Animal Model Evaluation Within Herd Linked to National Evaluations

G. R. WIGGANS and P. M. VanRADEN
Animal Improvement Programs Laboratory
Agricultural Research Service
United States Department of Agriculture
Beltsville, MD 20705

ABSTRACT

Computer programs with modest computational requirements were developed to calculate animal model evaluations on a herd basis. Evaluations are connected to the national system through parents with national evaluations and parents assigned to unknown-parent groups. The programs can be used to compute evaluations of animals in herds not included in national evaluations or to update evaluations between semiannual national evaluations for herds that are included. Computer memory requirements are reduced dramatically from those of the national system, because only data for animals with records or progeny in the herd are in memory at any time. Computational requirements also are reduced by processing only lactation records initiated after an adjustable date; earlier lactation records are replaced by management group deviations. For older animals, processing is reduced further by not updating the animal's evaluation. When progeny evaluations no longer are updated, the animal's evaluation no longer is required. Solutions for management group, permanent environmental, and herd-sire effects are by Gauss-Seidel iteration and for animal effects by approximate successive overrelaxation.

(Key words: animal model, genetic evaluation, updating)

INTRODUCTION

In the United States, national genetic evaluations of dairy cattle yield traits are computed semiannually by USDA. To ensure the integrity of these evaluations, only data from herds in supervised testing plans are included. Cows without sire identification are excluded because they cannot contribute information to evaluation of sires. Owners of unsupervised herds and of cows without sire identification also may desire genetic evaluations. More frequent evaluations that include latest data might aid in earlier location of outstanding cows and use of outstanding bulls. To be most useful, all evaluations should have the same genetic base and be computed with similar methods so that all animals can be compared easily.

An animal model has been used by USDA (9) to compute yield trait evaluations since July 1989. The simultaneous nature of animal model evaluations and the inclusion of more sources of information make on-farm approximation of national evaluations more difficult than with the previous evaluation method, the Modified Contemporary Comparison (MCC). More frequent evaluations similar to those of the national system and that include cows without sire identification and cows from unsupervised herds would be useful. Data for such evaluations are collected and stored by dairy records processing centers (DRPC) or in some cases by on-farm computers, and evaluations might be computed or approximated at these locations. Most DRPC have had ranking systems, but often these have been on a within-herd or regional basis and have emphasized predicted producing ability rather than transmitting ability. When genetic merit has been predicted, computing procedures and genetic bases have differed from those of the national system.

Mixed model procedures usually have required that all data be reprocessed at each evaluation (2, 3) even though only a small percentage of data are new. Computational requirements of the animal model could be reduced by updating strategies that concentrate processing on recent data. Jamrozik and Schaeffer (personal communication) studied some updating methods in a model with a single observation per cow. They concluded that those methods that were accurate did not save much computational effort and those that

Received November 17, 1989.
Accepted March 22, 1990.

significantly reduced computing time were not sufficiently accurate. However, with multiple observations per cow and a method of maintaining a summary of all information, adequate accuracy might be achieved with significant reduction in computing time. Some approximation may be necessary to solve the problem within available computing power, but effect of any approximation should be small. In the USDA system, data for animals from all herds for a breed are accumulated in computer memory. Modification of this approach is necessary to make animal model evaluations feasible on smaller machines.

Henderson (1) presented a method for intraherd prediction of breeding values using an animal model. That method required little computer memory and relied on nationally evaluated sires as ties to national evaluations. Recently, an unknown-parent grouping procedure (5, 8) has eliminated the need to assume that base animals in each herd are from a common base population. Advances in computational techniques (7, 9) suggest that improvements in computational efficiency also are possible. An updating strategy could limit further the computational resources required.

The purpose of this study was to develop a herd system to compute genetic evaluations comparable with those from the national system and within DRPC computing capabilities.

ANIMAL MODEL FOR INDIVIDUAL HERDS

Model

The model as given by Wiggans et al. (9) includes fixed management (m) and random herd-sire (c), permanent environmental (p), animal (a), and residual (e) effects:

$$y_{ijkl} = m_{ij} + c_{ik} + p_{kl} + a_{kl} + e_{ijkl}$$

where y_{ijkl} = milk, fat, or protein yield of cow kl (daughter l of sire k) in herd i in year-season, parity, and registry group j. Animal effect is breeding value and includes effects for unknown-parent groups. Variance components scaled to a phenotypic variance of 1 are .25 for genetic, .16 for permanent environmental, .14 for herd-sire, and .45 for residual variances, which results in heritability of .25 and repeatability of .55. The same variance components are used for milk, fat, and protein. Evaluations are reported as predicted transmitting abilities, $\hat{a}_{kl}/2$ (10).

Data Requirements

Data for genetic evaluations are stored in yield and pedigree files. These files are generated by the herd evaluation system and are input for the following evaluation after insertion of information from new animals and new or updated lactation records.

The pedigree file is a combination of information from the herd and national systems and contains herd, animal, sire, and dam identifications; birth year; previous or national evaluation; and reliability. Reliability is an estimate of squared correlation between predicted and true genetic values. Cows with national evaluations but relatively little information from other herds are evaluated in the herd system. Birth year for other nationally evaluated animals is coded as 0 to indicate that the animals' evaluations are not to be updated during evaluation. Parent information for such animals is not needed. Most of the pedigree information required is part of normal milk recording data. Pedigree data for progeny test bulls are included with national evaluations. For local bulls, sire and dam identification and birth date may have to be obtained from the producer or for registered animals from the breed association. The pedigree file is sorted by herd and birth year.

The yield file contains herd, animal, and sire identification; yield deviations (weighted average of $y_{ijkl} - \hat{m}_{ij} - \hat{c}_{ik} - \hat{p}_{kl}$ where a circumflex indicates estimate of effect); predicted producing abilities; herd-sire effects; and, for each lactation, calving year and month; lactation length weight (9); milk, fat, and protein yields; and management group deviations ($y_{ijkl} - \hat{m}_{ij}$). None of this information comes from the national system. Because the same file format is used to report results, the yield file also contains average number of lactations in management groups and daughter equivalents from progeny (10). The yield file is sorted by herd.

Computational Procedures

The programs developed for the national system (9) were the basis for the three programs in this herd system. The number of steps in this system is fewer because locating ancestors and associating national evaluations with pedigree records are completed before data enter the system. Three (or more) traits can be processed. Data for all traits are prepared at the same time. Iteration is sequential by trait. Solutions for unknown-parent groups and nationally evaluated animals with most of their information from outside the herd are not changed. This provides consistency between evaluations from the herd system and national evaluations and avoids the problem of including information again from the herd that also may have been included in national evaluations. Cows could have both a national evaluation and a more current herd evaluation.

The herd system processes only traits that are present. Any number of traits could be processed, but all traits must be observed on every cow in the herd. The same set of ratios of variance components are used for all traits. A herd may be processed separately by trait if information on some traits is missing for some cows. Control of the number of traits allows reduction in processing time for milk-only herds and herds not testing for protein.

Only modest computing resources are required for this herd system. Herds of 10,000 cows can be processed with 2 Mb of memory. The count of animals includes all cows that ever had a production record in the herd, not just the current cows. Larger herds would require proportionally more memory. Memory requirements could be controlled by dropping earlier animals from the system as explained for updating. The functions of the three computer programs follow.

Preparation Program. This program processes herds one at a time. For each herd, it creates vectors of previous solutions for animal, herd-sire, and permanent environmental effects and yields for each trait. These solutions are from previous herd evaluations; however, previous national evaluations are used for across-herd parents. A variable number of lactations per cow is accommodated by storing all lactation data in the same vector. A vector indexed by cow indicates how many lactations from that vector belong to each cow. This method avoids allocating computer memory for the maximum number of lactations for each cow and is simpler than the strategy of the national system, which stores identification and lactation information in a single vector with a hierarchical structure. Animals are coded by the sequence of the pedigree file. A parent is recoded as an animal before it occurs as a parent, because the pedigree file is sorted by birth year. Nationally evaluated animals are assigned the smallest numbers. Bulls and cows are interspersed. Unknown-parent group solutions are at the beginning of the vector of animal effects. A particular unknown-parent group solution is chosen based on breed, birth year, and sex of animal and parent. For Holsteins, country of origin (Canada or the United States) also is considered. Management groups are assigned in the same way as for the national system (9). Iteration on the data does not require formation of the coefficient matrix, only its diagonal, which is collected in this program. Each data vector is included as a long record in the same file. This approach enhances processing efficiency because large blocks of data are accessed with a single instruction.

Iteration Program. The iteration program processes the data one herd and one trait at a time. Separate processing by trait saves memory because information for fat or protein can use the same memory as for milk. Sequence of solving equations is management group, permanent environmental, herd-sire, and animal effects. Management group solutions are computed after accumulating adjusted right-hand sides from individual lactation data. Total management group deviations then are stored for each cow. Adjusted right-hand sides for permanent environmental and herd-sire effects are accumulated while processing these management group deviations. Because a cow is nested within her sire, herd-sire contributions can be collected after computation of a new permanent environmental solution but during the same processing of the data. Thus, solutions for management group, permanent environmental, and herd-sire effects are by Gauss-Seidel iteration.

Solutions for animal effects are computed by first processing total management group deviations to collect contributions of lactations to cows and then processing pedigree data from youngest to oldest. This sequence ensures that all progeny contributions are collected before an animal is processed. The current round solution for the animal then is computed and includes current round solutions for progeny and previous round solutions for parents. Rate of convergence is improved by using a relaxation factor to multiply the change applied to each round. The current value for the animal is subtracted as each contribution is made to the accumulator; therefore, its value approaches 0 at convergence. This accumulator is multiplied by the relaxation factor and divided by the diagonal of the coefficient matrix. The new solution is the sum of this value and the previous solution. Contribution from this animal then is added to the accumulators of its parents. This procedure is approximate successive overrelaxation because contribution of each parent's solution to the other parent's adjusted right-hand side is computed with the solution from the previous round of iteration. For exact successive overrelaxation (4), the solution from the current round would be used for one of the parents in computing its contribution to the other parent. The national system uses previous round solutions for all animals. The within-round use of new solutions was expected to improve rate of convergence (4). An advantage of using updated solutions for progeny in parent evaluations is that information from the youngest animals (those making the greatest changes) is spread to all ancestors each iteration. In a test that included two animals with information coming only from the same progeny, the system failed to converge. Solutions for those two parents alternated high and low in each round with a relaxation factor of 1.7 but converged quickly with a relaxation factor of 1.3. Thus, high relaxation factors should be avoided if using approximate successive overrelaxation. In exact successive overrelaxation, the updated solution for one of the parents would have been used in the evaluation of the other, thus limiting fluctuations in the evaluations.

The program computes yield deviation and predicted producing ability for each cow and management group deviation for each lactation. After processing all herds, an average number of rounds of iteration required to reach convergence criterion over herds, the maximum rounds for any herd, and the maximum difference between solutions by trait are printed.

Reliability Program. This program accesses the file created in the preparation program and computes reliability using values from previous evaluations primarily for calculating amount of information that an animal contributes to each of its parents. An adjustment for reliability of the other parent's evaluation is necessary to compute contribution to a parent. Lactation data are processed to collect the amount of information from records. After this, animals are processed first from youngest to oldest to measure how much information is contributed by progeny and then from oldest to youngest to compute reliability so that parent reliability is available for progeny. Lactations in management groups are counted to provide average number of lactations in a cow's management group as a supplement to reliability. Yield and pedigree files are updated with solutions from the iteration program.

Controlling Processing Expense Through Updating

The herd system was designed to minimize computer processing by maintaining some solutions at their value from the previous evaluation (freezing). Lactations before a particular calving date (cutoff date) are not processed; therefore, the management group effect for such lactations is not changed. All information from such animals contributes to evaluations of their relatives; therefore, no information is lost. However, management group effects are not allowed to adjust to changes in estimates of the other effects. Adequate accuracy probably can be obtained without processing all lactation records of current cows. Solutions for cows are frozen and cows eliminated from the system in stages as follows.

All Lactations After Cutoff Date. If all of a cow's lactation records are after the cutoff date, all effects are estimated.

Some Lactations Before Cutoff Date. Only management group effects with lactations after the cutoff date are estimated. For lactations before the cutoff date, management group deviations from a previous evaluation are used. The cutoff date is applied on a lactation basis instead of a cow basis so that management groups will not lose later lactations of older cows if their first lactations fall before the cutoff date.

All Lactations Before Cutoff Date. Cow's lactation data are represented as a single record that contains average management group deviation.

Parent Evaluations Not Updated. If evaluations of a cow's parents are frozen in addition to all her lactations being before the cutoff date, her evaluation also is frozen. The cow's evaluation continues to be included in herd evaluations of her progeny and serves as a tie among progeny in the same way as do national evaluations. Some increase in accuracy can be obtained by postponing the freezing of a cow's evaluation for several years if she has several progeny in the herd that are adding data to their evaluations.

Progeny Evaluations Not Updated. If progeny as well as parent evaluations are frozen and all lactations are before the cutoff date, the cow's evaluation no longer contributes to evaluations of other animals and, therefore, no longer is needed.

Bulls and cows with national evaluations based primarily on information from other herds serve as ties with the national evaluations and have their most recent evaluations from the national system used as "frozen" values. These animals are processed the same way as cows with all lactations before the cutoff date and parent evaluations not updated. Although their evaluations are not affected by the herd system, they contribute information. Initially, local bulls (bulls with progeny primarily in the herd being evaluated) and dams without records are processed the same way as cows with all lactations before the cutoff date.

Initial Processing

Some initial processing is required to collect lactation data from all cows (including dead cows) back to when the herd began testing. The cutoff date can be employed only if management group deviations are available following the first complete evaluation with the herd system. Some herds may not have enough information to justify applying the system. If no pedigree information has been recorded, the animal model adds little to within-herd deviations. With a moderate amount of pedigree information, animals without known parents will use unknown-parent solutions and will contribute to accuracy of estimation of management group effects.

The pedigree file is composed of data from all cows with lactations, their parents, and additional ancestors that provide ties. A parent record with sire and dam identification numbers of 0 is the eldest ancestor in every path. The pedigree file can be constructed by forming three records for each cow with lactations: her own cow-sire-dam record and records for her sire and dam with values of 0 for their parents' identification numbers. The birth year in these constructed parent records is that of the cow minus 3. A code differentiates between constructed and actual pedigree records. The records are sorted by animal, code, and birth year, which causes actual records to occur first. The first record for each animal is kept. This method generates the minimum pedigree file required.

If additional common ancestors for animals already identified can be found, their information will contribute to accuracy of evaluations. Bulls are matched with their national evaluations. Accuracy of evaluations is increased by using national evaluations of cows with more information from other herds than the new information from the herd being evaluated, e.g., cows with descendants in other herds and no additional lactations in the current herd and cows in other herds with progeny in the herd being evaluated. Birth year of these animals is set to 0, and sire and dam identifications are ignored. Pedigrees for local bulls are valuable because they allow the parents of local bulls to contribute to their evaluations. If such bulls' parents do not have pedigree records already, records for them must be added.

Routine Processing

When USDA evaluations are released, the new national evaluations of sires, some dams, and unknown-parent groups must be included. Before each evaluation of a herd, data for new cows and new, corrected, and updated lactation records must be added to the yield file with the appropriate lactation length weights. Pedigree records for new animals also must be placed in the pedigree file and records for their parents added if they are not there already.

SYSTEM TEST

Test Data and Methods

Yield and pedigree records were from 100 Ayrshire herds that included 27,467 cows in the Northeast and Midwest United States. Herds were required to have accumulated 100 or more cows with lactation records with management group mates since 1960 and have a most recent calving date of September 1987 or later. Solutions were computed from data available for January 1989 national evaluations. Updating strategy was tested by adding data received for July 1989 evaluations (latest calving date of March 1989). Pedigree information was collected from lactation records and national evaluations. Bulls and cows were matched with national evaluations from the January 1989 preliminary animal model evaluation. Bull evaluations were included to form ties and, therefore, were not updated; cow evaluations served as starting values for iteration.

Results from evaluations with January and July data were compared with national evaluations. The relaxation factor that yielded the most rapid convergence was determined for both data sets. Effect of length of period following cutoff date also was studied; cutoff date ranged from January 1960 (all data) through January 1988 (15 mo of calvings).

Test Results

Numbers of rounds of iteration required to reach the same convergence criterion used as a stopping point in the national system (sum of squared differences divided by the sum of squared current round solutions $< 1 \times 10^{-6}$) are in Table 1 for various relaxation factors. A larger relaxation factor was better for the initial evaluation. Because starting values for herd-sire and permanent environmental effects were 0, more change in solutions was required. For the update, a smaller relaxation factor was optimum, and number of rounds of iteration required was small. Fewer rounds were required for the update because solutions from the initial evaluation served as starting values for all effects.

TABLE 1. Average rounds of iteration and central processing unit (CPU) time required to reach a convergence criterion[1] value of 1×10^{-6} for various relaxation factors.

Evaluation	Relaxation factor	Rounds of iteration	CPU Time
			(s)
Initial[2]	1.5	23.8	44.25
	1.6	23.6	44.03
	1.7	23.5	44.00
Update[3]	1.2	6.9	19.13
	1.3	6.5	18.68
	1.4	6.7	18.92

[1]Sum of squared differences divided by the sum of squared current round solutions.

[2]Included data available for January 1989 evaluations; starting values were from the January 1989 preliminary national model for animal effects and 0 for herd-sire and permanent environmental effects.

[3]Included data available for July 1989 evaluations; starting values were from January 1989 herd evaluations for all effects.

Correlations between herd and national evaluations are in Table 2 for various cutoff dates. Highest correlation was between January evaluations as expected, because bull evaluations in the herd system were the same as those from the national system. Progeny in other herds would be the primary reason for differences in cow evaluations. For July data, January bull solutions also were used in calculations to simulate evaluations that were computed some months after the latest national evaluation. Correlation was reduced only minimally (.961 to .952) for a cutoff date as recent as January 1988. In computing these correlations, evaluations for all cows were allowed to change. The changing cutoff date controlled which management group effects were frozen at January 1989 values.

Central processing unit times reported for the iteration program were on an IBM 3090-600E. Times on an IBM 9370 model 90 were five to seven times greater. The modest time requirements of this system indicate that it may be practical for routine use. The preparation and distribution programs require about the same amount of time as does the iteration program when updating.

Savings in processing time from this updating strategy were greater than those observed by Jamrozik and Schaeffer (personal communication), because they investigated a single lactation case in which processing contributions from relatives and computing new solutions were the principal expenses. Savings from not processing all lactations if there are more than one is greater than for a system processing only first lactations.

TABLE 2. Average number of rounds of iteration and central processing unit (CPU) time required to reach a convergence criterion[1] value of 1×10^{-6} and correlations between herd and corresponding national evaluations[2] for various restrictions on length of period of data included in iteration.

Cutoff date[3]	Latest calving date	Rounds of iteration	CPU Time	Correlation between herd and national evaluation[4]	
				Milk	Fat
			(s)		
January 1960	September 1988	23.5	44.00	.989	.990
January 1960	March 1989	6.5	18.68	.961	.964
January 1980	March 1989	6.2	16.26	.966	.969
January 1985	March 1989	5.8	12.15	.966	.969
January 1988	March 1989	5.4	11.92	.952	.954

[1]Sum of squared differences divided by the sum of squared current round solutions.

[2]July 1989 evaluation was update of January 1989 evaluation with 6 mo of additional data.

[3]Earliest calving date of lactations included in management group effects that were estimated; management group effects before cutoff retained their estimate from the January evaluation.

[4]For September 1988 latest calvings, herd evaluations correlated with January 1989 preliminary national animal model evaluations; for March 1989 latest calvings, herd evaluation correlated with July 1989 national evaluations.

DISCUSSION

The herd system could provide more accurate and complete information to producers. The AI organizations may find daughter yield deviations of their bulls from the herd system useful in predicting the bulls' eventual evaluations. This program emphasizes the importance of national evaluations because it relies on them to tie evaluations for all herds together. National evaluations depend on testing programs that provide supervised data that are the combined observations of supervisor and producer. Supervised data are necessary because the national program has an impact on all producers, not just the one recording the data.

This herd system can serve two classes of herds. Producers with herds in testing plans not included in the national system can have evaluations for their cows. These evaluations would be for use within herds. Because herd evaluations are tied to national evaluations through nationally evaluated animals, these producers can compare their animals with others. Herd evaluations may be useful to potential buyers, but their value depends on validity of the data that the producer contributes. Source of the data for evaluations must be indicated clearly, e.g., with a special designation such as "unsupervised local". These evaluations could be computed on the farm as well as by a DRPC.

For herds that contribute to the national system, managers can have genetic evaluations based on the most current data available and can have them updated frequently. These evaluations could be designated as "supervised local". Such evaluations are a desirable complement to national animal model evaluations because they provide the best information for prediction of results of the next national evaluation. A concern about the animal model has been that approximating an animal's evaluation from DHI reports would be more difficult than with MCC because of the simultaneous estimation of various factors in the model. With the herd system, estimation is done uniformly and routinely. Its results are a direct extension of the national system and, therefore, can receive similar use if the high correlations in the system test are obtained in practice. If these local evaluations were computed each time that herd data were processed, the evaluations could be the basis for identifying animals for verification testing.

To best predict results of the next national evaluation, data included in the herd system should mimic those used in the national system. However, herds in supervised testing plans often include cows not included in national evaluations, such as cows without sire identification. Including these cows' data would provide them with predicted producing abilities and allow them to contribute to estimation of management group effects. Evaluations for other cows in the herd then might be more accurate, but including data from such cows could reduce the correlation between herd and national evaluations of other cows.

Some standardization will be desirable in implementation of the herd system. Factors used for age-season adjustment and projection to 305 d should be the same as those used nationally. The test herd that is processed periodically by the DRPC for checking accuracy of computing and reporting lactation records could be adapted for checking computation of herd evaluations. Such checking would be particularly useful to ensure that information from national evaluations was successfully incorporated after each semiannual computation.

The herd system is not sufficient by itself. A national evaluation is necessary to combine information across herds. The national evaluation provides the best information for bulls with progeny in more than one herd and for cows with sons and daughters in other herds. However, for young cows, which are adding lactation information, the herd evaluation could be more accurate than the national evaluation because of its currency. If producers can be served adequately by evaluations from the herd system, more stringent requirements for records to be included in national evaluations could be imposed without depriving producers of evaluations. This system is similar to the decentralized computing of estimates of genetic merit suggested by Robinson and Chesnais (6).

Accurate genetic evaluations for animals in individual herds can be computed without using a supercomputer. The herd programs can be used on personal computers with adequate memory. Minimization of memory requirements and a strategy for updating make this system attractive as a basis for national animal model evaluations.

ACKNOWLEDGMENTS

This research was conducted using the Cornell National Supercomputer Facility, a resource of the Center for Theory and Simulation in Science and Engineering at Cornell University, which is funded in part by the National Science Foundation, New York State, and IBM Corporation. Assistance of J. S. Clay in designing the system to meet the needs of the DRPC, suggestions of I. Misztal and L. D. Van Vleck, and the assistance of S. M. Hubbard in manuscript preparation are acknowledged.

REFERENCES

1 Henderson, C. R. 1975. Use of all relatives in intraherd prediction of breeding values and producing abilities. J. Dairy Sci. 58:1910.

2 Hudson, G.F.S. 1984. Extension of a reduced animal model to recursive prediction of breeding values. J. Anim. Sci. 59:1164.

3 Martinez, M. L., and M. F. Rothschild. 1983. Recursive procedures in sire evaluation. J. Dairy Sci. 66:1967.

4 Misztal, I., and D. Gianola. 1987. Indirect solution of mixed model equations. J. Dairy Sci. 70:716.

5 Quaas, R. L. 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71:1338.

6 Robinson, J. A. B., and J. P. Chesnais. 1988. Application of the animal model on a national basis to the evaluation of Canadian livestock. J. Dairy Sci. 71(Suppl. 2):70.

7 Schaeffer, L. R., and B. W. Kennedy. 1986. Computing strategies for solving mixed model equations. J. Dairy Sci. 69:575.

8 Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic groups in an animal model. J. Dairy Sci. 71:1310.

9 Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 1988. Implementation of an animal model for genetic evaluation of dairy cattle in the United States. J. Dairy Sci. 71(Suppl. 2):54.

10 Wiggans, G. R., and P. M. VanRaden. 1989. USDA-DHIA animal model genetic evaluation. Natl. Coop. Dairy Herd Improvement Progr. Handbook, Fact Sheet H-2, Washington, DC.