Data collection ratings and best prediction of lactation yields

J.B. Cole

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

Summary

Best prediction (BP) has been used in the US to calculate lactation yields of milk, fat, protein, and somatic cell score (SCS) from test day data since 1998, when it replaced the test interval method (TIM) used since 1969. It is more complex than TIM but also more accurate, particularly when testing is less frequent. The BP software was recently modified to use breed-, parity, and trait-specific lactation curves, accommodate lactations longer than 305-d, and provide additional output, such as daily yields. Data collection ratings (DCR), which describe the accuracy of lactation records from a variety of test plans, are calculated as a function of each trait's reliability. Predictions of daily yield were validated using daily milk weights from university research herds. Data collection ratings for fat and protein increased when multiple trait BP was used. Many cows can produce pro?tably for >305 DIM, and the revised BP program provides a flexible tool to model records of any reasonable length and testing pattern.

***Key words**: Best prediction, Daily yield, Data collection rating, Lactation yield.*

Introduction

Best prediction (VanRaden, 1997; Cole and VanRaden, 2006) is a computational method derived from selection index that allows test day data to be condensed into total lactation yields and persistencies. Lactation yields have been calculated using BP since February 1998, and the International Committee for Animal Recording (ICAR, 2006) approved BP for lactation record calculation in 2002. Best prediction requires only phenotypic covariances, and assumes that herd means and variances are known. It is simpler to compute than a test-day model but still provides accurate lactation records for a variety of test plans. Reverse prediction may be used to obtain daily yields from lactation yields and persistencies. Both single- and multiple-trait prediction are supported. However, the original BP programs could not accommodate lactations longer than 305-d, did not provide daily yields, and modelled covariances among test days using functions that did not have simple biological interpretations.

Materials and methods

Test-day data for six dairy breeds were extracted from the national dairy database (NDDB), and lactation lengths were required to be d”500-d (d”800-d for SCS). Average yield and SD at any day in milk (DIM) were estimated by fitting 3-parameter Wood's curves to milk, fat, protein data and 4-parameter exponential functions to SCS data for means and SD of 15- (d”300 DIM) and 30-d (>300 DIM) intervals. The resulting breed-, parity-, and trait-specific curves defined a set of standard curves to which individual cow data were compared.

Best prediction requires correlations among individual DIM within and between traits. Those correlations were estimated using an autoregressive matrix to account for biological changes over the course of a lactation and an identity matrix to model daily measurement error. Autoregressive parameters were estimated separately for first and later parities. Correlations between traits were modified so that correlations between SCS and other traits may be non-zero.

The new lactation curves and correlation functions were validated by extracting test day data from the NDDB, estimating 305-d yields using the original and new programs, and correlating those results. Daily BP of yield were validated using daily milk weights from on-farm meters in university research herds.

The accuracy of lactation records from a wide variety of test plans can be compared using a data collection rating (DCR). Data collection ratings are computed from test-day data using best prediction (VanRaden, 1997) as the squared correlation of estimated and true yields multiplied by a factor of 104 to give monthly testing a rating of 100 and daily testing a rating of 104. Separate DCR are calculated for milk yield, components yield, and somatic cell score. Herd DCR are calculated as the DCR for a cow tested in a given herd and whose lactation reached 305 days in milk on the herd's most recent test day. The ratio of the error variance from daily testing to the error variance from less complete testing is also computed from the DCR and used to weight lactation records in the animal model; it differs from the squared correlation and DCR because genetic and permanent environmental effects are excluded from the variance ratio.

Results and discussion

The lactation curves used for BP need to accommodate cows with long lactations in a reasonable manner, but it is more important that the first 365-d of lactation be modelled with great precision than, say, days 600 to 700. The use of 500-d lactations represents a reasonable compromise between cows milking only to 305-d and those milking longer, and across breeds only about 5% of cows have lactations longer than 500 DIM. Records up to 999-d in length were analyzed to demonstrate that the new curves and correlation functions can accommodate lactations of any reasonable length.

Lactation curves for milk, fat, and protein yield were similar. First parity cows in all breeds have lower peak yields and flatter lactation curves on average than later-parity cows, i.e., are more persistent. Later-parity cows consistently outperform first parity cows for daily milk, fat, and protein until late in lactation.

First parity cows had consistently lower SCS than older cows for all breeds but Milking Shorthorn, where the shape of the mature cow curve was clearly affected by a few extreme observations. The curves for Holstein cows were similar to those published by Schutz *et al*. (1995), although our curves for first parity cows did not cross the mature cow curves near 305-d as did theirs. Reents *et al*. (1995) found large differences between first and later parity lactation curves for Canadian Holsteins, as was the case with US Holsteins in this study, but they found that the older cows had lower SCS than first parity cows until about 100 DIM. These differences underscore the challenge of correctly modelling SCS curves.

Autoregressive parameters (r) were estimated separately for first- (r=0.998) and later-parities (r=0.995). These parameters were slightly larger than previous estimates due to the inclusion of the identity matrix.

Correlations within breeds and parities ranged from 0.900 to 0.988 for 305-d milk yield. The highest correlation among actual and predicted daily yields was 0.988, and the lowest was 0.015 on day 1 of lactation, which may be due to calving-related disorders that are not accounted for by BP. As expected, correlations were lowest early in lactation and highest later in lactation.

Data collection ratings for a number of testing plans are presented in table 1. Variation is affected by the test plan. Measurement errors are decreased if several daily yields are averaged as in labor efficient records. Errors are increased in AM-PM testing because only a fraction (2/3, 1/2, or 1/3) of the cow's daily yield is

Table 1. Data collection ratings (DCR), squared correlations of true with estimated milk yields, and lactation weights for various 305-day records.

Test plan	Test days (no.)	Lactation weight (%)	Squared correlation (%)	DCR
Daily	305	100	100	104
Labor efficient record				
10-d mean	100[1]	99	100	104
5-d mean	50[2]	98	99	103
Monthly supervised				
All milkings[3]	10	95	96	100
2 of 3 milkings	10	92	93	97
1 of 2 milkings	10	89	91	95
1 of 3 milkings	10	83	87	90
Monthly owner-sampler[3]				
All milkings[4]	10	66	72	75
2 of 3 milkings	10	64	70	73
1 of 2 milkings	10	63	69	72
1 of 3 milkings	10	60	66	69
Bimonthly supervised				
All milkings[4]	5	90	91	95
2 of 3 milkings	5	84	87	90
1 of 2 milkings	5	79	83	86
1 of 3 milkings	5	71	75	78
Bimonthly owner-sampler				
All milkings[4]	5	63	69	72
2 of 3 milkings	5	61	66	69
1 of 2 milkings	5	58	64	67
1 of 3 milkings	5	53	60	62

[1]A 10-day average reported in each of 10 mo.
[2]A 5-day average reported in each of 10 mo.
[3]Owner-sampler records are assigned a maximum possible DCR of 75.
[4]Milk weights were obtained from all milkings on each test day.

weighed or sampled. Measurements reported by the cow's owner instead of a supervisor also are assumed to be more prone to bias than supervised tests because supervisors are trained to provide accurate data. Finally, variation of lactation yield is affected by the number and pattern of tests within the lactation.

Lactation yields for cows calving on or after January 1, 1997 were computed using the revised programs and multiple-trait BP. The data were extracted from the database and genetic evaluations were re-calculated for Holstein bulls that received genetic evaluations in April 2008. Average reliabilities of milk and protein yield increased by about 0.5% and 1.25%, respectively, suggesting that estimates of genetic merit can be improved by using information from correlated traits.

Conclusions

Best prediction is a flexible tool for accurately modelling milk, fat, and protein yields and SCS in lactations of any length. Data collection ratings allow US dairy producers more choices of sampling programs than those in many other countries, and almost all data from the field can be used for genetic evaluation. The program provides lactation-to-date, 305-d, 365-d, and projected yields, as well as BP of daily yields. The software is in the public domain and may be downloaded from the AIPL website at: <*www.aipl.arsusda.gov/software/bestpred*>

List of references

Cole, J.B. and VanRaden, P.M. 2007. A Manual for Use of BESTPRED: A Program for Estimation of Lactation Yield and Persistency Using Best Prediction. Available: <*www.aipl.arsusda.gov/software/bestpred*> Accessed May 14, 2008.

Cole, J.B. and VanRaden, P.M. 2006. Genetic evaluation and best prediction of lactation persistency. J. Dairy Sci. 89: 2722–2728.

ICAR. 2006. International Agreement of Recording Practices. International Committee for Animal Recording, Rome, Italy.

Reents, R., Jamrozik, J., Schaeffer, L.R. and Dekkers, J.C.M. 1995. Estimation of genetic parameters for test day records of somatic cell score. Dairy Sci. 78: 2847-2857.

Schutz, M.M., VanRaden, P.M., Wiggans, G.R. and Norman, H.D. 1995. Standardization of lactation means of somatic cell scores for calculation of genetic evaluations. J. Dairy Sci. 78: 1843–1854.

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.