Implementation of an Animal Model for Genetic Evaluation of Dairy Cattle in the United States

G. R. WIGGANS
Animal Improvement Programs Laboratory
Agricultural Research Service
United States Department of Agriculture
Beltsville, MD 20706

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

Procedures were developed for animal model evaluation of dairy cattle for milk, fat, and protein yields as well as evaluation for fat and protein percentages as a function of yield evaluations. Data available for January and July 1987 Modified Contemporary Comparison evaluations were used to compute animal model evaluations. For each cow, records from her first five lactations are included. Cows without a first lactation record are evaluated separately. Relationships with all known male and female relatives influence evaluations. Solutions are computed by iteration on data for fixed management and random herd-sire, permanent environment, and breeding value effects. Iteration is by second-order Jacobi for breeding value and by Gauss-Seidel for other effects. Accuracy (expressed as repeatability) is computed as in Modified Contemporary Comparison procedures except that sire repeatability includes contributions from parents and dam repeatability is adjusted for number of daughters. A genetic base is imposed by subtracting average evaluation of cows born in 1984 from all evaluations. Results include predicted transmitting ability (one-half breeding value), predicted producing ability (sum of herd-sire, permanent environmental, and breeding value effects), parent average (average transmitting ability of parents), and average effect of management groups containing a cow's records.

INTRODUCTION

An animal model allows simultaneous genetic evaluation of male and female dairy animals with all relationships included. Previously computing constraints limited this approach to evaluation within herd (3). Recent work by Westell and Van Vleck (17), Wiggans and Misztal (19), and Wiggans et al. (20, 21) demonstrated that animal model evaluations are computationally feasible with large data sets across herds. However, for conversion of the national dairy cattle evaluation system to the animal model, further research was needed 1) to demonstrate feasibility of the animal model with records from over 10 million Holsteins, 2) to develop an updating procedure that minimizes processing expense through effective use of results from evaluations calculated 6 mo earlier, 3) to adapt the animal model system to evaluate protein yield for the subset of cows with protein data, 4) to develop a procedure to include contributions from parents in calculation of repeatability (R) of bulls, and 5) to determine which predictions of model effects should be distributed to the industry in addition to predictions of breeding value. The purpose of this study was to apply the animal model evaluation method to records from the major US dairy cattle breeds and develop a system appropriate for routine evaluations.

ANIMAL MODEL EVALUATION SYSTEM

Data

Data for an animal model evaluation system include yield and pedigree information. Records for lactations after fifth are excluded. Records from later herds and from cows without a first lactation record are evaluated separately. Later lactation data from these cows do not influence evaluations of other animals. Separation of cows without first lactation records avoids selection bias that could result from culling animals based on unreported first lactation performance. Such culling would result in availability of later records only from better animals. Records from later herds are separated to simplify computation of permanent environmental effects across herd. For cows without first lactation records, evaluation is restricted to those born in the last 10 yr; however, all cows with evaluated progeny are evaluated. Thus, some cows without first lactation records but born more than 10 yr ago are evaluated as dams without records.

Model

The model includes fixed management (m) and random herd-sire (c), permanent environmental (p), animal (a), and residual (e) effects as in Wiggans et al. (20):

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

where y_{ijkl} = milk yield of cow kl (daughter l of sire k) in herd i in year-season, parity, and registry group j. Management groups are defined by herd, year-month of calving (2 mo seasons), parity (first or later), and registry status (registered or grade). Animal effect is breeding value and includes effects for genetic groups (13, 16), which are defined for unknown parents. Every ancestral path ends at an unknown-parent group. Parents without yield records and not connecting at least two animals with yield records also are assigned to unknown-parent groups. The method of calculating inverse of the relationship matrix (A^{-1}) is extended to include these groups (13, 16) with elements of the inverse dependent on animal's number of known parents. Values of variance components (see Appendix) are the same for milk, fat, and protein.

Effect of herd environment is called management effect because it includes registered-grade distinction and parity group as well as the usual herd-year-season classification. Calving seasons of 2 mo were chosen to make management groups as homogeneous as practical. The Modified Contemporary Comparison (MCC) procedure has 5-mo seasons centered on cow's calving month, which results in 12 overlapping seasons per year. Short seasons were selected because of complexity of including overlapping seasons. To provide for at least five lactation records in a management group, adjacent seasons are joined if necessary. A registered-grade distinction was adopted because of different management afforded registered and grade cows in the same herd (10).

Herd-sire effect is included to account for environmental correlation among a sire's daughters in the same herd. As explained in Wiggans et al. (20), this effect has been credited with increasing the correlation between early evaluations of bulls with daughters in few herds and their later evaluations based on widespread data.

Operational Steps for Computing Evaluations

The main operational steps are 1) processing of yield data, 2) processing of pedigree data, 3) preparation for computation of solutions, 4) iteration until satisfactory convergence, 5) calculation of R, 6) computation of supplemental evaluations, and 7) adjustment of breeding value to a genetic base and calculation of transmitting ability. Details of these steps are in the Appendix.

Computation of Breeding Value

Solutions are computed for all effects in the model during each round of iteration. Solutions are obtained by iteration on the data (14) rather than by forming the coefficient matrix. Only diagonal elements for animals are saved. As contributions to off-diagonal elements are encountered, they are used to compute adjusted right-hand sides (ARHS). This approach simplifies the procedure and reduces processing because no factors are absorbed. Absorption creates many off-diagonal elements, which results in more computer operations than does processing the data (14). Data are processed by herd. New solutions for management, herd-sire interaction, and permanent environmental effects are computed by Gauss-Seidel iteration while processing the herd.

Management effects are estimated first:

$$m_{ij}^{(r)} = \sum_{kl} w_{ijkl} \left(y_{ijkl} - c_{ik}^{(r-1)} - a_{kl}^{(r-1)} - p_{kl}^{(r-1)} \right) / \sum_{kl} w_{ijkl}$$

where r = round number of iteration and w = weight for lactation length (18). Lactation length weights (Table 1) are a measure of the amount of information in the record. For the first round, solutions for other effects are from the previous evaluation. If no previous evaluation exists, parent average is used for animal effect and 0 for other effects. Because data are sorted by herd and sire rather than by herd management group, contributions to ARHS and diagonals for all management groups in a herd are accumulated while

processing the herd, and new solutions for each are computed at the end of the herd.

Next, processing by sire, sum of each daughter's lactations are adjusted for management effects (x) is computed:

$$x_{kl} = \sum_{j} w_{ijkl} \left(y_{ijkl} - m_{ij}^{(r)} \right)$$

The x_{kl} and $\sum w_{ijkl}$ are stored. Permanent environmental effects for each daughter are estimated:

$$p_{kl}^{(r)} = \left[x_{kl} - \sum_{j} w_{ijkl} \, c_{ik}^{(r-1)} + a_{kl}^{(r-1)} \right] / \left(\sum_{j} w_{ijkl} + k_p \right)$$

The divisor includes variance ratio $k_p = e/p$ because p is a random effect.

Herd-sire interaction effects then are estimated:

$$c_{ik}^{(r)} = \sum_{l} \left[x_{kl} - \sum_{j} w_{ijkl} \left(a_{kl}^{(r-1)} + p_{kl}^{(r)} \right) \right] / \left(\sum_{l} \sum_{j} w_{ijkl} + k_c \right)$$

where k_c = variance ratio e/c.

Last, ARHS for each daughter (z) are accumulated:

$$z_{kl} = x_{kl} - \sum_{j} w_{ijkl} \left(c_{ik}^{(r)} + p_{kl}^{(r)} \right)$$

Processing continues with the next sire.

New animal and group solutions are computed by Jacobi iteration after all yield data have been processed. Jacobi iteration is employed because updated solutions for previously evaluated animals are not required for the current animal's solution.

Contribution of relatives to evaluation of an animal can be illustrated by:

$$a_{kl} = \frac{z_{kl} + k_a u_{kl}(a_s + a_d) + k_a \sum_{o}^{n_o} u_o (a_o - .5 a_m)}{\sum_{ij} w_{ijkl} + 2 k_a u_{kl} + .5 k_a \sum_{o}^{n_o} u_o}$$

where k_a = variance ratio σ_e^2 / σ_a^2, s = sire, d = dam, o = offspring, m = mate, n_o = number of offspring, and u = 1 if both parents are evaluated, 2/3 if one parent is evaluated, and 1/2 if neither parent is evaluated. Evaluations for all animals are from the previous round of iteration. Solutions for animals without records are computed the same way except that z_{kl} and $\sum w_{ijkl}$ equal 0. This formula shows that parents are included directly. Progeny are adjusted for mate's genetic value. Other relatives are included through their effects

on predictions of breeding values for parents or progeny. Iteration allows contribution from each animal eventually to affect all its relatives.

For iterations after second, a relaxation factor is applied that adds a percentage of the difference between the previous two rounds to the solution for the current round. The relaxation factor is chosen to maximize rate of convergence.

Computation of Final Evaluations

Breeding values are adjusted so that cows born in 1984 average 0. Adjusted breeding values are divided by 2 to obtain predicted transmitting abilities (PTA). The term PTA is used to distinguish final evaluations from breeding values and because the same term can be used for both cows and bulls. The adjective "predicted" indicates that the solution is for a random effect.

Table 2 lists information generated from the evaluation system for distribution to the dairy industry. All solutions are reported. Predicted producing abilities (PPA) include herd-sire, permanent environmental, and animal effects. An-

cestor contributions are reported as parent average, the average transmitting ability of parents. Economic indexes (PTA, dollars) using animal model evaluations and percentiles are computed as for MCC (6). Other computations are described in the Appendix.

A measure of the amount of information in an evaluation is necessary. These animal model calculations do not provide a direct approximation for R as do MCC procedures, and separate calculations are used to approximate MCC R. For consistency with cow evaluations and to recognize contributions from ancestors, R for bulls was changed to include parent contribution. Cow R was changed to include contributions from daughters. Table 3 shows R for bulls from various combinations of R from parents (R of parent average) and R from progeny; Table 4 shows R for cows from various combinations of R from parents and records (current R) and numbers of daughters. One-fourth of the sum of parent R is R of parent average, which also is R for an animal with no records or progeny information.

TABLE 2. Information generated from animal model evaluations.

Trait	Animal model information	Description
Milk, fat, and protein yields	Predicted transmitting ability (PTA)	One-half breeding value; adjusted so that cows born in 1984 average 0; for protein, adjusted for compatibility with milk evaluation.
	Parent average (PA)	Average PTA of parents; if either parent unknown, unknown-parent group effect substituted.
	Repeatability (R)	Same for milk and fat PTA; contribution to accuracy from parents included.
	R_{PA}	One-fourth of sum of parents' R; accuracy of PA as a prediction of transmitting ability.
	Predicted producing ability (PPA)	Sum of predictions of breeding value, herd-sire, and permanent environmental effects; prediction of relative performance in future lactations.
	Average management group effect	For cows, average of management effects of groups in which cow has records; for bulls, average of daughters' management group effects.
	Average mature-equivalent yield	Averaged over lactations for cows and over daughters for bulls.
	PTA dollars	Economic index combining relative values of milk and components; calculated as for Modified Contemporary Comparison.
	Percentile	Ranking based on PTA dollars.
Fat and protein percentages	PTA	Derived from yield evaluations in combination with first lactation ME yield of cows born in 1984.

TABLE 3. Resulting repeatability (R) of animal if R of parent average[1] (R_{pa}) is considered with R calculated from progeny information only for bulls or from own records only for cows

R_{pa}	R from progeny									
	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.00	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.05	.05	.14	.23	.32	.42	.51	.61	.70	.80	.90
.10	.10	.18	.27	.35	.44	.53	.62	.71	.80	.90
.15	.15	.22	.30	.38	.46	.54	.63	.72	.81	.90
.20	.20	.27	.33	.40	.48	.56	.64	.72	.81	.90
.25	.25	.31	.37	.43	.50	.57	.65	.73	.81	.90
.30	.30	.35	.40	.46	.52	.59	.66	.73	.82	.90
.35	.35	.39	.44	.49	.55	.61	.67	.74	.82	.91
.40	.40	.44	.48	.52	.57	.63	.68	.75	.82	.91

[1]Parent average is average of parents' predicted transmitting abilities; R_{pa} is one-fourth sum of parents' R.

Average of management effects for a cow's records is computed. These values are averaged over a bull's daughters to indicate environment in which the bull's daughters produced. Difference between this average and average of daughter records is an unregressed indication of the performance of a bull's daughters, which is similar to Modified Contemporary Deviation from MCC. This supplemental information is provided to aid in understanding how a particular evaluation could result. Production average represents starting information. Subtracting average of management effects indicates what remains for genetic evaluation. The PPA is the regressed cow effect comparable to that unregressed difference. Comparison of twice PTA with PPA indicates degree that breeding value differs from permanent environmental and herd-sire predictions. Comparison of parent average with PTA indicates impact of progeny and records. Study of these values should enable owners to monitor accuracy of data and evaluation methods and understand why the evaluation of a particular animal may differ from that expected.

RESULTS OF ANIMAL MODEL APPLICATION

The animal model system was applied to data available for January and July 1987 MCC evaluations for all breeds except Holstein, for which only January data were processed. Processing successive runs provided experience with the updating strategy and information on usefulness of previous evaluations as starting values for current evaluations. Numbers of records available for January 1987 evaluations prior to editing and numbers excluded are in Table 5 for Ayrshires, Holsteins, and Jerseys. Each herd-cow combination is a record. For Brown Swiss, Milking Shorthorns, and Red and Whites, percentages of records lost were similar to those for Ayrshires and Jerseys. For breeds other than Holstein, all records were eliminated for a substantial number of herds. Lack of a first lactation record was the major reason for not including cows. Some of these cows are included in the separate evaluation of cows without first lactation records that were born in the last 10 yr and have management group mates. Flexible assignment of management groups resulted in few cows without management group mates.

Table 6 shows numbers of equations solved for model effects for January 1987 evaluations. Number of management group effects is related to number of lactations and, therefore, appears high in comparison with number of cows. Modal management group size is five as expected because of how groups are formed. Inclusion of all ancestors

TABLE 4. Resulting repeatability (R) of cow if number of daughters is considered.

R^1	Number of progeny				
	1	5	10	20	50
.1	.14	.26	.37	.52	.72
.2	.23	.33	.42	.55	.73
.3	.32	.40	.48	.58	.74
.4	.42	.48	.53	.62	.75
.5	.51	.55	.60	.66	.77

[1]No daughter records.

TABLE 5. Number of records of herd-cow combinations available for January 1987 evaluations and numbers excluded for Ayrshires, Holsteins, and Jerseys.

	Ayrshire	Holstein	Jersey
Herd-cow records available	214,032	10,976,127	858,851
Herd-cow records excluded because:			
No first lactation record for cow	76,428	2,315,730	238,272
All parities >5	1474	54,320	16,416
No management group mates for cow	4625	40,489	10,065
Herds available	6530	104,986	16,529
Herds included	2941	86,120	8946

TABLE 6. Numbers of equations solved for January 1987 data by breed.

Breed	Management group	Herd-sire	Permanent environment[1]	Animal Cows	Animal Dams without records	Animal Sires	Unknown parent groups	Total
Ayrshire	49,531	44,461	131,184	118,429	44,346	10,504	17	398,472
Brown Swiss	61,434	53,958	155,716	138,743	47,887	13,609	17	471,364
Guernsey	183,146	158,088	517,702	471,932	187,679	42,716	20	1,561,283
Holstein	2,621,233	3,156,992	8,117,088	7,577,712	1,658,439	291,587	38	23,423,089
Jersey	189,073	174,515	602,415	541,224	199,018	42,762	24	1,749,031
Milking Shorthorn	8383	7311	21,932	20,040	5826	3528	10	67,030
Red and White	2333	2982	6571	6381	2710	1193	7	22,177

[1]Numbers are larger than numbers of cow equations because of separate permanent environmental effect for each herd in which a cow has records.

resulted in more dams without records than sires. Number of sires included all bulls with progeny except those assigned to unknown-parent groups. Number of permanent environmental effects was sum of number of cows with records and number of herds with later records for each cow. Records in later herds will not be included in routine evaluations; they will be processed with data for cows without a first lactation record.

Computer time required for the first six steps in obtaining animal model evaluations for January 1987 data is in Table 7. Table 8 shows rounds of iteration required to reach various degrees of convergence for evaluations from July 1987 data if January 1987 solutions are used as starting values. For most breeds, value of convergence criterion

TABLE 7. Central processing unit time required by animal model evaluation system for processing January 1987 data by breed.

System operation	Holstein	Other breeds[1]
	(min)	
Processing of yield data	102.9	10.2
Processing of pedigree data	52.7	14.2
Preparation for iteration	16.6	15.7
Iteration (25 rounds)	209.5	21.7
Calculation of repeatability	56.7	11.7
Preparation of final results	. . .[2]	3.2

[1]Includes Ayrshire, Brown Swiss, Guernsey, Jersey, Milking Shorthorn, and Red and White breeds.

[2]Not available.

TABLE 8. Rounds of iteration[1] required to reach convergence criterion values of various magnitudes for July 1897 data[2] by breed.

Breed[4]	Rounds required to reach convergence criterion[3] value of:					
	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Ayrshire	1	2	11	22	36	>50
Brown Swiss		1	8	19	30	44
Guernsey		1	6	16	37	>50
Jersey		1	6	19	30	47
Milking Shorthorn	1	8	18	29	40	>50
Red and White	2	11	23	34	47	>50

[1]Relaxation factor of .8.

[2]January 1987 solutions used as starting values.

[3]Sum of squared round to round differences divided by sum of squared current round evaluations.

[4]Holstein information not available.

was below 10^{-2} after first round of iteration. This accuracy would require approximately 10 rounds of iteration with starting values of 0. For some breeds, the relaxation factor of .8 may be too high; a factor of .7 will be used for future Milking Shorthorn and Red and White evaluations. Central processing unit time for each round of iteration is in Table 9 by breed. These times do not include additional time required for swapping between disk and memory that is associated with large memory utilization. This additional time was included in Table 7 and is significant for Holsteins.

Figure 1 shows average breeding value by birth year for each breed. Slopes of quadratic curves jointed at 1965 are in Table 10 for various years. The joint at 1965 allows for the substantial increase in rate of genetic progress that started about that time. Before 1965, relatively little progress occurred. Slopes changed with year, and annual increase in slope (kg/yr²) is in Table 10. Average breeding values approached 0 in the early 1980's as did MCC values used as starting values for iteration.

Restrictions to reduce distribution costs while still making evaluation information available for animals of current interest were investigated. These restrictions did not affect which animals were evaluated, only which animals were considered for distribution of evaluation information. To provide information on bulls being proven, all bulls aged less than 8 yr were included. Bulls 8 yr and older were required to have 10 or more daughters, each with a usable first lactation record; this number of daughters required is the same as

for MCC (1). Cow evaluations considered for distribution were limited to cows born in the past 10 yr. Numbers and percentages of animals that met these restrictions are in Table 11. Distribution of cow evaluations probably will be expanded to include all cows with evaluated progeny born in the last 20 yr. Dams without records will be included.

DISCUSSION AND CONCLUSIONS

The animal model evaluation system developed to replace the MCC system is similar to MCC in 1) inclusion of later lactation records, 2) inclusion of a herd-sire effect, 3) provision for comparing a lactation with appropriate first or later lactations of contemporaries, 4) variance component values (heritability of .2), and 5) general method of computing R. Differences are in Table 12.

TABLE 9. Central processing unit (CPU) time per round of iteration for January 1987 data by breed.

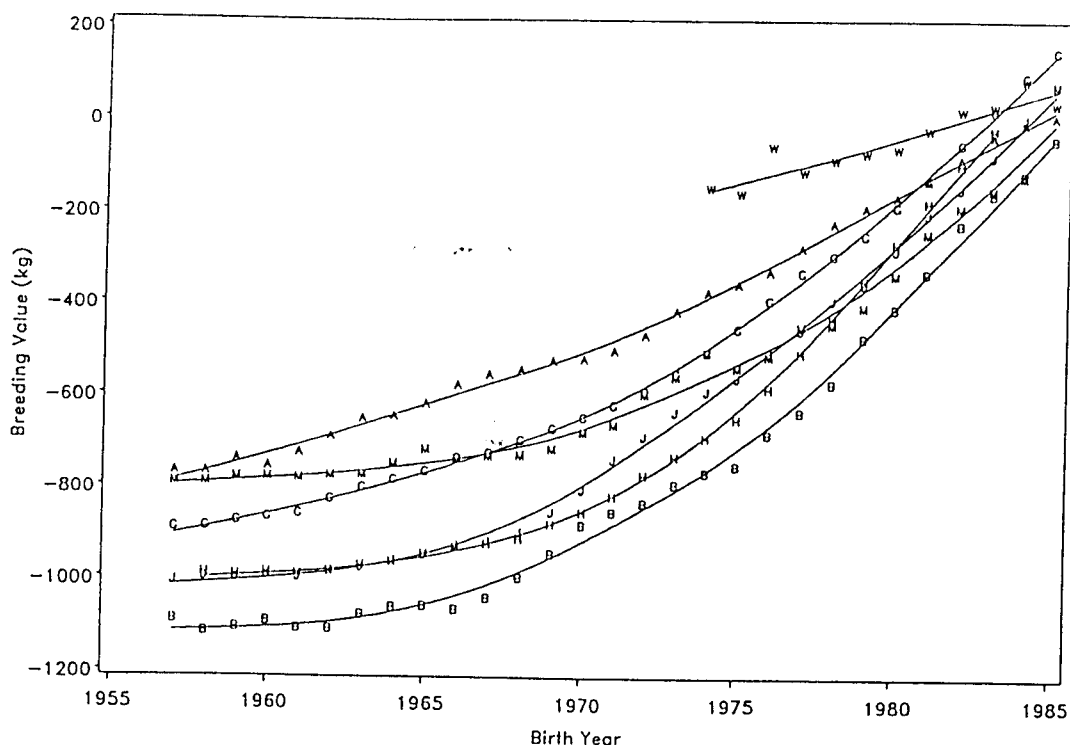| Breed	CPU Time
	(s)
Ayrshire	4.8
Brown Swiss	5.7
Guernsey	19.1
Holstein	353.6
Jersey	22.3
Milking Shorthorn	.7
Red and White	.2

Breeding Values from Animal Model

Figure 1. Average breeding values for milk yield from animal model evaluation system by birth year of animal and breed (A = Ayrshire, B = Brown Swiss, G = Guernsey, H = Holstein, J = Jersey, M = Milking Shorthorn, and W = Red and White).

Cows that do not have first lactation records are excluded from routine evaluation to avoid potential selection bias. Such cows are evaluated separately using management group and breeding value solutions from routine evaluations. These separate evaluations do not contribute to evalua- tions of other animals. The requirement for first lactation data is relaxed for protein evaluations (although first lactation milk and fat data are required) because of the relatively recent intro- duction of widespread protein testing and, there- fore, smaller likelihood for direct selection. Poten-

TABLE 10. Slopes for quadratic curve of average breeding value for milk yield over time by breed and birth year.

Breed	Birth year						Annual increase in slope
	1960	1965	1970	1975	1980	1985	
	(kg/yr)						(kg/yr^2)
Ayrshire	21.3	20.9	27.3	33.7	40.2	46.6	1.4
Brown Swiss	3.5	15.8	34.2	52.5	70.9	89.3	3.7
Guernsey	17.2	17.7	32.1	46.5	60.9	75.3	2.9
Holstein	5.4	7.6	32.7	57.7	82.7	107.8	5.0
Jersey	−1.9	25.1	38.2	51.3	64.3	77.4	2.6
Milking Shorthorn	3.7	8.5	22.0	35.5	49.1	62.6	2.7
Red and White	. . .	. . .	. . .	3.9	23.5	43.1	3.9

TABLE 11. Number of bulls born in the last 8 yr or with 10 or more daughters, numbers of cows born in the last 10 yr, and percentages of total bulls and cows by breed.[1]

Breed	Bulls		Cows	
	Number	Percentage	Number	Percentage
Ayrshire	1972	19	32,690	27
Brown Swiss	2362	19	46,759	33
Guernsey	8257	19	90,086	19
Jersey	8767	23	205,624	37
Milking Shorthorn	569	18	6722	33
Red and White	254	23	6089	87

[1]Holstein information not available.

tial for selection bias is reduced for protein because protein and milk evaluations from the protein subset are used only to compute a protein percentage evaluation. Evaluation of protein yield is a function of the protein percentage evaluation and the routine evaluation for milk yield.

Number of lactations per cow is restricted to five to limit the time period over which a cow can contribute data to her own evaluation and because later records gradually become less useful in estimating genetic merit (12). Further reduction in records might be possible. However, dairy producers are accustomed to having all records included in national genetic evaluations; therefore, any further reduction should be justified.

Records from herds other than the herd in which the first lactation was made are processed separately to simplify computation of permanent environmental effect across herd. Lactations in later herds do not affect evaluations of relatives and do not contribute to estimation of management effects.

Number of daughters is included in dam R. This addition provides a meaningful R for dams without records. Calculation of R includes several approximations. Improvements are possible, but the gain may not justify additional processing cost because R is only a guide to risk, not an estimate of genetic merit.

ACKNOWLEDGMENTS

TABLE 12. Differences between animal model (AM) and modified contemporary comparison (MCC) evaluations.

Characteristic	AM	MCC
Merit of mates considered	Yes	No
Sons contribute to parents	Yes	No
Daughters contribute to dams	Yes	No
Dams contribute to sons	Yes	No (ancestor merit includes maternal grandsire)
First lactation required	Yes (cows without first lactation records evaluated separately)	No (cows without first lactation records receive less weight)
Lactations included	1 – 5	1 – 15
Later herd lactations included	No, they are included in a separate evaluation	Yes
Repeatability components		
Parents for males	Yes	No
Daughters for females	Yes	No
Animals evaluated	All (simultaneously)	Recent only (sires, then cows)
Environmental group definition	Management group (registered-grade, 2 mo, flexible)	Contemporary group (5 mo centered)
Base definition	Birth year	Calving year

Number US-805-84 and the Holstein Association. Data preparation was by M. J. Govern and C. J. Shanefelter. Many helpful suggestions were contributed by B. G. Cassell, I. Hoschele, S. Hudson, and P. M. VanRaden.

REFERENCES

1 Dickinson, F. N. 1985. USDA-DHIA Sire Summaries for yield traits: data sources and summary requirements. Natl. Coop. Dairy Herd Improvement Prog. Handbook Fact Sheet H-3, Washington, DC.

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

4 Knuth, D. E. 1973. Searching and sorting. The art of computer programming. Vol. 3. Addison-Wesley, Reading, MA.

5 Misztal, I., and H. Haussmann. 1985. Comparison of alternative computational procedures in BLUP sire evaluation. Z. Tierz. Zuchtungsbiol. 102:142.

6 Norman, H. D. 1986. Sire evaluation procedures for yield traits. Natl. Coop. Dairy Herd Improvement Prog. Handbook, Fact Sheet H-1, Washington, DC.

7 Powell, R. L. 1978. A procedure for including the dam and maternal grandsire in USDA-DHIA Cow Indexes. J. Dairy Sci. 61:794.

8 Powell, R. L. 1981. Possible effects of embryo transfer on evaluation of cows and bulls. J. Dairy Sci. 64:2476.

9 Powell, R. L., and H. D. Norman. 1976. Procedures for approximating components of Predicted Difference. Page 41 in USDA Prod. Res. Rep. No. 165, Washington, DC.

10 Powell, R. L., and H. D. Norman. 1986. Genetic and environmental differences between registered and grade Holstein cows. J. Dairy Sci. 69:2897.

11 Powell, R. L., H. D. Norman, and F. N. Dickinson. 1976. The USDA-DHIA Modified Contemporary Comparison Cow Index. Page 35 in USDA Prod. Res. Rep. No. 165.

12 Powell, R. L., H. D. Norman, and R. M. Elliott. 1981. Different lactations for estimating genetic merit of dairy cows. J. Dairy Sci. 64:321.

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

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

15 Snader, J. C. 1987. Look it up faster with hashing. Byte 12:1:129.

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

17 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.

18 Wiggans, G. R., and F. N. Dickinson. 1985. Standardization of NCDHIP dairy cattle lactation records. Natl. Coop. Dairy Herd Improvement Prog. Handbook, Fact Sheet G-2.

19 Wiggans, G. R., and I. Misztal. 1987. Supercomputer for animal model evaluation of Ayrshire milk yield. J. Dairy Sci. 70:1906.

20 Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 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.

21 Wiggans, G. R., J. W. J. Van Dijk, and I. Misztal. 1988. Genetic evaluation of dairy goats for milk and fat yield with an animal model. J. Dairy Sci. 71:(in press).

APPENDIX

A system appropriate for routine animal model evaluations and with provision for processing breeds successively was developed. The main steps are 1) processing of yield data, 2) processing of pedigree data, 3) preparation for computation of solutions, 4) iteration until satisfactory convergence, 5) calculation of R, 6) computation of supplemental evaluations, and 7) adjustment of breeding value to a genetic base and calculation of transmitting ability. These steps are illustrated in Figure 2.

Processing of Yield Data

New and corrected records are combined with records present for the evaluation 6 mo earlier (master file). A corrected record replaces the previous incorrect version. Ability to correct records eliminates need for replacing all data for each evaluation. Lactation records are converted to a 305-d, twice daily milking, mature-equivalent basis in the preceding data acquisition and maintenance phase. Herd-cow records not including a first lactation are separated. Records are excluded if the cow is more than 96 mo at calving or parity is greater than 5.

Through a hashing scheme (4, 5, 15), identification numbers are converted to numbers that serve as subscripts for storing information for an animal in computer memory. The hashing process converts registration and eartag numbers to numbers in the desired range. Eartag numbers are accommodated by using the extended binary coded decimal interchange code (EBCDIC) equivalent for all identification. In this implementation of hashing, identification numbers are converted initially to hash numbers by calculating the modulus. Range for hash numbers is total number of animals plus 20% more for computational efficiency. If a collision occurs (i.e., number has been used previously), 17 is added and the new number checked. Adding and checking continues until an

Update ——→ | 1. Process yield | ←——→ MGT
yield | | groups
| Merge update records |
| Assign MGT groups | ←——→ Master
| Convert ID to hash numbers | yield

                    Grade   Hash   Later
      Yield   pedigree   table   cows
                     (S)

Update ——→ | 2. Process pedigree |
pedigree | |
| Merge update records | ←——→ Master
Hash ←——— | Find parents of parents | pedigree
table | Assign some parents to groups |
| Add hash number to later only records |

                              Equation
              ———— Female   numbers
          ——— Male   pedigree   Parents
           pedigree   Later
                                                    cows

→ | 3. Prepare for iteration |
| |
| Assign equation numbers |
| Compute diagonal elements |
| Prepare starting values | ←

      Yield Parents DIAG PE HS   BV

| 4. Iterate |
| |
| Obtain solutions for all model elements |
| Get diagonals and ARHS for later only evaluation |

                                    Later   Later
     MGT   PE   HS   indicator   cows

→ | 5. Calculate R | ——→ PE-HS
| | ——→ BV cow
| Compute R | ——→ BV sire
| Associate hash numbers with evaluations | ——→ BV unknown
| ←   parent group

     BV   BV   BV   Later
     cow   bull   dam   R   cows (S)

→ | 6. Compute supplemental evaluations |
| |
| Calculate evaluations from later herds and records | ——→ Sire R
| Add R |
| Convert hash numbers to ID |

            BV   BV
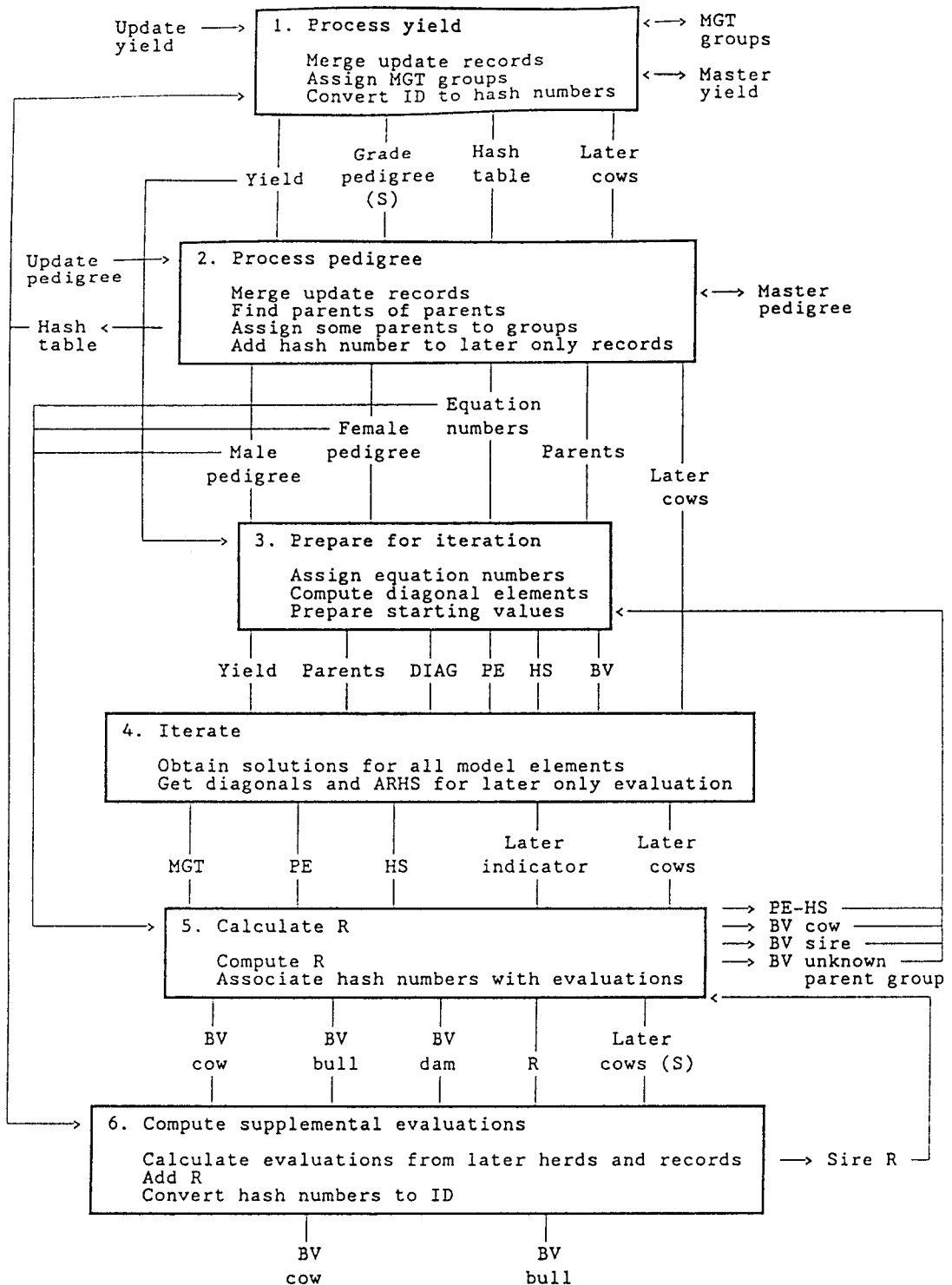   cow   bull

Figure 2. Main operational steps for animal model evaluation system (ARHS = adjusted right-hand sides, BV = breeding value, DIAG = diagonal, HS = herd-sire, ID = identification, MGT = management, PE = permanent environment, R = repeatability, and S = sort).

available number is found. The constant, 17, ensures that any remaining available number will be found. Identification numbers of cow, sire, and dam are hashed and a count of progeny accumulated. The hash number is assigned permanently to allow matching of animals with their solutions from the previous evaluation.

Each lactation record is assigned to a management group. The groups are defined by year-month of calving (2-mo periods), registry status (registered or grade), and parity (first or later). If management group size is less than five, grouping requirements are relaxed successively until a group size of five is reached: 1) season is increased to 4 mo, 2) registered grade distinction is removed, 3) season is increased to 6 mo, 4) parity distinction is removed, and 5) season is increased to 12 mo in increments of 2 mo. If removing parity distinction, required group size is reduced to three. If a management group with more than one record cannot be formed, that record is deleted. If a cow's first lactation record is lost because of no management group mates, later records for the cow are included in evaluation of cows without first lactation records.

Lactation data for protein evaluations also are selected from those records meeting requirements for milk and fat. Later lactation records are included even if no first lactation protein data are available. Management group assignments for milk and fat are applied to protein data. Protein testing usually is applied to all cows in a herd; therefore, if any cows in the management group have protein data, all would, except for those in transition periods.

Pedigree data from grade cows excluded because of a missing first lactation record or no management group mates are collected for use if the cow is the dam of an animal being evaluated.

Processing of Pedigree Data

New pedigree data including sire, dam, and birth date for males and females are combined after receipt from breed associations, sorted in descending birth year order, and merged with the master pedigree file. Pedigree data for grade cows are available only from yield data. For cows being evaluated, pedigree data are associated with yield data. Pedigree data collected from yield data for grade cows excluded because of a missing first lactation record or no management group mates are merged with pedigree data for registered animals. This file identifies parents of all registered animals. This file identifies parents of all registered

cows and bulls and grade cows with excluded yield data. By matching this pedigree file with parents of cows with yield data, grandparents are identified. The file is sorted from youngest animal to oldest. If information for an animal is found in the pedigree file, its parents are added to the group of animals being matched with the pedigree file. In this way, pedigree of every animal is searched back to unknown ancestors. Including both sexes insures that all known ancestors will be found with one pass through the pedigree file. Pedigree data for parents are stored in memory as located. A count of progeny number for each parent is maintained.

Pedigree data are processed in ascending birth year order to determine which parents should be assigned to unknown-parent groups. A sire or dam is assigned to one of these groups if all the following are true: 1) both parents assigned to unknown-parent groups, 2) only one progeny, and 3) no yield record. These animals do not contribute information. While determining which parents should be assigned to unknown-parent groups, a table for translation of hash numbers to equation numbers is created. This second translation allows for the tight sequence of animal and group identification numbers needed for iteration. Sequence order is cows with lactation records, dams without records, sires, and groups. For protein evaluations, searching through all pedigree data is not necessary. Only pedigree information in yield data without protein and pedigree data selected for evaluations for milk and fat yields need to be searched, because animals to be evaluated for protein are a subset of those being evaluated for milk and fat.

Unknown-parent groups are assigned separately for each selection path: sire of sire, dam of sire, sire of cow, and dam of cow. These major divisions allow for differing selection pressure and, therefore, differing average genetic values for different paths. Within path, division is by birth year of the animal with unknown parent. Number of years per group for each selection path is adjusted to equalize group size and avoid groups that are too small for accurate estimation. The first group usually is the largest because it includes base animals. Recent groups tend to be small, especially for bulls, because pedigree information usually is complete and most parents are being evaluated. Unknown-parent groups serve to complete the relationship matrix. Every animal evaluated has two parents, either known or unknown.

Preparation for Iteration

To minimize number of rounds of iteration, predicted breeding value, herd-sire, and permanent environmental effects from the previous evaluation serve as starting values. Average of parent breeding values is substituted for new animals. Each factor evaluated requires that yield data be adjusted by current solutions of all other factors. New solutions for the first effect (management group) are available when evaluating later effects; therefore, previous solutions for management effect are not required.

Data for many herds are accumulated into a single record. This record must be at least large enough to contain the largest herd, and larger size may contribute to computational efficiency. This record has a hierarchical structure with repeating fields as necessary. Fields are number of herds included in record, number of sires in a herd, sire equation number, number of sire's daughters, daughter equation number, equation number of this daughter's dam, number of lactations for this daughter, management group assignment, lactation length weight (18), milk yield, and fat (or protein) yield. This method of coding minimizes memory required.

Diagonals for animal and group effects are computed (20, 21) prior to iteration because they do not need to be recomputed with each iteration. Lactation length weights determine contributions from lactations to diagonals and ARHS. A table of equation numbers for parents of animals without records is prepared to enable computation of relationship contributions to ARHS during iteration. Only these parents' equation numbers need to be stored because equation numbers for animals without records are defined by parents' location in the table; i.e., number of cows with records is added to row subscript to determine animal's equation number. Processing for the data subset with protein information for this and later steps is identical to that for milk and fat. The system is designed to do single-trait evaluations for two traits at the same time. While processing protein, evaluations also are computed for milk from the protein subset.

Iteration

Data and previous solutions are loaded into memory. A simple implementation of the animal model is possible with solutions of all animals in memory. Input and output operations are minimized by retaining solutions for all effects and all data in memory, because no reading or writing is required during iteration. This design required 840 Mb of memory for Holsteins.

Processing is by herd. For effects other than animal, iteration includes successively solving for each effect in the model by subtracting current estimates of all other effects from each record and averaging these residuals to update estimates for current effect (20). Averaging is across all records contributing to that effect. For random effects, variance ratio is included in the denominator. Iteration order is management, permanent environmental, and herd-sire effects. For animal effects, contributions to right-hand sides are collected by subtracting solutions for other effects as above. New solutions are computed after processing all yield data and parents without records. Data are sorted by herd, sire, cow, and calving date. The herd-sire major sort facilitates computation of herd-sire effects. Cow's permanent environmental effect can be estimated after her last record because all her records are together. Records in other herds are considered in a separate evaluation. Contributions to equations for all management effects in the herd are collected while processing the herd. This approach avoids sorting by management group.

Gauss-Seidel iteration, which uses new solutions from each effect in computing following effects, is applied for management, permanent environmental, and herd-sire effects. For animal and group effects, Jacobi iteration, which does not require use of new solutions during current round of iteration, is applied to simplify incorporation of relationship contributions. In this method, a percentage of difference between estimates for the two previous rounds is added to the current estimate. Progress of iteration is monitored by computing a convergence criterion—sum of squared differences between solutions for successive rounds divided by sum of squared solutions for current round (20).

Repeatability and Supplemental Information

Equation numbers are translated back to hash numbers. Starting values for evaluations 6 mo later are prepared for breeding value, permanent environmental, and herd-sire effects. Averages for lactation records and management effects are

computed; PPA (21) are calculated and one record per bull, dam without records, and cow written.

Calculation of R involves processing cow data to collect contributions to accuracy of evaluations of the cow and her sire and then computing animal R with parent R incorporated. Computational method is similar to present MCC approximation for R (2, 9, 11). The MCC method was retained because no indication of accuracy is generated during animal model calculations; therefore, no obvious successor method exists. Because no absorption is done, diagonal elements of the coefficient matrix are not a good basis for approximation of prediction error variance.

In computing R, the following variance components scaled to a phenotypic variance of 1 are used: $\sigma_g^2 = .2$ (genetic), $\sigma_c^2 = .14$ (herd-sire), $\sigma_p^2 = .16$ (permanent environmental), and $\sigma_e^2 = .5$ (residual). This results in $h^2 = .2$ (heritability), $\sigma_s^2 = .05$ (sire), and $\sigma_x^2 = .31$ (cow variance excluding sire contributions). Yield data are processed one herd at a time. Each herd is processed twice, first to collect information for average sire R of management group mates and second to compute contributions to information on cows and sires.

Weighted sire R of other cows with records in a management group is computed by accumulating lactation length weights (w_{len}), product of sire R, and w_{len} ($R*w_{len}$). The number of sires (n_s) for each management group also is accumulated. A default R of .3 is used for bulls not previously evaluated. For each lactation, contribution from that lactation is subtracted to compute weighted average sire R of management group mates ($\overline{R}_{mgt}$):

$$\overline{R}_{mgt} = [\Sigma(R*w_{len}) - R*w_{len}]/[(\Sigma\, w_{len}) - w_{len}]$$

To measure information from one lactation, lactation weight (w_{lac}) is calculated as:

$$w_{lac} = 1/\{(\sigma_e^2/w_{len}) \\ + [\sigma_s^2(1 - \overline{R}_{mgt}) + \sigma_c^2]/(n_s - 1) \\ + (\sigma_e^2 + \sigma_s^2)/[\Sigma\, w_{len}] - w_{len}]\}$$

If only one sire has daughters in the management group, $n_s - 1$ is set to 1 to avoid division by 0.

To measure information on a sire from a daughter, daughter weight (w_{dau}) is calculated as:

$$w_{dau} = 1/[\sigma_x^2 + (1/\Sigma\, w_{lac})]$$

For each sire, w_{dau} are accumulated across herds. Herd-sire weights (w_{hs}) are calculated as:

$$w_{hs} = 1/[\sigma_c^2 + (1/\Sigma\, w_{dau})]$$

where summation is over daughters in that herd. These weights are accumulated over all herds in which a sire has daughters.

Parent identification for a cow is collected while processing the yield file. The bull file is processed to collect parent identification of bulls. Birth year is collected at the same time. Then, following the procedure of Powell (7), R is calculated with contributions from parents included. First, accuracy (acc) from progeny information for bulls and from own records for cows is computed. For bulls:

$$acc = \sigma_s^2/[\sigma_s^2 + (1/\Sigma\, w_{hs})]$$

and for cows:

$$acc = h^2/[.5 + (1/\Sigma\, w_{lac})]$$

Contribution to accuracy from parents is R of parent average (R_{pa}), which is sum of parent R divided by 4. A weighting factor (w_R) is calculated as:

$$w_R = (1 - R_{pa})/[(1/acc) - R_{pa}]$$

R for the animal is calculated as:

$$R = w_R + (1 - w_R)R_{pa}$$

For dams with daughters, an additional computation includes daughter contribution. Powell (8) suggested a method to include daughter information in the MCC Cow Index and the associated R. For the animal model, interest is only in R; therefore, formulas are simplified as follows:

$$R^* = w_{dam}R + 1 - w_{dam}$$

where R^* is R with daughter information included and w_{dam}, a weighting factor for including daughter information in dam R, is calculated as:

$$w_{dam} = (1 - .25n_{dau}B)/(1 - .25R(n_{dau})B]$$

with n_{dau} = number of daughters with first lactation records and

$$B = 4/(9 + 20/\Sigma\ w_{lac} + n_{dau})$$

If each daughter is assumed to be in a separate herd and have $\Sigma\ w_{lac}$ = 1.7, then B = $4/(9 + 20/1.7 + n_{dau})$ = $4/(20.765 + n_{dau})$,

$$
\begin{aligned}
w_{dam} &= \{1 - .25n_{dau}[4/(20.765 + n_{dau})]\}/ \\
&\quad \{1 - .25R(n_{dau})[4/(20.765 + n_{dau})]\} \\
&= [1 - n_{dau}/(20.765 + n_{dau})]/ \\
&\quad [1 - R(n_{dau})/(20.765 + n_{dau})] \\
&= 20.765/[20.765 + n_{dau} + R(n_{dau})]
\end{aligned}
$$

and:

$$
\begin{aligned}
R^* &= \{20.765/[20.765 + n_{dau} + R(n_{dau})]\}R \\
&\quad + 1 - 20.765/[20.765 + n_{dau} + R(n_{dau})] \\
&= [20.765R + n_{dau} - R(n_{dau})]/ \\
&\quad [20.765 + n_{dau} - R(n_{dau})] \\
&= [20.765R + n_{dau}(1 - R)]/ \\
&\quad [20.765 + n_{dau}(1 - R)]
\end{aligned}
$$

Calculation of R is repeated so that new R_{pa} can contribute to progeny. This second round is necessary because animals are not sorted by birth year. For the first round of calculation, current parent R are used if they have been computed already. If progeny R is calculated before parent R, an R of .3 is used for dams; for sires, R from the previous evaluation is used if available, and otherwise an R of .3 is used. The second round uses current R for all parents.

Averages of breeding value and R are computed separately for cows and bulls for each birth year. Average breeding value of cows born in 1984 is used to set the genetic base. Average of first lactation records of cows born in 1984 is required for computing evaluations for fat and protein percentages.

Supplemental Evaluations

Hash numbers are translated back to registration or eartag numbers. Evaluations for cows without first lactation records and evaluations including records in later herds are computed using management group solutions already computed. If the cow has an evaluation as a dam without records or based on her first herd, existing ARHS and diagonal are augmented by contribu-

tions from later records. For cows with a first lactation record, ARHS is adjusted by adding weighted solutions for herd-sire and permanent environmental effects associated with records in the first herd.

For cows with records in more than one herd, information is collected from all herds. For each such cow, a herd-sire effect is estimated for each herd assuming that the cow is the only daughter in that herd. Simultaneous solutions are computed for animal and permanent environmental effects. These steps are repeated to improve their estimates. For bulls, averages for daughter yields and management group effects are computed. These values include only data from the first herd. The R are associated with records. A file of bull R is created for computing R for the next evaluation.

Genetic Base and Transmitting Ability

Average breeding value of cows born in 1984 is subtracted from predicted breeding value to impose the genetic base and the result halved to obtain PTA. The base adjustment also is subtracted from average breeding value of parents and the result halved to form parent averages (21). The PPA (21) is not halved, but the base adjustment is subtracted. The base adjustment is added to average management effects so that overall sum of effects in the model equals records. Effect of this adjustment is to transfer amount of base adjustment from breeding value to management effect.

The PTA for fat and protein percentages are computed. For each breed, PTA for fat percentage is calculated as:

$$
PTA_F\% = [(PTA_F + F_{84})/PTA_M + M_{84} - (F_{84}/M_{84})]100
$$

where $F\%$ = fat percentage, F = fat yield, F_{84} = average first lactation mature-equivalent fat yield of cows born in 1984, M = milk yield, and M_{84} = average first lactation mature-equivalent milk yield of cows born in 1984.

Protein and milk/fat results are merged, and the PTA for protein percentage is calculated similarly to that for fat percentage using PTA for protein and milk yields from the protein data subset.

A PTA for protein yield for the full data set is derived from the protein percentage evaluation

computed from the protein data subset and the milk yield evaluation from the full data set:

$$PTA_{P(full)} = (PTA_P \%/100)(PTA_{M(full)} + M_{x4}) - P_{x4}$$

where P = protein yield, $P\%$ = protein percentage, P_{x4} = average first lactation mature-equivalent protein yield of cows born in 1984, and full refers to the full data set.

To aid in finding potential errors in data or computer programs, current evaluations are compared with previous evaluations to determine if large changes have occurred in transmitting ability relative to change in amount of available information.