



Short communication: Projecting milk yield using best prediction and the MilkBot lactation model

J. B. Cole,^{*1} J. L. Ehrlich,[†] and D. J. Null^{*}

^{*}Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

[†]Dairy Veterinarians Group, Argyle, NY 12809

ABSTRACT

The accuracy and precision of 3 lactation models was estimated by summarizing means and variability in projection error for next-test milk and actual 305-d milk yield (M305) for 50-d intervals in a large Dairy Herd Improvement Association data set. Lactations were grouped by breed (Holstein, Jersey, and cross-bred) and parity (first vs. later). A smaller, single-herd data set with both Dairy Herd Improvement Association data and daily milk weights was used to compare M305 calculated from test-day data with M305 computed by summing daily milk weights. The lactation models tested were best prediction (BP), the nonlinear MilkBot (MB) model, and a null model (NM) based on a stepwise function. The accuracy of the models was ranked (best to worst) MB, BP, and NM for later-parity cows and MB, NM, and BP for first-parity cows, with MB achieving accuracy in projecting daily milk of 0.5 kg or better in most groups. The models generally showed better accuracy after 50 d in milk. Best prediction and NM had low accuracy for crossbred cows and first-parity Holstein and Jersey cows. The MB model appears to be more precise than BP, and NM had low precision, especially for M305. Regression of model-generated M305 on summed M305 showed BP and MB to be equally efficient in ranking lactations, but MB was better at quantifying differences.

Key words: best prediction, lactation model, MilkBot

Short Communication

Many lactation models have been proposed for interpolating between test days to construct lactation curves and for projecting complete lactation production from partial data. Model accuracy and precision vary because of the influence of management, genetics, and environment on lactating cows. In the United States, lactation yields are calculated using best prediction (BP; Cole

and VanRaden, 2006), a method in which test-day data are compared with breed- and parity-specific herd lactation curves. However, Cole and VanRaden (2006) reported that BP systematically underestimates daily yield in early and late lactation, and overestimates daily yields in the middle of lactation. This may be due to the fact that BP uses a 3-parameter lactation model (Wood, 1967), and that it is optimized for the calculation of lactation yields from test-day data, rather than the prediction of daily yields. The MilkBot (MB) model developed by Ehrlich (2011) is a nonlinear lactation model designed to predict daily yields using a lactation curve model that is flexible enough to accommodate disease and management effects, and may provide more accurate estimates of daily milk yields. The objectives of this study were to calculate daily and lactation milk yields from a large database of test-day data using BP, MB, and a null model, and to compare the models against one another to determine which method provides the most accurate predictions of daily milk yield.

Data were obtained from 2 sources: a large data set of test-day records from many herds, and a small data set of actual daily milk yields from a single herd. A large data set of test-day data (DRMS) was collected from about 6 million lactations between 2005 and 2008 by Dairy Records Management Systems (Raleigh, NC), and included primarily herds from the eastern half of the United States. Six subsets of approximately 100,000 lactations each were selected from the DRMS data set, and included 3 breed categories (Holstein, crossbred, and Jersey) and 2 parity groups (first vs. later). Lactations with fewer than 9 test days were excluded. The second data source included 809 lactations recorded between 2003 and 2008 in a single Holstein herd in Wisconsin with an average 305-d milk yield of 10,048 kg. Data from this herd (DLY) were processed by Ag-Source Cooperative Services (Madison, WI), and both test-day data and daily milk weights were available. Summary statistics for each data set are shown in Table 1.

Daily and 305-d lactation yields using BP were calculated by the US Department of Agriculture Animal

Received September 6, 2011.

Accepted March 18, 2012.

¹Corresponding author: john.cole@ars.usda.gov

Table 1. Number of lactations (n), average actual 305-d milk yield (M305, kg), and average test-day milk yields (kg) for 50-d intervals by breed and parity group

| Breed ¹ | Parity ² | n ³ | M305 (kg) | DIM ⁴ | | | | |
|--------------------|---------------------|----------------|-----------|------------------|------------|------------|------------|------------|
| | | | | 51 to 99 | 100 to 149 | 150 to 199 | 200 to 249 | 250 to 305 |
| Crossbred | 1 | 14,622 | 8,251 | 27.7 | 29.4 | 28.6 | 27.6 | 26.3 |
| | 2+ | 17,112 | 9,307 | 36.0 | 35.8 | 32.8 | 29.8 | 26.7 |
| Holstein | 1 | 100,000 | 9,540 | 31.0 | 33.6 | 33.2 | 32.1 | 30.9 |
| | 2+ | 100,000 | 11,063 | 41.8 | 42.4 | 39.1 | 35.7 | 32.2 |
| Jersey | 1 | 20,883 | 6,621 | 22.8 | 23.7 | 22.8 | 21.8 | 20.8 |
| | 2+ | 40,399 | 7,650 | 29.4 | 29.1 | 26.5 | 24.2 | 22.0 |
| Daily | 1 | 485 | 10,198 | 31.4 | 35.5 | 35.7 | 34.6 | 33.2 |
| | 2+ | 400 | 11,603 | 43.5 | 44.7 | 42.0 | 37.6 | 33.1 |

¹The Daily breed group includes Holstein cows from a single herd in the state of Wisconsin.

²Parity groups include first- (1) and later-parity cows (2+).

³All records for crossbred and Jersey cows were used. For Holsteins, records were selected at random until 100,000 lactations were included in the data set.

⁴Average daily milk yields were calculated using test days in successive 50-d intervals (51 to 99, 100 to 149, 150 to 199, 200 to 249, and 250 to 305 DIM).

Improvement Programs Laboratory (Beltsville, MD) using the BESTPRED software package (Cole and VanRaden, 2007). The same lactation data were also fitted using the MilkBot engine (DairySight LLC, Argyle, NY), and 305-d yields calculated from those data. The parameters used by BP and MB models are default values computed using data sets independent of those used in this study (Cole and VanRaden, 2006; Ehrlich, 2011). A null model (NM) based on a stepwise function that assumed future milk yields were equal to the previous test-day observation also was used. Test-day data were grouped into 50-d periods (1 to 50 DIM, 1 to 100 DIM, . . . , 1 to 305 DIM), and projected 305-d yields were calculated to determine the effect of incomplete data on predictions. The resulting lactation yields were correlated with 305-d yields from complete lactations (M305), and projection errors were calculated as differences among the projected 305-d yields and the M305 for each method. For each 50-d interval, the yield on the next test day was predicted and compared with the actual observation. Milk weights were summed for the DLY set to obtain 305-d yields (S305) for each lactation, and missing values were filled in using interpolated values for each model. The number of missing points between 0 and 305 DIM averaged 9 for first-parity animals (ranging from 0 to 53) and 12 for later parities (ranging from 0 to 10). The mean error of prediction (MEP) and root mean square mean error of prediction (RMSEP) were calculated for each breed-parity-DIM group using each model for next-test milk and M305 (Tedeschi, 2006).

The accuracy of each model in predicting daily milk yield was measured using MEP, which measures nonrandom error, with zero MEP being equivalent to perfect accuracy. All of the models tested in this study

were constructed by minimizing MEP during their design phase, but mathematical characteristics of a model may prevent it from fitting data perfectly at all DIM. Variables that are not included within these models, such as use of somatotropin or changes in milking frequency, are expected to influence lactation curve shape and will affect the accuracy of projection.

Precision is the variability in prediction error, and was measured as RMSEP. No model can eliminate variability caused by factors outside the model, and the lack of precision attributable to the model is the difference between the total variability and variability from effects outside of the model domain. For a lactation model, outside effects include normal day-to-day variation in milk production, as well as changes in animal health or environment during a lactation. Variation among consecutive milkings in the DLY set was calculated using 512,622 consecutive lactation days, and had a standard deviation of 4.1 kg for later-parity cows and 3.7 kg for first-parity cows. Variability attributable to changes in animal health or environment is difficult to estimate, but RMSEP of about 5.0 kg for first-parity cows and 6 kg for later-parity cows may approach the maximum precision theoretically attainable.

Mean errors of prediction and RMSEP for M305 in Holsteins by parity are presented in Table 2. MilkBot had the best performance for first parity, and was similar to BP for later parities. Models were least accurate early in lactation when few data were available, and increased in accuracy as additional test-day observations were added. Notably, BP appears to substantially underpredict M305 when only early lactation data are available. Best prediction regresses individual performance toward a standard lactation curve (Cole and VanRaden, 2006), and when only 1 test day is available,

Table 2. Mean error of prediction (MEP) and root mean square error of prediction (RMSEP) of actual 305-d Holstein milk yield

| Measure ¹ | Method ² | Parity ³ | DIM ⁴ | | | | | |
|------------------------|---------------------|---------------------|------------------|------------|------------|------------|------------|------|
| | | | 51 to 99 | 100 to 149 | 150 to 199 | 200 to 249 | 250 to 300 | >300 |
| MEP (kg ²) | BP | 1 | -1,531 | -390 | -190 | -127 | -77 | -23 |
| | | 2+ | 80 | -2 | -42 | -71 | -39 | 1 |
| | MB | 1 | -200 | -7 | 0 | 21 | 37 | 3 |
| | | 2+ | -315 | -148 | 8 | 73 | 76 | 6 |
| | NM | 1 | -144 | 339 | 366 | 250 | 127 | 0 |
| | | 2+ | 1,559 | 1,707 | 1,148 | 674 | 296 | 0 |
| RMSEP (kg) | BP | 1 | 1,452 | 1,071 | 747 | 482 | 300 | 155 |
| | | 2+ | 1,806 | 1,223 | 861 | 582 | 366 | 183 |
| | MB | 1 | 1,586 | 1,229 | 836 | 518 | 276 | 57 |
| | | 2+ | 1,736 | 1,256 | 922 | 596 | 322 | 63 |
| | NM | 1 | 1,642 | 1,228 | 840 | 583 | 331 | 0 |
| | | 2+ | 2,073 | 1,616 | 1,151 | 802 | 485 | 0 |

¹Accuracy was measured as the MEP and precision was measured as the RMSEP.

²Methods for calculating lactation yields from test-day data were best prediction (BP), the MilkBot (MB) model, and the null model (NM).

³Parity groups include first- (1) and later-parity cows (2+).

⁴Accuracy and precision were calculated using test days in successive 50-d intervals (51 to 99, 100 to 149, 150 to 199, 200 to 249, and 250 to 305 DIM).

as is the case during the first 50 DIM, the resulting prediction will be very similar to the population average. The MilkBot model also uses a 4-parameter model, which is able to model a greater variety of curves than the BP 3-parameter Wood's curve. The performance of the NM was poor, especially with respect to precision, because the underlying model does not accurately reflect the shape of typical lactation curves. Results were generally similar for Jerseys and crossbred cows (data not shown), although BP overestimated rather than underestimated milk yield in early lactation, possibly because BP uses Holstein lactation curves for crossbred animals by default (Cole and VanRaden, 2007).

Due to the large sample size for prediction in the Holstein data set, the models were compared with one another using the MSE and RMSEP as discussed in Analla (1998). More sophisticated methods for model comparison have been presented in the literature (e.g., Diebold and Mariano, 1995); all differences among models would likely have been statistically significant despite the small magnitude of most differences. The SEP and RMSEP should, however, be interpreted with caution because they are sensitive to individual data points with large deviations from the average.

Predicted test-day milk yields were compared with observed values to assess model accuracy. Mean errors of prediction and RMSEP for Holsteins by parity are presented in Table 3. Accuracy and precision varied between models and parities, and improved with increasing DIM. This is expected because early in lactation few data are available and normal changes in milk production are large. The MB model consistently showed both better accuracy and better precision than BP or the NM. For first-parity lactations, the NM has

better accuracy and precision than BP in estimating next-test milk. As was the case for M305, results were generally similar for Jerseys and crossbred cows (data not shown).

Relationships among M305, which is predicted from monthly test-day observations, and S305, which is based on summed daily milk weights, were examined using the DLY data set. The S305 were regressed on M305 for each lactation model-parity group combination and model fit assessed using the coefficient of determination statistic (Cook and Weisberg, 1998). Best prediction and the MB model both had high coefficients of determination, with values ranging from 0.91 to 0.93, but MB results had intercepts closer to 0 and slopes nearer 1 than did BP. The NM performed very poorly, with coefficient of determination of 0.15 and 0.22 for first and later parities, respectively, because NM calculates M305 with very low precision.

Best prediction currently is used to calculate 305-d milk yields from test-day data, which are inputs for the calculation of genetic evaluations. The ability of BP to accurately predict daily milk yields is less important than its ability to predict M305 without error. MilkBot, on the other hand, was designed for prospective applications such as detecting and quantifying effects of disease or management interventions. Recent work by Hostens et al. (2012) suggests that MB can model changes in the shape of individual lactation curves attributable to metabolic diseases. Appuhamy et al. (2007, 2009) have recently shown, for example, that diseases affect the shape of the lactation curve, and prediction models that do not account for such events may have decreased accuracy or precision, or both. For such applications, predictive accuracy and precision both are important,

Table 3. Mean error of prediction (MEP) and root mean square error of prediction (RMSEP) of next test-day Holstein milk yield

| Measure ¹ | Method ² | Parity ³ | DIM ⁴ | | | | |
|------------------------|---------------------|---------------------|------------------|------------|------------|------------|------------|
| | | | 51 to 99 | 100 to 149 | 150 to 199 | 200 to 249 | 250 to 305 |
| MEP (kg ²) | BP | 1 | -7.05 | -3.59 | -3.24 | -3.66 | -3.75 |
| | | 2+ | 0.71 | 0.61 | -0.09 | -0.85 | -0.60 |
| | MB | 1 | -0.04 | -0.23 | -0.21 | -0.21 | 0.33 |
| | | 2+ | -0.32 | -0.54 | -0.55 | -0.16 | 0.91 |
| | NM | 1 | -2.27 | 0.12 | 0.65 | 0.74 | 1.20 |
| | | 2+ | -1.14 | 1.88 | 2.23 | 2.32 | 2.78 |
| RMSEP (kg) | BP | 1 | 7.55 | 6.75 | 6.42 | 6.17 | 6.08 |
| | | 2+ | 10.29 | 8.20 | 7.33 | 6.90 | 6.96 |
| | MB | 1 | 5.76 | 5.26 | 5.05 | 5.05 | 5.09 |
| | | 2+ | 7.53 | 6.59 | 6.17 | 6.00 | 6.15 |
| | NM | 1 | 6.18 | 5.61 | 5.45 | 5.45 | -5.46 |
| | | 2+ | 8.27 | 7.34 | 6.98 | 6.58 | 6.37 |

¹Accuracy was measured as the MEP and precision was measured as the RMSEP.

²Methods for calculating lactation yields from test-day data were best prediction (BP), the MilkBot (MB) model, and the null model (NM).

³Parity groups include first- (1) and later-parity cows (2+).

⁴Accuracy and precision were calculated using test days in successive 50-d intervals (51 to 99, 100 to 149, 150 to 199, 200 to 249, and 250 to 305 DIM).

with poor accuracy likely to result in type 1 errors, and poor precision more likely to lead to type 2 errors. In the context of a herd management tool used to monitor individual lactation curves, a type 1 error represents the incorrect conclusion that veterinary treatment or management intervention is necessary when it is not, whereas a type 2 error represents a failure to take such an action when it is necessary. This is of increasing importance as dairy farmers continue to seek out new tools for better managing their herds, and those tools will require accurate and precise predictions of milk, fat, and protein production.

The nonlinear MB may have several applications in the dairy industry. Its incorporation into herd management software could provide farm managers with a tool that alerts them quickly to changes in the shape of a cow's lactation curve in response to health or management difficulties. It also may provide more accurate estimates of daily yields than BP, which is of interest to dairy records processing centers and other providers of herd management tools. However, although MB provides predictions of daily milk yield with greater accuracy and precision than BP, the latter currently provides calculations for more breed and trait combinations than MB.

ACKNOWLEDGMENTS

The cooperation of Dairy Records Management Systems (Raleigh, NC, and Ames, IA) and AgSource Cooperative Services (Madison, WI) in supplying lactation yield data is gratefully acknowledged. This research was conducted in support of Cooperative

Research and Development Agreement No. 58-3K95-9-1349-M between Dairy Veterinarians Group (Argyle, NY) and the US Department of Agriculture (Beltsville, MD). The authors appreciate the comments of 2 anonymous reviewers that helped improve the quality of this manuscript.

REFERENCES

- Analla, M. 1998. Model validation through linear regression fit to actual versus predicted values. *Agric. Syst.* 57:115-119.
- Appuhamy, J. A. D. R. N., B. G. Cassell, and J. B. Cole. 2009. Phenotypic and genetic relationships of common health disorders with milk and fat yield persistencies from producer-recorded health data and test day yields. *J. Dairy Sci.* 92:1785-1795.
- Appuhamy, J. A. D. R. N., B. G. Cassell, C. D. Dechow, and J. B. Cole. 2007. Phenotypic relationships of common health disorders in dairy cows to lactation persistency estimated from daily milk weights. *J. Dairy Sci.* 90:4424-4434.
- Cole, J. B., and P. M. VanRaden. 2006. Genetic evaluation and best prediction of lactation persistency. *J. Dairy Sci.* 89:2722-2728.
- Cole, J. B., and P. M. VanRaden. 2007. A Manual for Use of BEST-PRED: A Program for Estimation of Lactation Yield and Persistency Using Best Prediction. Accessed Aug. 16, 2011. <http://www.aipl.arsusda.gov/software/bestpred/>.
- Cook, R. D., and S. Weisberg. 1998. *Applied Regression Including Computing and Graphics*. John Wiley & Sons, New York, NY.
- Diebold, F. X., and R. S. Mariano. 1995. Comparing predictive accuracy. *J. Bus. Econ. Stat.* 13:253-263.
- Ehrlich, J. L. 2011. Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. *Bov. Pract.* 45:88-95.
- Hostens, M., J. Ehrlich, B. Van Ranst, and G. Opsmer. 2012. On-farm evaluation of the effect of metabolic diseases on the shape of the lactation curve in dairy cows through the MilkBot lactation model. *J. Dairy Sci.* <http://dx.doi.org/10.3168/jds.2011-4791>.
- Tedeschi, L. O. 2006. Assessment of the adequacy of mathematical models. *Agric. Syst.* 89:225-247.
- Wood, P. D. P. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216:164-165.