

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Papers and Publications in Animal
Science

Animal Science Department

January 1988

Animal Model Evaluation of Ayrshire Milk Yield with All Lactations, Herd-Sire Interaction, and Groups Based on Unknown Parents

G. R. Wiggans

Agricultural Research Service, United States Department of Agriculture, Beltsville, MD

I. Misztal

University of Illinois, Urbana

L. Dale Van Vleck

University of Nebraska-Lincoln, dvan-vleck1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/animalscifacpub>



Part of the [Animal Sciences Commons](#)

Wiggans, G. R.; Misztal, I.; and Van Vleck, L. Dale, "Animal Model Evaluation of Ayrshire Milk Yield with All Lactations, Herd-Sire Interaction, and Groups Based on Unknown Parents" (1988). *Faculty Papers and Publications in Animal Science*. 305.

<https://digitalcommons.unl.edu/animalscifacpub/305>

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Papers and Publications in Animal Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Animal Model Evaluation of Ayrshire Milk Yield with All Lactations, Herd-Sire Interaction, and Groups Based on Unknown Parents¹

G. R. WIGGANS

Animal Improvement Programs Laboratory
Agricultural Research Service
United States Department of Agriculture
Beltsville, MD 20705

I. MISZTAL

Department of Animal Sciences
University of Illinois
Urbana 61801

L. D. VAN VLECK

Department of Animal Science
Cornell University
Ithaca, NY 14853

ABSTRACT

An animal model was applied to predict genetic merit for Ayrshire milk yield. The model included fixed herd-year-season (32,287) and random herd-sire interaction (32,159), permanent environment, animal, and residual effects. Animals evaluated included 119,541 cows with 301,799 records, 5762 sires, and 11,893 dams without records. Genetic groups (36) were defined for unknown parents and parents not contributing ties or records. Groups were defined by sex of parent and by birth year and sex of animal with unknown parent. Evaluations included combinations of these group effects derived from tracing each path in pedigree back to an unknown parent group. Iteration was by Gauss-Seidel for herd-year-season, permanent environment, and herd-sire interaction effects and by second-order Jacobi for animal and genetic group effects. Iteration was conducted without forming mixed model equations; instead one copy of data sorted by herd and sire was read each

round. About 3.5 s of central processing unit time on a Cray X-MP/48 was required per round for the complete model; herd-sire interaction contributed .7 s and later lactations, 1.6 s. Large memory requirements were reduced by evaluating most animals in groups of herds including up to 2500 cows. Information for animals with progeny outside their herd group remained in memory throughout iteration. Genetic evaluations by an animal model that includes factors presently in the national evaluation system can be computed with present computer resources. Application to major dairy cattle breeds appears possible.

INTRODUCTION

Wiggans and Misztal (26) found that an animal model was computationally feasible when applied to first lactations. A review by Cassell et al. (1) and a study by Powell and Norman (13) found that records from later lactations provided more complete information on lifetime performance than did those from first lactations. Later lactation records routinely are included in USDA's Modified Contemporary Comparison (MCC) evaluations. The first goal of this study was to allow for inclusion of later lactation records by adding a permanent environment effect to the model and to determine additional computer resources required.

Genetic grouping is important in achieving accurate evaluations by defining different

Received July 6, 1987.

Accepted November 9, 1987.

¹Supported in part by US-Israel Binational Agricultural Research and Development Project Number US-805-84; supercomputer support by National Center for Supercomputer Applications, Champaign, IL, and Cornell Theory Center, Ithaca, NY.

populations from which animals arise (12). The relationship matrix has reduced need for groups by accounting for most of genetic trend. However, groups still are necessary to represent differences in genetic values of unknown parents that possibly may be from different populations. A grouping strategy was proposed by Thompson (21) and discussed by Robinson (19), Westell (23), Westell and Van Vleck (25), and Westell et al. (24), in which genetic group effect for an animal is derived from a combination of group effects for its ancestors. Each of these underlying group effects is the average of genetic values of unknown parents grouped, for example, by sex of parent and birth year and sex of the animal with the unknown parent.

This approach to grouping was made computationally feasible by Westell (23), Westell and Van Vleck (25), and Westell et al. (24), who discovered simple rules for constructing group equations and group contributions to animal equations. Their procedure applies to mixed model equations transformed so that animal solutions include appropriate group effects (18). These rules generate the same contributions as those for constructing inverse of the numerator relationship matrix (A^{-1}) (5, 16) for the animal and known parents. Quaas (17) also has provided an alternative derivation of these rules. The second goal of this study was to apply the Westell grouping procedure in conjunction with iteration on data (8, 26) so that creation of the coefficient matrix is not required.

Environmental correlation among a sire's daughters in the same herd has been identified as an important factor in sire evaluation (7, 10). This correlation can be accounted for by including a herd-sire interaction in the model (4) as in MCC evaluations (2). Including this interaction has been credited with causing initial evaluations of bulls proven through natural service to be adequate predictors of the bulls' eventual AI evaluations (11). The third goal of this study was to determine if a herd-sire interaction could be included in the model without an unacceptable increase in processing time per round of iteration or in number of iterations required.

The procedure of Wiggans and Misztal (26) required four values in memory for each animal during iteration. This memory requirement

might make the animal model impractical for large populations. The final goal of this study was to discover ways to reduce this requirement by storing intermediate results on disk without greatly increasing processing time.

The overall objective of this study was to determine if animal model evaluations that include later records, herd-sire interaction, and grouping of unknown parents are computationally practical in both processing time and memory limitations.

MATERIALS AND METHODS

Data

The Ayrshire breed was selected for this study because of its small population size, which allowed for realistic testing of procedures on data for a complete breed. Data were organized into records that contained all lactations for a cow in a herd. Lactations in other herds were in different records. Herd-cow records (212,801) available for July 1986 USDA-DHIA genetic evaluations were the data source for this study as well as that of Wiggans and Misztal (26). Records of cows born before 1958 (23,863) were excluded because data contained only calving dates in 1960 or later; therefore, cows born before 1958 were unlikely to have first lactations included. Also excluded were records of cows that had no reported first lactation (61,639) or no herdmates (7758). A first lactation was required to avoid bias due to selection. This requirement may not be acceptable for national evaluations. Resulting data consisted of 301,799 lactation records for 119,541 cows (in 3196 herds) that were daughters of 7895 sires. Of these sires, 5637 had at least two daughters. There were 125 sires without daughters with records but with at least two progeny (sons or daughters without records themselves but with daughters with records); similarly, 11,893 cows were dams of at least two animals but had no records themselves.

Groups of unknown parents were defined based on sex of parent and birth year and sex of animal with unknown parent. Parents with no records and only one offspring also were treated as unknown parents to reduce number of animals evaluated. These animals did not contribute any direct ties although some would have provided ties through a grandparent.

However, little loss of information occurred by combining them with unknown parents. Of the 5762 sires evaluated, 3353 had unknown dams and 2147, unknown sires; 31,010 cows had unknown dams and 2258, unknown sires. All 2258 of these sires of cows actually were known, but they had no other progeny. Cows with unknown parents remained in the analysis because they may have been dams and were herdmates of other animals. They also contributed to estimation of unknown parent effects. For unknown parents of sires, 5 groups each for sires and dams were formed at about 5-yr intervals; for unknown parents of cows, 13 groups each were formed at 2-yr intervals. These 36 groups were based on birth year of the animal rather than the parent and were fixed effects. [See Appendix for development of group equations with Westell grouping procedure (23, 24, 25).]

Model

The complete model with all effects included (model 1) was:

$$y_{ijkl} = h_{ij} + c_{ik} + u_{kl} + p_{ikl} + e_{ijkl}$$

where y_{ijkl} = milk yield of daughter l of sire k (cow kl) in herd i in year-season j ; h_{ij} = fixed effect (total of 38,287) of herd i in year-season j (seasons were December through April and May through November); c_{ik} = random effect of herd-sire group ik , with a separate effect assigned for each animal with an unknown sire (total of 32,159); u_{kl} = breeding value of cow kl , a random additive genetic effect (a_1 in Appendix with variance $A_{11}\sigma_a^2$ where A_{11} is portion of the numerator relationship matrix relating animals being evaluated and σ_a^2 is additive genetic variance) plus contributions from genetic groups ($A_{10}Qg$ in Appendix); p_{ikl} = random permanent environmental effect of cow kl in herd i ; and e_{ijkl} = random residual associated with each record. Variance components for c , u , p , and e scaled to a phenotypic variance of 1 were $\sigma_c^2 = .14$, $\sigma_a^2 = .2$, $\sigma_p^2 = .16$, and $\sigma_e^2 = .5$, which resulted in variance ratios (k) of $k_c = 3.57$, $k_a = 2.5$, and $k_p = 3.125$. Later records of cows that changed herds were not included, although this model would allow for them by predicting a separate p for each herd.

This model is equivalent to a model with explicit groups. Every predicted breeding value includes a combination of group effects and is not unique because a dependency exists between herd-year-season and group effects.

Three reduced models also were applied to the data. Model 2 did not include c . Its variance was assigned to p ; therefore, $\sigma_p^2 = .3$ retaining a repeatability of .5. Model 3 was model 1 but with only first lactation records included; variance of p was assigned to e ($\sigma_e^2 = .64$). Model 4 was model 2 but for first lactation records only; variances of c and p were assigned to e ($\sigma_e^2 = .8$).

Computing Strategy

Mixed model equations were solved iteratively by Gauss-Seidel for h , c , and p and by second-order Jacobi for animal and group effects (8). Solutions were obtained by iteration on the data as follows. Data were sorted by herd and sire so that c could be calculated one sire at a time. Data for each herd were read once for each round but processed once each for h , p , c , and u . This procedure was necessary for Gauss-Seidel iteration so that new solutions for one effect were available when computing solutions for the next effect during the same round.

First, h was estimated by averaging adjusted right-hand sides (ARHS), which were yields adjusted for other effects:

$$h_{ij} = \frac{\sum_{kl} (y_{ijkl} - c_{ik} - u_{kl} - p_{ikl})}{n_{ij..}}$$

where r = number of iteration round and $n_{ij..}$ = number of cows in herd i in year-season j . Because data were sorted by herd and sire only, contributions to ARHS and diagonals for all year-seasons in a herd were accumulated, and new solutions for each year-season were computed at the end of the herd. This approach eliminated a requirement to have a copy of the data sorted by herd-year-season.

Second, p was estimated similarly; however, the divisor included the appropriate variance ratio:

$$p_{ikl} = \frac{\sum_j (y_{ijkl} - h_{ij} - c_{ik} - u_{kl})}{(n_{i.kl} + k_p)}$$

where $n_{i,kl}$ = number of lactations of cow kl in herd i .

Third, c was estimated as:

$$c_{ik} = \frac{\sum_j (y_{ijkl} - h_{ij} - u_{kl} - p_{ikl})}{(n_{i,k} + k_c)} \quad (r) \quad (r-1) \quad (r)$$

where $n_{i,k}$ = number of daughters of sire k in herd i .

Last, ARHS for animal equations were accumulated. Contribution to ARHS from yield records for animal kl was:

$$\sum_j (y_{ijkl} - h_{ij} - c_{ik} - p_{ikl}) \quad (r) \quad (r) \quad (r)$$

Contributions from W_{11} , a matrix analogous to A^{-1} (see Appendix), to ARHS for animal and group equations were:

$$\text{cow (ARHS}_{kl}): -D_{kl}k_a \begin{pmatrix} (r-1) \\ (-.5 u_s - .5 u_d) \end{pmatrix} \quad (r-1)$$

$$\text{sire (ARHS}_s): -D_{kl}k_a \begin{pmatrix} (r-1) \\ (-.5 u_{kl} + .25 u_d) \end{pmatrix} \quad (r-1)$$

$$\text{dam (ARHS}_d): -D_{kl}k_a \begin{pmatrix} (r-1) \\ (-.5 u_{kl} + .25 u_s) \end{pmatrix} \quad (r-1)$$

where s = sire of cow kl , d = dam, and D_{kl} = 2, 4/3, or 1 depending on whether neither, one, or both parents of cow kl are unknown. The same contributions were calculated for parents without records.

Diagonal elements of the coefficient matrix for group and animal equations were accumulated before iteration was started and read each round as needed. Diagonals (diag) were the only part of the coefficient matrix explicitly formed and consisted of number of records ($n_{i,kl}$) plus contributions from W_{11} of:

$$\begin{array}{ll} \text{cow } (\text{diag}_{kl}): & D_{kl}k_a \\ \text{sire } (\text{diag}_s): & .25D_{kl}k_a \\ \text{dam } (\text{diag}_d): & .25D_{kl}k_a \end{array}$$

If a parent was unknown, ARHS and diag contributions were made to the appropriate unknown parent group. Diagonals were computed similarly for parents without records.

Groups of herds were defined to contain similar numbers of animals. These groups assisted in minimizing computer memory requirements. At the end of a group of herds, solutions for animals with no progeny outside those herds and, at the end of a round, solutions for all other animals were computed as:

$$u_{kl} = f(u_{kl}^{(r-1)} - u_{kl}^{(r-2)}) + (\text{ARHS}_{kl}/\text{diag}_{kl}) \quad (r) \quad (r-1) \quad (r-2) \quad (r)$$

where a relaxation factor (f) of .85 was used. At the end of a round, group solutions were adjusted to average 0. Only group effects were averaged, but the adjustment applies to all animal solutions because group effects are included in all animal solutions. This constraint was imposed to speed convergence by maintaining a base so that solutions could not drift from round to round.

Convergence (C_p) was measured as sum of squared differences in u_{kl} by round divided by sum of u_{kl}^2 from current round:

$$C_p = \frac{\sum_{kl} (u_{kl}^{(r)} - u_{kl}^{(r-1)})^2}{\sum_{kl} u_{kl}^2} \quad (r) \quad (r) \quad (r-1) \quad (r)$$

Accuracy (C_t) was defined as sum of squared differences between solutions at convergence and current round solutions divided by sum of squared converged solutions. The C_p criteria substantially underestimates error but may be used to estimate C_t indirectly (9):

$$\text{Est } C_t = \frac{\sum_{kl} \left[\frac{2n}{(r-n) \quad (r)} \right]^2}{\sum_{kl} C_p} \quad (r-n) \quad (r) \quad C_p$$

where r = number of current round of iteration and n = number of rounds between current and preceding iteration ($n \geq 1$). Estimates of C_t were calculated for $r \geq 40$.

Memory Saving Strategy

In the study of Wiggans and Misztal (26), four values (current and previous solutions, ARHS, and diagonals) remained in memory for each animal throughout iteration. With all animals in memory, progeny encountered anywhere could contribute to their parents, and new solutions could be computed at the end of a round from values residing in memory. Ways of reducing this memory requirement were developed so that the animal model could be applied to larger populations. Many animals, particularly females, do not have progeny outside their own herd, and many of those that do have them in nearby herds. Therefore, herds with consecutive herd codes were combined into superherd groups of up to 2500 cows. The value 2500 was chosen to minimize memory by finding the point at which reduction in number of animals with outside ties equaled increase in size of herd group. This value would vary with population size; at a minimum, it must be as large as the number of cows in the largest herd.

The superherd group was the unit for computing solutions for animals with no progeny outside the group (local animals). Computing these solutions at the end of a group rather than at the end of a round allowed for dramatic reduction in memory requirements because local animals reused memory. For data in this study, number of animals with ties outside local herd group (tie animals) was 14,993, only 10.6% of total.

Memory requirements for tie animals also were reduced from those of Wiggans and Misztal (26) by retaining in memory only solutions from the previous round and ARHS. Diagonals and solutions from two rounds earlier for tie animals were read and processed in blocks. This method was implemented with three files of solutions, one each for the previous two rounds and one for those computed in the current round. For c and p solutions, only files of previous and new solutions were required as Gauss-Seidel iteration does not require solutions from two previous rounds. Solutions for h were not saved as they were computed first each round; therefore, new solutions were available for processing other effects.

Trend

Means of evaluations by birth year were computed separately for cows and bulls at various iterations to assess effect of iteration on estimates of trend.

RESULTS

Table 1 contains indicators of convergence for the complete model (model 1). By 150 rounds, C_p had dropped to 3.11×10^{-10} and C_t was 1.3×10^{-7} . After 50 rounds, solutions changed little as indicated by high correlations between solutions for a given round and those for round 150. Average change also was small after 50 rounds. Effect of iteration on estimates of genetic trend can be seen by the increasing values from 10 to 50 rounds. Evaluations by birth year are in Figure 1 for 1958 through 1984. Change in slope in Figure 1 was contrary to findings of Van Vleck et al. (22), who found similar trend at 10, 20, and 30 rounds for Holstein data. Their trend estimates probably were similar because they started with previous evaluations. In these evaluations, all solutions started at 0. Trend after 1970 changed little after 30 rounds; however, before 1970, average predicted breeding values after 30 rounds of iteration were substantially higher than those after 150 rounds. Trend reported in Table 1 was from means for 1968 and later because trend was nearly linear from 1968 on. These trend estimates were larger than those reported by Powell et al. (15).

Table 2 shows a comparison of the four models processed on a Cray X-MP/48. Correlations between sets of solutions from the various models were large and increased as the model became more similar to the complete model (model 1) and included the same number of parities. Central processing unit time for model 4 (first lactations only without c) was slightly greater than that for the model in Wiggans and Misztal (26), probably because of steps involved in reducing memory requirement. Estimation of c (models 1 and 3) added about .7 s regardless of how many parities were included. Including later lactations (models 1 and 2) added about 1.6 s, which more than doubled the time required for a model without c (model 4). Time advantage of models with only first lactations (models 3 and 4) was less

TABLE 1. Convergence indicators and genetic trend¹ in females at various rounds of iteration for the complete model for milk yield.²

Rounds of iteration	C_p^3	Correlation with C_p at 150 rounds	Average absolute change from evaluations at 150 rounds	Linear trend in breeding value for milk yield	Estimated C_t^4
			(kg)	(kg/yr)	
10	3.89×10^{-2}	.800	194.4	12.8	...
20	5.69×10^{-3}	.934	118.4	26.7	...
30	1.38×10^{-3}	.987	56.1	32.9	...
40	3.19×10^{-4}	.998	19.6	35.2	6.0×10^{-2}
50	6.28×10^{-5}	1.000	5.0	36.8	9.5×10^{-3}
75	1.70×10^{-6}	1.000	2.8	37.4	3.2×10^{-4}
100	8.91×10^{-8}	1.000	.5	37.2	2.6×10^{-5}
125	3.64×10^{-9}	1.000	.0	37.1	8.9×10^{-7}
150	3.11×10^{-10}			37.1	1.3×10^{-7}

¹ From 1968 through 1984 birth years.

² Model includes herd-year-season, herd-sire interaction, genetic, and permanent environmental effects.

³ C_p is convergence measured as sum of squared differences in solutions between rounds divided by sum of squared solutions from current round.

⁴ C_t is accuracy defined as sum of squared differences between true and current round solutions divided by sum of squared true solutions.

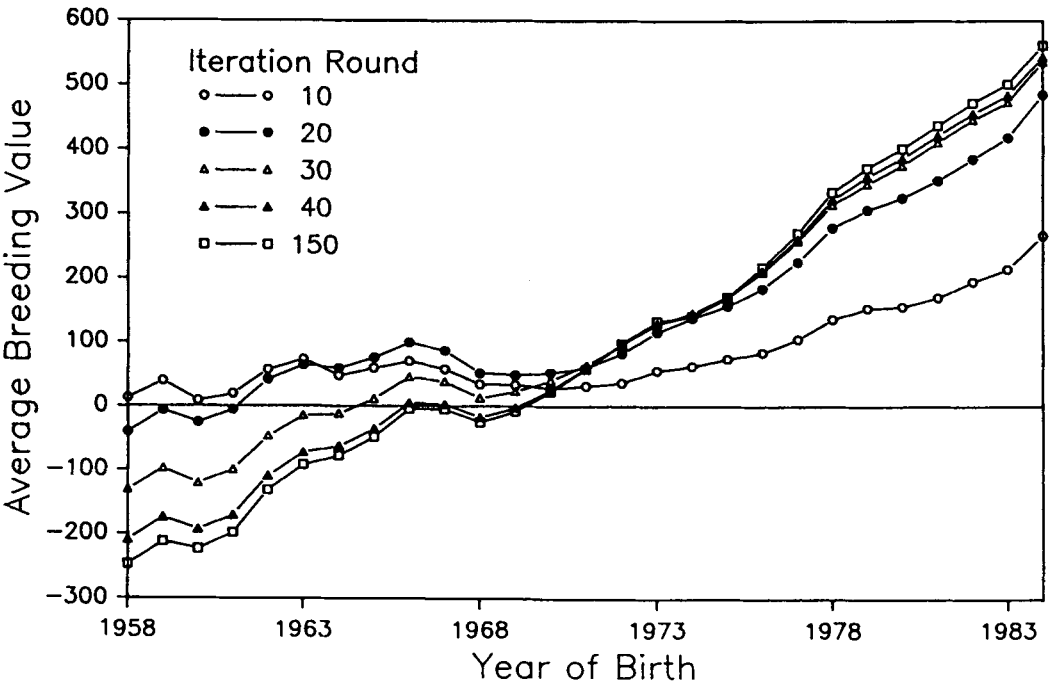


Figure 1. Average breeding values for milk yield by birth year for 10, 20, 30, 40, and 150 rounds of iteration.

TABLE 2. Comparison of evaluations from four animal models.

Model	Parities included	Effects excluded ¹	Correlation with model 1	Average absolute change from model 1 evaluations	Central processing unit time per round	Linear trend in breeding value for milk yield ²
1	All ³	None			3.5	37.1
2	All	c	.98	.61.1	2.8	40.3
3	First ⁴	p	.92	117.2	1.9	38.4
4	First	c, p	.91	130.2	1.2	39.4

¹ Complete model (model 1) includes herd-year-season (h), herd-sire interaction (c), genetic (u), and permanent environmental (p) effects.

² From 1968 through 1984 birth years.

³ At 150 rounds of iteration.

⁴ At 250 rounds of iteration.

than round time suggests because more rounds were required to reach the same value of C_p . All models produced similar estimates of genetic trend; those with c (models 1 and 3) had slightly lower estimates.

DISCUSSION

This study demonstrates that an animal model incorporating desirable aspects of present sire models is practical. Comparison of models provides information on cost of including various factors. Cost of including later records is not as great as single record times suggest because rate of convergence is more rapid. The most complete model is preferred if it is within computational capabilities.

Several areas require further research before genetic evaluations can be computed routinely with this method. One area is assignment of year-seasons. In MCC, comparisons are within parity group; first lactations generally are compared with other first lactations and later lactations with later lactations. This practice makes evaluations less sensitive to inaccuracies in age correction. Although separate year-seasons could be assigned for first and later lactations, this would result in more animals without contemporaries. Another concern is that the two seasons are fixed. In MCC, an animal is compared with animals calving in a 5-mo period centered on calving month, which results in 12 overlapping seasons per year.

Registry status also has been shown to affect environment of an animal (14). Flexible rules for assignment of year-season could maintain contemporaries in small herds but define more categories in larger herds with lactation group, 2-mo seasons, and registry status considered.

Another possibility is to evaluate fat yield at the same time as milk yield in single trait evaluations that are computed concurrently. This would save some processing compared with separate computer runs but would require additional memory for solutions and ARHS for fat.

Cows without first lactation records were excluded from this analysis to avoid bias due to selection. However, genetic estimates for all animals are desirable in a routine evaluation system. These cows could be excluded from initial evaluations but initial results used to compute evaluations for them separately. This approach would prevent such cows from affecting evaluations of other animals but provide genetic estimates for them.

Assignment of animals to unknown parent groups should be revised to allow for ties through two generations. In this study, such ties were not considered. For example, if a sire had only one daughter, the sire would have been grouped as an unknown parent. However, this sire might have had siblings that had progeny with records. These "cousin" records could have contributed information to the daughter's evaluation.

In multiple lactation evaluations, later records of cows that change herds are of interest both for added accuracy and as herd-mates. Although the method in this study could accommodate these records, a separate permanent environment effect would have to be estimated for each herd. This requirement is necessary to allow for processing of permanent environment effect by herd. Records in another herd outside a cow's herd group would cause her to be a tie animal, thus increasing storage required. A cow's contributions due to relationships are computed only once. Therefore, they should not be computed again when processing her records in later herds.

Accuracy of evaluation of individual animals has not been estimated. Such estimates will be required in routine evaluations. Accuracy should reflect contributions from progeny and parents and the animal's own records. Procedures that use diagonal elements after absorption have been developed for sire models (3, 20). Absorption was not done in this study; however, some absorption may be required to estimate accuracy.

No updating strategy has been proposed. In a semiannual update cycle, only a small percentage of animals with records add new information, but all data are reprocessed. This reprocessing allows information from descendants to feed back to ancestors. A method that directs iteration to those solutions that are subject to significant change while allowing readjustment of all solutions occasionally might allow for reduction in processing expense while not compromising accuracy. Solutions from previously computed evaluations may be sufficiently good as starting values that relatively few iterations are required. Usefulness of previous solutions as starting values may be improved by adding pedigree estimates for new animals and estimates based on trend for new groups.

Evaluations resulting from an animal model are breeding values in contrast to present sire models that estimate transmitting ability. Reporting evaluations as breeding values is desirable because it is consistent with the model and estimates of genetic trend are computed from averages of evaluations by year. The sometimes overlooked need to multiply trend in transmitting abilities by 2 is eliminated. Real

producing ability is $u + c + p$ in this model. If evaluations are expressed as breeding values, analysis of environmental effects by comparison of breeding value with real producing ability will be simplified.

No particular base was selected for these evaluations. Group effects determine level of all other animal evaluations because every evaluation includes a weighted combination of group effects, for which sum of the weights is 1. A fixed base could be maintained by adjusting a fixed combination of unknown parent effects to the same value for each computation of evaluations.

CONCLUSIONS

Results with Ayrshire data indicate that an animal model is practical and that estimates of genetic trend from resulting evaluations are reasonable. Genetic groups can replace unknown parents, which results in groups that complement the relationship matrix. The reduction in memory achieved and short computing times suggest that the complete model may be feasible for the national Holstein population.

ACKNOWLEDGMENTS

Support provided by the Department of Animal Sciences, University of Illinois, for a sabbatical stay; assistance with literature search and manuscript preparation by S. M. Hubbard; assistance with applying genetic grouping procedures by S. Hudson; and many helpful suggestions by R. Fernando are acknowledged.

REFERENCES

- 1 Cassell, B. G., and B. T. McDaniel. 1983. Use of later records in dairy sire evaluation: a review. *J. Dairy Sci.* 66:1.
- 2 Dickinson, F. N., H. D. Norman, R. L. Powell, L. G. Waite, and B. T. McDaniel. 1976. Procedures used to calculate the USDA-DHIA Modified Contemporary Comparison. Page 18 in *USDA Prod. Res. Rep. No. 165*, Washington, DC.
- 3 Greenhalgh, S. A., R. L. Quaas, and L. D. Van Vleck. 1986. Approximating prediction error variances for multiple trait sire evaluations. *J. Dairy Sci.* 69:2877.
- 4 Henderson, C. R. 1973. Sire evaluation and genetic trends. Page 10 in *Proc. Anim. Breed. Genet. Symp.* in Honor of Dr. J. L. Lush, Am. Soc. Anim.

- Sci. and Am. Dairy Sci. Assoc., Champaign, IL.
- 5 Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69.
 - 6 Henderson, C. R. 1976. Best linear unbiased prediction of breeding values not in the model for records. *J. Dairy Sci.* 60:783.
 - 7 Meyer, K. 1987. Variances due to sire x herd interactions and common environmental covariances between paternal half sisters for first lactation production. *J. Dairy Sci.* 70(Suppl. 1):126. (Abstr.)
 - 8 Misztal, I., and D. Gianola. 1987. Indirect solution of mixed model equations. *J. Dairy Sci.* 70:716.
 - 9 Misztal, I., D. Gianola, and L. R. Schaeffer. 1987. Extrapolation and convergence criteria with Jacobi and Gauss-Seidel iteration in animal models. *J. Dairy Sci.* 70:2577.
 - 10 Norman, H. D. 1974. Factors that should be considered in a national sire summary model. *J. Dairy Sci.* 57:955.
 - 11 Norman, H. D., R. L. Powell, and J. R. Wright. 1985. Changes of evaluation for natural-service sampled bulls brought into artificial insemination. *J. Dairy Sci.* 68:1513.
 - 12 Pollak, E. J., and R. L. Quaas. 1983. Definition of group effects in sire evaluation models. *J. Dairy Sci.* 66:1503.
 - 13 Powell, R. L., and H. D. Norman. 1981. Different lactations for estimating genetic merit of dairy cows. *J. Dairy Sci.* 64:321.
 - 14 Powell, R. L., and H. D. Norman. 1986. Genetic and environmental differences between registered and grade Holstein cows. *J. Dairy Sci.* 69:2897.
 - 15 Powell, R. L., H. D. Norman, and F. N. Dickinson. 1977. Trends in breeding value and production. *J. Dairy Sci.* 60:1316.
 - 16 Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32:949.
 - 17 Quaas, R. L. 1987. Notes on animal models with groups and relationships. Page 216 in *Genet. Res.* 1986-1987 Rep. to Eastern AI Coop., Inc. Dept. Anim. Sci., Cornell Univ., Ithaca, NY.
 - 18 Quaas, R. L., and E. J. Pollak. 1981. Modified equations for sire models with groups. *J. Dairy Sci.* 64:1868.
 - 19 Robinson, G. K. 1986. Group effects and computing strategies for models for estimating breeding values. *J. Dairy Sci.* 69:3106.
 - 20 Robinson, G. K., and L. P. Jones. 1987. Approximations for prediction error variances. *J. Dairy Sci.* 70:1623.
 - 21 Thompson, R. 1979. Sire evaluation. *Biometrics* 35:339.
 - 22 Van Vleck, L. D., R. A. Westell, and J. C. Schneider. 1986. Genetic change in milk yield estimated from simultaneous genetic evaluation of bulls and cows. *J. Dairy Sci.* 69:2963.
 - 23 Westell, R. A. 1984. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. Ph.D. Thesis, Cornell Univ., Ithaca, NY.
 - 24 Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic groups in an animal model. *J. Dairy Sci.* 71:(in press).
 - 25 Westell, R. A., and L. D. Van Vleck. 1987. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. *J. Dairy Sci.* 70:1006.
 - 26 Wiggans, G. R., and I. Misztal. 1987. Supercomputer for animal model evaluation of Ayrshire milk yield. *J. Dairy Sci.* 70:1906.

APPENDIX

The following description of the grouping procedure for this study is based on Westell's development (23, 24, 25) of a grouping procedure in which genetic group effect on an animal's record is a weighted combination of genetic group effects for animals that are unknown parents of animals being evaluated or known parents without records or ties to other animals, e.g., animals with only one offspring. Determination of which unknown parents to include in a genetic group may be based on time period, path of selection, region, origin, etc. Group effects for other animals are sums of fractions of these group effects of unknown parents; i.e., for animal j , group effect is $\sum_r q_{jr}g_r$ where q_{jr} is fraction of animal j 's genes expected from ancestors in genetic group r ($\sum_r q_{jr} = 1$) and g_r is mean genetic value of unknown parents assigned to genetic group r . For example, if both parents of animal j are unknown, $q_{jr} = .5$ for each parent; if all four grandparents are unknown, $q_{jr} = .25$ for each grandparent. If sire and maternal grandparents are unknown, $q_{jr} = .5$ for sire and .25 for each maternal grandparent. Because each path of an animal's pedigree ends in an unknown ancestor, q_{jr} corresponding to that path is .5 to the power of the number of steps from the animal to the unknown ancestor in that path. For the unknown animals, number of steps is 0 and $q_{jr} = 1$.

In matrix notation, the model for an animal with records is:

$$y = Hh + Cc + Zp + Za + ZA_{10}Qg + e$$

where y is a vector of records; h is a vector of fixed herd-year-season effects; c is a vector of random herd-sire interaction effects; p is a vector of random permanent environmental effects; a is a vector of random additive genetic values; H , C , and Z are incidence matrices that associate elements of h , c , p , and a with those of y ; A is the numerator relationship matrix for all animals partitioned into A_{00} (relationships among unknown parents; $A_{00} = I$, an identity matrix), A_{10} (relationships among identified animals and unknown parents; $A_{10} = A'_{01}$), and A_{11} (relationships among identified animals); Q is a matrix with a 1 in each row that assigns an unknown parent to a group and 0's elsewhere; g is the vector of mean genetic effects associated with groups of unknown parents; and e is a vector of residual effects associated with y . Thus, $ZA_{10}Qg$ is contribution of genetic effects of groups of unknown parents to records.

Total genetic merit (u) for animal j is:

$$u_j = a_j + \sum_r q_{jr} g_r$$

where a_j is additive genetic value of animal j expressed as a deviation from the function of group effects. In matrix notation, the vector of genetic merit of unknown animals (u_0) is:

$$u_0 = a_0 + Qg$$

where a_0 is a vector of additive genetic values of unknown animals. The vector of genetic merits of identified animals (u_1) is:

$$u_1 = a_1 + A_{10} Qg$$

The mixed model equations augmented to include unknown parents and animals without records (6) are:

$$\begin{bmatrix} H'H & H'C & H'Z & H'Z & 0 & H'ZA_{10}Q \\ C'H & C'C + Ik_C & C'Z & C'Z & 0 & C'ZA_{10}Q \\ Z'H & Z'C & Z'Z + Ik_p & Z'Z & 0 & Z'ZA_{10}Q \\ Z'H & Z'C & Z'Z & Z'Z + A^{11}k_a & A^{10}k_a & Z'ZA_{10}Q \\ 0 & 0 & 0 & A^{01}k_a & A^{00}k_a & 0 \\ Q'A_{01}Z'H & Q'A_{01}Z'C & Q'A_{01}Z'Z & Q'A_{01}Z'Z & 0 & Q'A_{01}Z'ZA_{10}Q \end{bmatrix} \begin{bmatrix} \hat{h} \\ c \\ p \\ a_1 \\ a_0 \\ g \end{bmatrix} = \begin{bmatrix} H'y \\ C'y \\ Z'y \\ Z'y \\ 0 \\ Q'A_{01}Z'y \end{bmatrix}$$

where A^{00} , A^{01} , A^{10} , and A^{11} are submatrices of A^{-1} with dimensions that correspond to A_{00} , A_{01} , A_{10} , and A_{11} . If an animal included in a_1 does not have a record (e.g., a bull), corresponding column of Z is null and corresponding right-hand side is 0.

Equations are transformed with the QP transformation of Quaas and Pollak (18) to solve for genetic merit directly. If V is coefficient matrix with solution vector s and right-hand side vector f , then $Vs = f$. The transformation matrix P for this model is:

$$P = \begin{bmatrix} I & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & A_{10}Q \\ 0 & 0 & 0 & 0 & I & Q \\ 0 & 0 & 0 & 0 & 0 & I \end{bmatrix}$$

with P^{-1} constructed by negating terms containing Q . Transformed equations are:

$$(P^{-1})'VP^{-1}Ps = (P^{-1})'f$$

The new solution vector (s^*) becomes:

$$s^* = Ps = \begin{bmatrix} h \\ c \\ p \\ u_1 \\ u_0 \\ g \end{bmatrix}$$

where u_1 and u_0 are solved for directly. Pre-multiplication of the left-hand and right-hand sides by $(P^{-1})'$ preserves symmetry. The new right-hand side vector (f^*) becomes:

$$f^* = (P^{-1})'f = \begin{bmatrix} H'y \\ C'y \\ Z'y \\ Z'y \\ 0 \\ 0 \end{bmatrix}$$

The new coefficient matrix (V^*) becomes:

$$V^* = (P^{-1})'VP^{-1} = \begin{bmatrix} H'H & H'C & H'Z & H'Z & 0 & 0 \\ C'H & C'C + Ik_c & C'Z & C'Z & 0 & 0 \\ Z'H & Z'C & Z'Z + Ik_p & Z'Z & 0 & 0 \\ Z'H & Z'C & Z'Z & Z'Z + A^{11}k_a & A^{10}k_a & 0 \\ 0 & 0 & 0 & A^{01}k_a & A^{00}k_a & -Qk_a \\ 0 & 0 & 0 & 0 & -Q'k_a & Q'Qk_a \end{bmatrix}$$

Key steps follow to show that submatrices of the sixth row and column of the transformed coefficient matrix are as shown. Because:

$$\begin{bmatrix} A_{11} & A_{10} \\ A_{01} & A_{00} \end{bmatrix}^{-1} = \begin{bmatrix} A^{11} & A^{10} \\ A^{01} & A^{00} \end{bmatrix}$$

then:

- 1) $A_{01}A^{11} + A_{00}A^{01} = 0$ and $A_{00} = I$; thus, $A_{01}A^{11} + A^{01} = 0$;
- 2) $A_{01}A^{10} + A_{00}A^{00} = I$ and $A_{00} = I$; thus, $A_{01}A^{10} + A^{00} = I$;
- 3) $A^{01}A_{10} + A^{00}$ is transpose of $A_{01}A^{10} + A^{00}$ and thus equals I ; and
- 4) other terms that sum to zero in $(P^{-1})'VP^{-1}$ are obvious.

Solutions for unknown parents are not needed. Thus, equations for genetic merit of unknown parents could be absorbed algebraically. As can be seen by inspection of the transformed coefficient matrix, absorption affects only blocks of coefficients associated with a_1 and g . After absorption, this block becomes:

$$\begin{bmatrix} Z'Z + [A^{11} - A^{10}(A^{00})^{-1}A^{01}]k_a & A^{10}(A^{00})^{-1}Qk_a \\ Q'(A^{00})^{-1}A^{01}k_a & Q'[I - (A^{00})^{-1}]Qk_a \end{bmatrix}$$

Let this block be denoted as:

$$\begin{bmatrix} Z'Z + W_{11}k_a & W_{10}k_a \\ W_{01}k_a & W_{00}k_a \end{bmatrix}$$

which is equivalent to:

$$\begin{bmatrix} Z'Z + A_{11}^{-1}k_a & -A_{11}^{-1}A_{10}Qk_a \\ -Q'A_{01}A_{11}^{-1}k_a & Q'A_{01}A_{11}^{-1}A_{10}Qk_a \end{bmatrix}$$

Procedures developed by Westell (23) and Westell and Van Vleck (25) for calculating elements of W make it easy to include group effects defined through unknown parents. The group that represents an unknown parent of an animal can be treated nearly the same as a parent in the computations rather than

having to accumulate terms such as $Q'A_{01}Z'ZA_{10}Q$ and $Z'ZA_{10}Q$, which would involve many multiplications and additions of fractions.

Each row or column of W corresponds to an animal, its sire, or its dam. The corresponding group replaces an unknown sire or dam. As discovered by Westell (23), contributions to elements of W for an animal are:

$$\begin{array}{l} \text{Animal} \\ \text{Sire} \\ \text{Dam} \end{array} \begin{bmatrix} 1 & -.5 & -.5 \\ -.5 & .25 & .25 \\ -.5 & .25 & .25 \end{bmatrix} Dk_a$$

where $D = 4/(\text{number of unknown parents} + 2)$. Thus, $D = 2$ if both parents are known, $4/3$ if one parent is unknown, and 1 if both parents are unknown.