

## BREEDING VALUE PREDICTION WITH THE ANIMAL MODEL IN THE BELGIAN BLACK AND WHITE<sup>1</sup>

P. L. LEROY, F. FARNIR

Faculté de Médecine Vétérinaire,  
Université de Liège  
45 rue des Vétérinaires, B-1070 Bruxelles.

### SUMMARY

An animal model was applied to estimated breeding values of Belgian dairy cattle. 252,793 cows limited to their 3 first lactations were evaluated. The model includes fixed environmental effects, groups, additive genetic, permanent environment and residual effects. Genetic groups were defined for unknown parents. Groups were defined by origin, sex and year of birth. Solutions were obtained by iteration without forming the equations and using a Gauss-Seidel in sequential and in parallel using LCAP approach. The speedup was 2.99 with 4 processors indicating that 90 % of the computing time was in disconnected virtual machines considered as slaves.

### INTRODUCTION

In Belgium, as in other countries, the presence of animals from different populations and non-random mating cause bias in the evaluation of dairy cattle for genetic merit. To correct for these sources of error in the estimations different approaches have been used.

Grouping introduced by Henderson (1973) to account for differences between populations and subpopulations has been used in Belgium in single trait and multiple trait additive models developed to evaluate native Black and White and Holstein Friesian sires (sire models). Grouping was by origin or by year of use within group of origin assuming common genetic variances (Leroy et al., 1978). Later, maternal grandsire effects (Quaas, Everett, McLintock (1979)) were included in the models to correct for non-random mating, especially for imported sires. These models considered relationships among males.

The animal model (AM) (Henderson 1972, Quaas and Pollak 1980, Westell and Van Vleck 1984), allows simultaneous evaluation of sires and cows for additive genetic values even for animals with no records but which are tied by relationships (Henderson 1977).

Westell and Van Vleck (1984) used the animal model in the northeast U.S. Holstein population. As indicated by Quaas (1988) this model with groups is the most complete model to date and will turn out to be the Westell-Robinson model, a model proposed independently by the two authors (Westell (1984,1987) and Robinson (1986)). An other version of the model including interaction terms has been used in the USA by Wiggins *et al.* (1988).

At the end of the 80's programming strategies were developed in different ways in order to reduce turn around time (parallel approach) and CPU time (vectorization, numerical analysis). An important reduction in CPU time has been found by Ducrocq *et al.* (1990) using special properties of the linear model. The study presented here gives the results of the breeding value predictions in Belgian dairy cattle using the animal model including groups of ancestors and permanent environmental effects. The linear model has been updated since 1987, year of the first AM cow evaluation in Belgium (Leroy and Van Vleck (1987)). The study gives also the results of the gain in computing time (elapsed time) due to LCAP.

<sup>1</sup> Research sponsored by IRSIA - IWONL

## MATERIAL AND METHODS

### Data

Data included records of 252,793 cows (482,003 305-day lactation records) taken from the recording program of the Black and White breed of Belgium. Cows were limited to their three first recorded lactations.

Cows were daughters of native Black and White, US, Canadian Holstein-Friesian and European sires (NL, F, D) put in different groups. Sires were also divided as: imported semen, tested in Belgium, in testing (> or = 75% HF and <75%HF).

The averages and standard deviations are given for each trait by lactation, in table 1.

### Methods

All animals, whether their parents were known or unknown, were included in the analysis. Sires were grouped in categories according to their origin (Native, European, USA, Canadian).

Pedigree information was picked from the pedigree file of the U.S. Northeast Dairy Records Processing Laboratory and Canadian file.

When an ancestor was unknown it was assigned to a group considered to be a representative parent according to Westell (1984,1987). When the unknown ancestor was a sire, the animal was tied to a special group (representative parent) according to its year of birth. When the unknown parent was a dam, the group was assigned according to the origin of the sire and the year of birth of the daughter to account for selection of that kind of dam. The number of groups was 32 and correspond to the different year of birth of the daughters, to generations of sires and to exporting countries.

**Table 1. Averages and standard deviations of production (milk, fat, protein, 305-day records Kilos) for cows in first, second and third lactations.**

	n	Lactation 1 252,793	Lactation 2 142,296	Lactation 3 86,914
Milk	mean	4480.9	5125.5	5602.9
	S.D.	1139.4	1232.0	1267.2
Fat	mean	179.1	206.3	224.5
	S.D.	49.0	53.3	55.4
Prot	mean	145.4	170.0	183.2
	S.D.	36.9	39.3	40.3
Fat %	mean	3.99	4.03	4.00
	S.D.	0.41	0.43	0.42
Prot %	mean	3.25	3.33	3.28
	S.D.	0.20	0.22	0.20

The complete animal model describing the record is :

$$y = X\beta + Wg + Za + Zp + e$$

where y is the vector of records,

$\beta$  is the vector of fixed herd-year effects associated with the records, age of calving, season of calving and calving interval

g is the vector of fixed group effects,

a is the vector of random additive genetic effects associated with the animals,

p is the vector of random permanent environmental effects associated with the animals,

e is the vector of residuals,

and X, Z and W are incidence matrices for the expression of  $\beta$ , g, a, and p in y with the entire row of Z (and of X) consisting of zeros when the animal has no records.

The same methodology described in Leroy and Van Vleck (1987) using the results of Quaas and Pollak (1981) and Westell and Van Vleck (1984) was applied.

#### PROGRAMMING STRATEGY

First, the data were sorted by herd because solutions are obtained by iteration starting with cows in individual herds and finishing with sires and groups in successive rounds. The situation allows a parallel approach (Leroy *et al.* 1987).

Two different approaches have been studied. A classical methodology using herds and animals within herds as blocks (Gauss-Seidel) and a parallel method using blocks of herds and cows (blocks of blocks) corresponding to animals and herds of a particular region of the country have been compared. In the second approach using more than one processor, the blocks of herds were treated separately on different virtual machines using the Master and Slaves LCAP approach (Clementi *et al.* (1988)). The ancestors and fixed effects were solved in the Master after updating the corresponding right hand sides by herd and cow equations solved previously in the slaves. That operation takes place at the end of each round of iteration.

Before the iteration process the data are prepared so that identification numbers of animals become indexes of vectors. Elements of A-inverse corresponding to sires and groups, merged with A-inverse elements separately computed for sires and groups of representative parents are sorted and written on the sire-group file containing all the elements for the sire-group iteration which is performed after the last iteration of the last herd at each round.

A-inverse elements corresponding to animals within herds (cows) accumulated together with the elements of the incidence matrix of the animals are written separately.

Permanent environmental effects are absorbed and can be obtained directly as functions of herd-year and cow solutions obtained previously in the same round of iteration.

The iteration process involves 4 vectors. These 4 vectors correspond to 2 work areas for herd-year and cow effects (array of variable length) and 2 fixed length (during execution time) vectors for sires, groups and fixed effects like age, season and calving interval. In each case, the two vectors correspond to right hand sides (RHS) which separately are adjusted throughout a round of iteration. Equations are never written and current solution vector containing RHS at the start are continuously adjusted for solutions obtained earlier in the round.

Convergence criterion is computed according to Jennings (1977).

#### RESULTS AND DISCUSSION

After 50 rounds of iteration the convergence value was comparable to previous results (Leroy and Van Vleck (1987)).

From the previous work (Leroy and Van Vleck (1987)), it has been found that, increasing within a round the number of iterations per herd and using a relaxation factor for herd-year and cow equation was not efficient during the iteration process. With this data set the previous "disturbance" did not appear but only 32 (instead of 172 previously) group equations were present.

Averages of cow additive values, permanent environmental effects, group effects and sire additive values solutions are given in table 2.

A positive estimated genetic trend for milk (58.57), fat (0.059) and protein content (0.047) has been observed in the Province of Liège.

Central Processing Unit time (CPU) for an IBM 3090 (Vector not used) for 1 complete round of iteration was 2.7 minutes. The CPU time could be reduced for routine runs by the vector and by using the memory of the IBM VMXA. Since memory size needed is not a limiting factor with this material the iteration system could be easily developed for microcomputers like the IBM PS/2 series.

On unloaded system specific for benchmarks using up to 4 CPUs, the speedup, defined as the elapsed time gain due to 2 and 4 processors, was 2.99 for 4 processors and 1.48 for 2 processors. According to

Amdahl's law (Clementi 1989) the Fraction sequential was estimated to 0.1 indicating that 90% of the work was done in the slaves (parallel) and 10% in the master (sequential).

A lot of work is to be done to incorporate pedigree in file and to retrieve data from all the foreign countries exporting semen. To achieve that objective, we think that the Interbull group should indicate the informations to create a pedigree file of well known sires used all over the world.

**Table 2. Average solutions (Kg) and standard deviations for cow additive genetic values, groups and sire additive genetic values at the end of the iteration process.**

	n	Milk Average (S.D.)	Fat% Average (S.D.)	Prot.% Average (S.D.)
Groups (n=32) and ancestors (n=983)		99.56 (190.93)	0.088 (0.187)	0.071 (0.135)
Cows additive genetic values	252,793	-0.61 (558.56)	-0.006 (0.25)	-0.008 (0.14)

#### References

- CLEMENTI, E. 1989 Modern Techniques in Computational Chemistry (MOTEC). Edited by IBM.
- CLEMENTI, E., LOGAN, D., and SAARINEN, J. 1988. LCAP/3090 : Parallel processing for large-scale scientific and engineering problems. IBM Coporation ISSN 18-8670.
- DUCROCQ, V., BOICHARD, P., BONAITI, B., BARBA, A. and BRIEND, M. (1990). Pseudo-absorption strategy for solving animal equations for large data files. *J. Dairy Sci.* (in press).
- HENDERSON, C.R. 1972. Sire evaluation and genetic trends. *Proc. Anim. and Genet. Symp. in Honour of Dr. Jay L. Lush ASAS-ADSA, Champaign, IL.*
- HENDERSON, C.R. 1975. Rapid method for computing the inverse of a relationship matrix. *J. Dairy Sci.*, 58: 1727
- HENDERSON, C.R. 1977. Best linear unbiased prediction of breeding values not in the model for records. *J. Dairy Sci.*, 66: 1503
- LEROY, P., VAN VLECK, L.D., Scannapieco M. and Verdier, F. 1987. Solving the animal model with the Supercomputer IBM 3090 (Preliminary Report IBM).
- LEROY, P.L., VAN VLECK, L.D. 1987. Estimation of breeding values of dairy cattle in an importing country using an animal model. In 38th Annual Meeting of the EAAP, Lisbon, Portugal, September 28-October 1, Vol I, Commission of Animal Genetics, Paper N° G2.6 2p.
- QUAAS, R.L., POLLAK, E.J. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.*, 51:1277-1287
- QUAAS, R.L.1988. Additive genetic model with groups and relationships. *J. Dairy Sci.*, 71: 1338-1345.
- POLLAK, E.J., QUAAS, R.L., 1981. Modified equations for sire models with groups. *J. Dairy Sci.*, 64: 1868-1872
- ROBINSON, G.K., 1986. Group effects and computing strategies for models for estimating breeding values. *J. Dairy Sci.*, 69: 3106.
- SCHAEFFER, L.R., KENNEDY, B.W. 1986. Computing strategies for solving the mixed model equations. *J. Dairy Sci.*, 69: 575-579
- WESTELL, R.A. , 1984. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. Ph. D. Thesis, Cornell University, Ithaca,N.Y.
- WESTELL, R.A., VAN VLECK, L.D., 1987. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. *J. Dairy Sci.*, 70: 1006
- WIGGANS G.R., MISZTAL I., VAN VLECK, L.D. 1988. 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.
- Acknowledgments** The authors express their appreciation to Dr E. Clementi, Dr D. Logan and A. Trannoy for their help in the development of computing strategies using the LCAP on IBM 3090 in IBM Kingston.