oplus and \sum^{\oplus} represent a direct sum of matrices (SEARLE, 1965).