Mathematical Representations of Correlations Among Yield Traits and Somatic Cell Score on Test Day

H. D. NORMAN, P. M. VANRADEN, J. R. WRIGHT, and L. A. SMITH
Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

Prediction of lactation yields and accuracies of yields for use in genetic evaluation can be improved by including information from test day correlations, especially for milk recording plans that vary in the numbers of milk weights recorded and component samples taken. Daily milk weights for 658 lactations of Canadian cows and monthly test records of milk, fat, and protein yields and somatic cell scores for 500,000 lactations of US cows were used to estimate phenotypic correlations between test days within herd-year. Correlations between daily yields for a designated interval between test days generally were highest for midlactation and were lowest for early and late lactation. Regression (two linear, two quadratic, and interaction effects) on mean DIM and interval between test days predicted correlations with a squared correlation of 0.94 for daily milk yields. Similar relationships were found for US monthly data. Variation in sampling was reduced, computer memory was minimized, and positive definiteness was guaranteed by fitting regressions on simply defined sources of correlation. An autoregressive matrix represented the within-trait correlations very well. The equations developed could be used to derive covariances and, subsequently, to estimate lactation yields and accuracies from combinations of individual daily milk, fat, and protein yields and somatic cell score.
(**Key words:** best prediction, correlation, milk yield, test day)

Abbreviation key: DCR = data collection rating.

INTRODUCTION

More than 30 new test plans for milk recording were proposed and introduced in the US between 1989 and 1995 (P. Dukas, 1995, personal communication) in response to market requests. Some new plans are being used in a single state, whereas others have been implemented in nearly all states. The various test plans differ widely in the numbers of milk weights recorded and component samples taken. Some plans have fewer milk weights and samples, which makes accurate prediction of yields more difficult. Electronic capture of milk weights has stimulated the formation of new test plans and in many cases has provided information more frequently for milk yields even though component samples were fewer.

The most profitable test plan for each producer is a balance between the value of additional information and the cost of additional testing. In theory, an optimal sampling plan would have more tests scheduled in early lactation, especially near the time when milk yield is expected to peak, instead of at equal intervals. In practice, all cows in a herd are sampled at the same time, regardless of their stage of lactation. The optimal sampling plan for component yields would be similar to that for milk yield, whereas the optimal plan for estimation of component percentages for the entire lactation could be different.

Until recently, predictions of 305-d lactation yield included a cumulative yield calculated with the test interval method (8, 11). From January 1980 to November 1998, records in progress were extended to 305 d for use in USDA-DHIA genetic evaluations by including the cumulative yield and the yield on last test day in a regression equation (14). For records of ≤155 d, a regression coefficient for herd mean also was included. However, equations that include all test day yields can provide more accurate estimates of lactation yield, especially if correlations between the individual test days are used in the prediction equations (9).

Analysis of test day data requires estimation of many parameters. VanRaden (12) developed a procedure with best prediction properties that included a correlation matrix between test day yields to predict lactation yields (6). Norman et al. (5, 7) showed that this procedure had 4 to 10% less error in prediction of 305-d milk yield than did the test interval method (8) for monthly, a.m.-p.m., and trimonthly (or quarterly) testing.

The best prediction procedure provided an indication of the accuracy of prediction, which was referred to as a data collection rating (**DCR**). A DCR is the squared correlation of predicted and true lactation yields multiplied by 100 and divided by the squared correlation for a standard, supervised plan with 10 monthly tests

Received November 23, 1998.
Accepted May 28, 1999.

(13). Monthly testing produced a DCR of 100 for a 305-d lactation, and the accuracy of records from all other test plans could be compared with this DCR. Certain minimum levels of DCR have been selected by some breed associations as a criterion necessary for recognition of cows with high yield.

A half-stored correlation matrix for 305-d milk yields excluding diagonals contained 46,360 unique elements. If the matrix also included fat and protein, 418,155 correlations were present; if the matrix included SCS as well, the number of correlations was 743,590. Although computers today have the memory, storage, and speed to process large matrices, far fewer parameters may be required to represent those correlations mathematically.

Covariance functions were introduced by Kirkpatrick et al. (4) to represent large or even infinite matrices. A large matrix might be represented by the 10 or so largest eigenvalues and a 10×1220 matrix of eigenvectors, but many parameters are still required, which makes interpretation difficult. Jamrozik and Schaeffer (3) reduced a 305×305 matrix to a 5×5 variance matrix using random regression, and Gengler et al. (2) used covariance functions and Legendre polynomials to obtain a similar 3×3 system of equations. Simpler functions could reduce sampling variation for prediction of individual correlations and make the resulting computer programs more portable to others in research and industry. Industry groups have requested software to derive the correlations, lactation predictions, and DCR for field use, and the preference of those groups is for uniform and manageable functions. Computer programs were developed by VanRaden (12) to predict lactation yields, including 305-d or 365-d records, using single-trait or multitrait correlation matrices.

Desirable properties of functions are simplicity and reduction of sampling variation, thereby supplying estimates closer to the true values than the individual correlations originally calculated. Functions that are too simple to fit the large matrices adequately could generate biased predictions. Functions that supply correlations between individual daily observations also can supply the correlation between an individual daily observation and the 305-d total. Covariance of daily yield with lactation yield is the sum of 305 daily covariances.

Preliminary examination of the effectiveness of the best prediction procedure was based on correlations among milk yields from Canadian Holstein cows (5, 6, 7). No information was available on whether this procedure was appropriate for fat yield, protein yield, or SCS or for milk yield of US populations. The first objective of this study was to determine the correlations among individual daily yields of milk, fat, and protein and SCS within herd test day or herd-year. The second objective was to determine a method of expressing the relationships among all test day observations accurately and in a manner that minimized computer memory requirements. The final objective was to guarantee that the estimated correlation matrices were positive definite. The resulting parameter estimates then could be used to improve predictions of lactation yields and SCS nationally, regionally, or for on-farm computer use.

METHODS

Correlations Derived from Daily Milk Yield

Data to calculate correlations were daily milk weights for 658 cows from 17 Canadian herds (5, 7). Lactation length was required to be between 250 and 305 d. A milk weight for at least 90% of days in the lactation also was required for each cow. These daily milk weights allowed correlations for every DIM subclass to be based on hundreds of observations.

Correlations were calculated separately within herd-year for each pair of milk weights for the first 305 d of lactation. This approach eliminated the need to delete data for cows without milk weights for each day of lactation. Correlations also were calculated between daily yields and 305-d lactation yield.

A multiple regression analysis was conducted to determine how well the correlations between pairs of test days could be predicted from DIM variables (6). The model was

$$y_{ij} = b_1 \overline{DIM}_{ij} + b_2(\overline{DIM}_{ij})^2 + b_3(DIM_j - DIM_i) + b_4(DIM_j - DIM_i)^2 + b_5[(\overline{DIM}_{ij}\,(DIM_j - DIM_i)] + e_{ij}$$

where y_{ij} = 46,360 correlations within herd-year between daily yields on test days i and j (i = 1, 2, ..., 304; and j = i + l, i + 2, ..., 305), b = a partial regression coefficient, $\overline{DIM}_{ij}$ = mean DIM for test days i and j, $DIM_j - DIM_i$ = interval between test days i and j, and e_{ij} = random residual. The squared correlation between the predicted and observed correlations for pairs of test days indicated how effectively the prediction equation regenerated the correlations.

Daily data for fat, protein, and SCS were not available. The correlations developed from Canadian data also may not represent US correlations exactly. Therefore, a larger set of US data was examined. New techniques were also derived to guarantee that the correlation matrix including all four traits would remain positive definite.

Correlations Derived from Monthly Yield and SCS

Data to calculate correlations for all four traits at each DIM were obtained from US herds that had milk weights

and component samples taken at approximately 30-d intervals. A lactation record was included if 1) fat yield, protein yield, or SCS was reported in addition to milk yield, 2) parity was less than six, 3) the record was made in a single herd, 4) at least five tests were reported per lactation, 5) the lactation length was 250 d, 6) only twice daily milking was reported, 7) calving date was before 1997 (to exclude records in progress), and 8) the herd was enrolled in DHI test plan 00 or 20 (1) for which all milkings in a 24-h period are weighed and sampled. Because of computational limitations, only the first 500,000 lactation records in Animal Improvement Programs Laboratory (USDA, Beltsville, MD) database (based on herd number) that met those requirements were used to study the correlations between test days.

The use of monthly data to calculate correlations between daily observations for milk, fat, protein, or SCS would result in correlations based on highly variable number of observations. The number would be limited when the interval between test days was not approximately a multiple of 30 d and especially when the interval was a small number. Thus, the accuracy of individual correlations would differ considerably, and those correlations with few observations for the pair of test days would have large sampling variation. However, if the true correlations between test days were continuous and changed in a predictable manner, prediction of those correlations by variables based on DIM groups rather than specific DIM might be possible without losing much precision. The use of DIM groups would result in more observations in the calculation of each correlation. The correlations based on the smaller numbers would remain less accurate than those based on the larger numbers but in most cases would be more accurate than from the daily alternative. Therefore, prior to calculation of correlations, test days were divided into 61 groups based on 5-d increments (1 to 5 DIM, 6 to 10 DIM, ..., 301 to 305 DIM). Correlations for the test day group then were calculated in the same manner as for the Canadian daily data except that correlations for first and later parities were calculated separately.

A multiple regression analysis similar to that for Canadian daily data also was conducted, but DIM variables were based on DIM group rather than DIM. Groups with less than 100 observations were excluded from the analysis.

Because correlations and covariance matrices of large dimensions sometimes have some poorly estimated elements and because matrices must be at least positive definite to be useful, procedures to guarantee positive definite results were developed and applied. A sum of positive definite matrices must be positive definite because a sum of quadratic forms cannot be less than the smallest quadratic form in the sum. Also, the Kronecker product of two positive definite matrices must be positive definite because the eigenvalues of the Kronecker product are the products of the eigenvalues of the two individual matrices (10). Thus, simple components of variance could be combined into a composite that fit well with the estimated correlation or covariance matrix. Several sources of variation could be defined, and matrices for those sources then could be combined by regressing elements of the estimated correlation matrix on the elements of each source. Variance sources that receive negative regression coefficients would be removed from the model to preserve a positive definite result.

For analysis of US monthly data, a 4×4 correlation matrix (**T**) across traits was assumed to equal the phenotypic correlation matrix among lactation records for milk, fat, and protein yields and SCS. A series of 365×365 matrices described the correlation patterns within traits. The matrices were set up to be positive definite or positive semi-definite across the range of 1 to 365 d. Test days >305 were expected to be included when the procedure was implemented even though the research file was limited to 305 d. Each of the matrices had a simple definition.

Seven possible sources of correlation were defined and were fitted to the observed correlations. Sources included an identity (**I**) matrix; an intercept (**J**) matrix; a first order autoregressive (**E**) matrix; matrices indicating increased covariance among the first (**F**), middle (**M**), or last (**L**) days of a lactation; and a persistency (**P**) matrix to indicate the tradeoff between early and late lactation yields. Matrix **J** was defined as $\mathbf{1}(\mathbf{1}')$ where each element of $\mathbf{1} = 1$. Elements of **E** represented an exponential decline in the correlation for each day between test days. Matrices **F**, **M**, and **L** were each constructed using a diagonal matrix (**V**) to specify variation in common and a residual ($\mathbf{I} - \mathbf{V}$) to specify variation not in common. A positive definite result was guaranteed by limiting the diagonals of **V** to be >0 but <1 so that **V** and $\mathbf{I} - \mathbf{V}$ were both positive definite. Matrix **P** included covariance proportional to the difference from 182.5 d (half of a 365-d lactation) plus residual uncorrelated variation similar to **F**, **M**, and **L**.

For test days i and j, mathematical definitions for the matrices were $E_{ij} = r^{|i-j|}$, where $0 < r < 1$; $\mathbf{F} = \mathbf{1V1}' + (\mathbf{I} - \mathbf{V})$, where $V_{ii} = 1 - i/365$; $\mathbf{M} = \mathbf{1V1}' + (\mathbf{I} - \mathbf{V})$, where $V_{ii} = (i - i^2/365)/365$; $\mathbf{L} = \mathbf{1V1}' + (\mathbf{I} - \mathbf{V})$, where $V_{ii} = i/365$; and $\mathbf{P} = \mathbf{pp}' + \mathbf{I} - \text{diag}(\mathbf{pp}')$, where $p_i = (i - 182.5)/182.5$, and diag indicated the diagonal elements. An optimal r was determined by maximizing the squared correlation of actual and estimated elements by simple trial and error. Each matrix was then multiplied by its regression coefficient (b_1, b_2, ..., b_7) and summed into a matrix of correlations within trait (**B**):

$$\mathbf{B} = b_1\mathbf{I} + b_2\mathbf{J} + b_3\mathbf{E} + b_4\mathbf{F} + b_5\mathbf{M} + b_6\mathbf{L} + b_7\mathbf{P}.$$

A complete 1460 × 1460 correlation matrix (**C**) was obtained as the Kronecker product

$$\mathbf{C} = \mathbf{B} \otimes \mathbf{T}$$

where $C_{ik,jl}$ = correlation of trait k at DIM_i with trait l at DIM_j.

Sources of correlation were removed if regressions were negative or if they contributed very little to the squared correlation. For comparison, the regression function described for Canadian daily data was used also to obtain a **B** and then substituted into the Kronecker product with **T**, but this approach did not guarantee a positive definite result.

An improved estimate of each T_{kl} was obtained by dividing each correlation estimate ($C_{ik,jl}$) by the corresponding element of **B** and obtaining the mean for each trait combination

$$T_{kl} = (C_{ik,jl}/B_{ij})/n$$

where summation was across all correlations involving traits k and l, and n = number of correlations.

Because the traits analyzed might not all follow the same correlation pattern across DIM, a separate **B** could be fitted for each trait or for each combination of traits. However, this procedure would require many more parameters and would not guarantee a positive definite result. Instead, matrix **S** was defined to allow the correlation pattern within SCS to differ from the other traits. For milk, fat, and protein yields, elements of **S** were set equal to those of **E**. For the SCS submatrix, an identity variance matrix was substituted. Daily fluctuations of SCS could then be larger and independent of fluctuations of the other traits because off-diagonals between yield and SCS were set to 0. Thus, **S** allowed the model to account for less uniformity for SCS data than for yield data.

After prediction of correlations by this method, DCR were derived for 19 current or anticipated US test plans.

RESULTS

Correlations Derived from Daily Milk Yield

Correlations between milk yields at designated test days and test days at intervals of 30, 60, and 90 d later are shown in Table 1. For both observed and predicted correlations, the greater the interval was between test days, the lower the correlations were. Correlations tended to be lowest for test days during early and late lactation and highest for test days during midlactation. The lower correlations for early and late lactation indicated the value of more frequent weighing and sampling during those stages of lactation. However, not only the correlations but also the variances of test day yields affect the accuracy of the estimated lactation yield.

When the correlations among test day yields were predicted based on linear, quadratic, and interaction effects of mean DIM and interval between test days, all five variables were significant ($P < 0.0001$) and produced a squared correlation of 0.936. Prediction based solely on the linear effect of interval between test days gave a squared correlation of 0.907. If the correlations between test days during the first 20 d of lactation were excluded, all five variables again were significant ($P < 0.0001$), and the squared correlation increased to 0.964. However, the interaction between mean DIM and the interval between test days resulted in a squared correlation of 0.940, and addition of the linear effect of interval between test days increased the squared correlation to 0.959. Those two variables accounted for almost all of the squared correlations. The predicted correlations in Table 1 were based on linear and quadratic effects of mean DIM because the interval between test days was held constant at 30, 60, or 90 d.

Test day yields early and late in lactation had the lowest observed correlations (<0.60) with 305-d lactation yield (Table 1). Test day yields in midlactation (100 to 200 DIM) had the highest (0.80 to 0.86) correlations with 305-d lactation yield.

Correlations Derived from Monthly Yield and SCS

Few parameters were needed to describe extremely large correlation matrices. Within traits, correlations followed an autoregressive pattern. The best value for parameter r was 0.995 for first parity and 0.992 for a later parity. With those values, the simple Kronecker product $\mathbf{E} \otimes \mathbf{T}$ provided a squared correlation of 0.952 for the 24,496 filled cells in the correlation matrix among all test days and traits of first parity. For a later parity, the squared correlation was 0.945. The correlations that were the elements of matrix **T** are shown in Table 2.

Use of the matrix formulas can be demonstrated using the simple model $\mathbf{C} = \mathbf{E} \otimes \mathbf{T}$. For example, the correlation of milk yield at 60 DIM with protein yield at 90 DIM for a later parity was estimated as the correlation across traits (T_{ij} = 0.88) multiplied by the correlation within trait ($E_{ij} = 0.992^{30} = 0.786$) to obtain the total correlation [$C_{ik,jl} = (0.88)0.786 = 0.69$]. With this simple model, an intercept was not included, and the regression on **E** (b_3) was set equal to 1 because its estimated value was 1.005.

Slightly higher squared correlations were obtained by adding other terms to the model. For first parity, a

TABLE 1. Observed and predicted[1] correlations of milk yield on test days at designated DIM with milk yield on test days 30, 60, and 90 d later and observed correlations of test day with 305-d lactation yield for Canadian daily data.

	DIM + 30 d		DIM + 60 d		DIM + 90 d		Observed correlation with 305-d lactation yield
DIM	Observed	Predicted	Observed	Predicted	Observed	Predicted	
5	0.64	0.66	0.58	0.56	0.47	0.44	0.45
10	0.72	0.67	0.61	0.57	0.50	0.46	0.54
15	0.68	0.68	0.59	0.58	0.56	0.47	0.52
45	0.80	0.73	0.75	0.64	0.68	0.55	0.74
75	0.77	0.76	0.72	0.69	0.66	0.60	0.78
105	0.75	0.79	0.71	0.71	0.61	0.62	0.81
135	0.84	0.79	0.66	0.72	0.73	0.63	0.80
165	0.79	0.79	0.68	0.70	0.49	0.61	0.86
195	0.76	0.77	0.56	0.67	0.52	0.57	0.82
225	0.73	0.74	0.62	0.62	. . .	0.50	0.75
255	0.68	0.69	. . .	0.56	. . .	. . .	0.57
285	. . .	0.64	. . .	. . .	. . .	. . .	0.44
295	. . .	. . .	. . .	. . .	. . .	. . .	0.41
300	. . .	. . .	. . .	. . .	. . .	. . .	0.39

[1]Predictions based on linear and quadratic effects of mean DIM for the pair of test days.

squared correlation of 0.964 resulted from a model including **E**, **L**, and **S** to estimate the test day correlations. For later parity, a squared correlation of 0.970 resulted from a model including **E**, **M**, and **L**. Some of the other terms (**I**, **J**, **F**, and **P**) had negative regression coefficients, and together these terms increased the squared correlation by a total of only 0.009. For first parity, the final model was

$$\mathbf{C} = (0.665\mathbf{E} + 0.256\mathbf{L}) \otimes \mathbf{T} + 0.085\mathbf{S}.$$

For later parity, the final model was

$$\mathbf{C} = (0.737\mathbf{E} + 0.158\mathbf{M} + 0.167\mathbf{L}) \otimes \mathbf{T}.$$

These functions obtained from monthly American data replaced those from Canadian daily data for calculating lactation records and DCR by best prediction (12) beginning in February 1999.

For comparison, the regression equation used for Canadian test day data in a Kronecker product with **T** had almost identical accuracy and gave a squared correlation of 0.963 for first parity. Thus, correlations among test day data for the four traits had similar patterns that could be described with only a few parameters. The 5×5 matrix of Jamrozik and Schaeffer (3) produced a squared correlation of 0.973, equal to the squared correlation of our complete model, but theirs required estimation of more parameters. Also, the 3×3 matrix of Gengler et al. (2) produced a squared correlation of 0.967, approximately equal to our final model, but theirs required two parameters more.

Fewer parameters may be needed with correlation functions than with covariance functions because correlations and variances are estimated separately. Also, correlation functions may reduce the problem of high variances occurring at the beginning or end of lactation. Disadvantages of the correlation sources used here were that the inverses might not be trivial to compute. Surprisingly, five additional parameters to allow yield and somatic cell correlations to differ produced only a 0.005 increase in squared correlation while losing the guarantee of positive definiteness. Matrix **S** produced a 0.002 increase for first parity and almost no increase for later parity but did preserve positive definiteness.

The DCR for a number of test plans are shown in Table 3 as calculated from the correlation matrices from US monthly data. The accuracy of traditional monthly testing was defined to be 100%; therefore, the accuracy of the other test plans was shown in relation to it. The DCR for the individual test plans were nearly identical to those reported by VanRaden (12) for Canadian daily data, a tribute to the robustness of the best prediction procedure. Also, DCR for fat, protein, and SCS were all very similar to milk DCR for lactations with all traits recorded. Because the DCR for daily and labor efficient

TABLE 2. Correlations (matrix **T**) among yield traits and SCS for first (above diagonal) and later (below diagonal) parities for US monthly data.

Trait	Milk yield	Fat yield	Protein yield	SCS
Milk	1.00	0.67	0.91	–0.05
Fat	0.67	1.00	0.74	–0.08
Protein	0.88	0.73	1.00	–0.02
SCS	–0.12	–0.17	–0.11	1.00

TABLE 3. Data collection ratings (DCR) for various test plans.

Test plan	Test days	DCR		
			US parity	
		VanRaden (12)[1]	First	Later
	(no.)			
Daily	305	103	104	104
Labor efficient record				
10-d Mean	100[2]	103	104	104
5-d Mean	50[3]	102	103	103
Monthly supervised milkings				
All[4]	10	100	100	100
2 of 3	10	97	97	97
1 of 2	10	95	95	95
1 of 3	10	90	90	90
Monthly owner-sampler[5] milkings				
All[4]	10	77	75	75
2 of 3	10	75	74	73
1 of 2	10	74	72	72
1 of 3	10	71	69	69
Bimonthly supervised milkings				
All[4]	5	97	96	95
2 of 3	5	92	91	90
1 of 2	5	88	87	86
1 of 3	5	81	79	78
Bimonthly owner-sampler[5] milkings				
All[4]	5	75	73	72
2 of 3	5	72	70	69
1 of 2	5	69	67	67
1 of 3	5	65	63	62

[1]Based on Canadian daily data (5, 7).

[2]A 10-d mean reported in each of 10 mo.

[3]A 5-d mean reported in each of 10 mo.

[4]Milk weights were obtained from all milkings on each test day.

[5]For USDA-DHIA genetic evaluations, owner-sampler tests are assumed to be less accurate than supervised tests (13).

test plans was >100%, the weights given to records from these plans was >1.0 in genetic evaluations.

Table 4 shows DCR based on correlations from US monthly data for first and later parities according to length of the record in progress and frequency of milk recording (all daily milkings recorded or only one of two milkings recorded). Early records in progress were more accurate for first parity than for later parity, regardless of recording frequency. Accuracy increased rapidly throughout the first 6 mo. With those increases in accu-

TABLE 4. Data collection ratings for first and later parities of various lengths with all milkings or half of milkings recorded on test day.

Tests	DIM	All milkings recorded		1 of 2 milkings recorded	
		First parity	Later parity	First parity	Later parity
(no.)					
1	15	35	23	27	18
2	45	49	39	41	32
3	75	61	54	53	47
4	105	71	67	63	59
5	135	80	78	72	70
6	165	87	86	80	79
7	195	93	92	85	85
8	225	97	96	90	90
9	255	99	99	93	93
10	285	100	100	95	95
11	315	100	100	95	95

racy, predictions of lactation yield for individual cows were expected to change substantially between first test and 305-d total. However, the expected variation between genetic evaluations based on records in progress and complete records for individual cows would be considerably less because of a large component of information coming from pedigree data. Variation in genetic evaluations of sires also should be of lesser concern because combining information from several daughters generally reduced sampling variance.

All 1460 eigenvalues of matrix **C** were positive as guaranteed. The eigenvalue mean was 1.006 with a range of 0.001 to 393.09. The simpler function developed earlier for daily milk yield (6) did not provide a positive definite correlation matrix when adapted for use on four traits. The eigenvalue mean was 0.214 with a range of –185.38 to 209.31. Four of the 1460 eigenvalues were negative. In best prediction, use of a **C** that was not positive definite tended to inflate estimates away instead of regressing estimates toward the herd mean for some lactation records. The guarantee of positive definiteness was very valuable for such large correlation matrices.

CONCLUSIONS

Phenotypic correlations among test day observations of milk, fat, and protein yields and SCS can be represented mathematically with simple functions. Relationships between test days that were calculated from daily milk yields of 658 Canadian cows were determined to be appropriate for the US population. Correlations for milk, fat, and protein yields and SCS were also obtained for US cows based on monthly data from 500,000 lactations.

Positive definite correlations were guaranteed by adding together a series of simple, positive definite matrices multiplied by positive regression coefficients. The most important matrix was autoregressive, and the correlation declined slightly for each additional day between observations. When compared with covariance functions, correlation functions required fewer parameters to obtain the same accuracy in representing phenotypic correlations. Instead of a matrix of unknowns, only a series of unknown regressions was estimated.

Correlation functions may be a simpler alternative to covariance functions (4) for use in research. The methods developed to approximate phenotypic correlations might also be useful in genetic evaluation of test day data. The computer programs of VanRaden (12) were updated in February 1999 with the correlation functions developed in this study and are now in use nationally to predict daily yields or lactation totals for milk, fat, and protein yields and SCS and to provide DCR to indicate the accuracy of the predictions.

ACKNOWLEDGMENTS

The authors thank R. K. Moore, Québec Dairy Herd Analysis Service (Ste. Anne de Bellevue, PQ, Canada); L. R. Schaeffer, University of Guelph (Guelph, Ontario, Canada); and T. R. Meinert, formerly of National DHIA (Columbus, OH) for file preparation and transfer of Canadian data to USDA. The suggestions of R. H. Miller, University of Maryland, and S. M. Hubbard, Animal Improvement Programs Laboratory, USDA, for manuscript improvement are appreciated.

REFERENCES

1 Benson, R. H. 1985. The NCDHIP record plans. Handbook, Fact Sheet A-4. Natl. Coop. DHI Prog. Handbook, Fact Sheet A-4, Ext. Serv., USDA, Washington, DC.
2 Gengler, N., A. Tijani, G. R. Wiggans, and I. Misztal. 1999. Estimation of (co)variance function coefficients for test day yield with an expectation-maximization restricted maximum likelihood algorithm. J. Dairy Sci. 82:(Aug.). Online only. Accessed Aug. 26, 1999.
3 Jamrozik, J., and L. R. Schaeffer. 1997. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. J Dairy Sci. 80:762–770.
4 Kirkpatrick, M., W. G. Hill, and R. Thompson. 1994. Estimating the covariance structure of traits during growth and aging, illustrated with lactation in dairy cattle. Genet. Res. Camb. 64:57–69.
5 Norman, H. D., P. M. VanRaden, J. R. Wright, and J. S. Clay. 1999. Comparison of test interval and best prediction methods for estimation of lactation yield from monthly, a.m.-p.m., and trimonthly testing. J. Dairy Sci. 82(2):438–444.
6 Norman, H. D., P. M. VanRaden, J. R. Wright, and T. R. Meinert. 1996. Mathematical representation of relationships between daily milk yields. J. Dairy Sci. 79 (Suppl.1):143.(Abstr.)
7 Norman, H. D., J. R. Wright, and J. S. Clay. 1998. Comparison of the test interval method with best prediction for estimating lactation yield. Proc. 6th World Congr. Genet. Appl. Livest. Prod. 23:343–346.
8 Sargent, F. D., V. H. Lytton, and O. G. Wall, Jr. 1968. Test interval method of calculating Dairy Herd Improvement Association records. J. Dairy Sci. 51:170–179.
9 Schaeffer, L. R., and J. Jamrozik. 1996. Multiple-trait prediction of lactation yields for dairy cows. J. Dairy Sci. 79:2044–2055.
10 Searle, S. R. 1966. Matrix Algebra for the Biological Sciences. 1st ed. John Wiley and Sons, Inc., New York, NY.
11 Shook, G. E., L. P. Johnson, and F. N. Dickinson. 1980. Factors for improving accuracy of estimates of test-interval yield. Dairy Herd Improvement Lett. 56(4):9–24.
12 VanRaden, P. M. 1997. Lactation yields and accuracies computed from test day yields and (co)variances by best prediction. J. Dairy Sci. 80:3015–3022.
13 VanRaden, P. M., G. R. Wiggans, H. D. Norman, and R. L. Powell. 1997. Changes in USDA-DHIA genetic evaluations (February 1997). AIPL Res. Rep. CH7(2–97) Anim. Improvement Progr. Lab., Beltsville, MD. http://aipl.arsusda.gov/memas/pdf/chng972.pd. Accessed Aug. 26, 1999.
14 Wiggans, G. R., and R. L. Powell. 1980. Projection factors for milk and fat lactation records. Dairy Herd Improvement Lett. 56(1):1–15.