Asian-Aust. J. Anim. Sci.
Vol. 19, No. 7 : 915 - 921
July 2006

www.ajas.info

Prediction of Future Milk Yield with Random Regression Model Using Test-day Records in Holstein Cows

Byoungho Park and Deukhwan Lee*

National Livestock Research Institute, RDA, Seonghwan, Chungnam 330-801, Korea

ABSTRACT : Various random regression models with different order of Legendre polynomials for permanent environmental and genetic effects were constructed to predict future milk yield of Holstein cows in Korea. A total of 257,908 test-day (TD) milk yield records from a total of 28,135 cows belonging to 1,090 herds were considered for estimating (co)variance of the random covariate coefficients using an expectation-maximization REML algorithm in an animal mixed model. The variances did not change much between the models, having different order of Legendre polynomial, but a decreasing trend was observed with increase in the order of Legendre polynomial in the model. The R-squared value of the model increased and the residual variance reduced with the increase in order of Legendre polynomial in the model. Therefore, a model with 5th order of Legendre polynomial was considered for predicting future milk yield. For predicting the future milk yield of cows, 132,771 TD records from 28,135 cows were randomly selected from the above data by way of preceding partial TD record, and then future milk yields were estimated using incomplete records from each cow randomly retained. Results suggested that we could predict the next four months milk yield with an error deviation of 4 kg. The correlation of more than 70% between predicted and observed values was estimated for the next four months milk yield. Even using only 3 TD records of some cows, the average milk yield of Korean Holstein cows would be predicted with high accuracy if compared with observed milk yield. Persistency of each cow was estimated which might be useful for selecting the cows with higher persistency. The results of the present study suggested the use of a 5th order Legendre polynomial to predict the future milk yield of each cow. (**Key Words :** Test Day, Holstein Cows, Random Regression Model, Future Milk Yield)

INTRODUCTION

For avoiding a shortfall or an overproduction of milk, the dairy producers and milk processors of Korea need an accurate prediction of future milk yield from each cow of their herds. The early prediction will also help the Korean dairy industry to take appropriate measures well in time to ensure normal milk supply to consumers. The accurate measurement of daily milk production of each cow and of thus each herd may not be feasible due to high expenses incurred towards data collection.

Test day models (TDM) based on test-day (TD) records of milk production at regular interval of time is now gaining importance not only due to its cost effectiveness but also due to its accuracy to predict future milk yield with fewer test-day records. The TDM can include the individual test-day effects, which affect the test-day yields substantially and it can account for individual differences in the shape of the lactation curves of cows (Jamrozik et al., 1996). Thus, the TDM can be used to analyze individual test-day records of cows in place of full 305 days lactation model (Mayeres et al., 2004).

Several methods have been suggested to predict future milk yield from a limited number of test-day records. Jones (1997) used Empirical Bayesian method (EBM) in which the milk yields from lactation in progress were combined with prior information gathered from herd mates. Macciotta et al. (2002) used Autoregressive Integrated Moving Average (ARMA) model to predict TD milk production data, which required large run of time series data and also required technical expertise on the part of forecaster. Schaeffer and Dekkers (1994) developed random regression model, where, the shape of the lactation curve was modeled by a random regression function. The shape of the lactation curve for an individual cow was divided as two sets of regressions on days in milk (DIM). The fixed regressions for all cows of the same subclass described the general

* Corresponding Author: D. H. Lee. Dept. of Animal Life and Resources, Hankyong National University, Seokjeong-Dong 67, Ansung City, Kyeonggi-Do, 456-749, Korea. Tel: +82-31-670-5091, Fax: +82-31-676-5091, E-mail: dhlee@hknu.ac.kr
Received October 27, 2005; Accepted February 25, 2006

Table 1. General information for test day milk records at first lactation on the study in Holstein cows

	No.	Mean	SD	Minimum	Maximum
Records/herd	1,090	236.6	233.9	11	3,134
Cows/herd	1,090	25.8	23.0	8	336
Records/cow	28,135	9.2	268.0	1	14
DIM (d)		180.4	101.6	1	399
TD milk yield (kg)	257,908	26.6	6.5	2	68.5

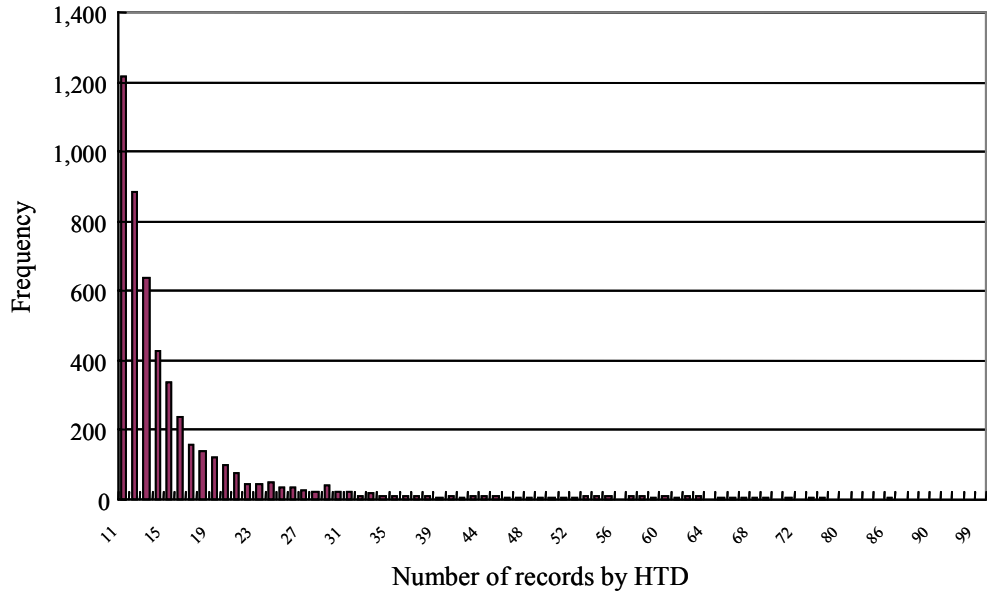

Figure 1. Frequency by number of records on milk yields by herd-test-date (HTD) in Holstein cows.

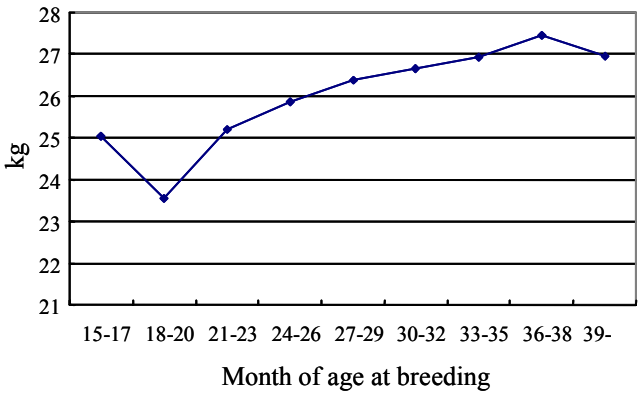

Figure 2. A plot of the average daily milk yields by month of age at breeding on first lactation in Holstein cows.

shape for that cow and the random regressions for a cow described the genetic deviations from the fixed regressions, which allowed each cow to have a different shape of the lactation curve on a genetic level (Jamrozik et al., 1996). This was the random regression model for TD yield. Shadparvar and Yazdanshenas (2005) reported genetic parameters for TD milk yields estimated by REML procedure with assuming different TD milk yields to different traits in a multiple traits test day model. Later on, Pool and Meuwissen (1999) developed phenotypic TD models incorporating Legendre polynomials being able to interpolate and extrapolate missing records. The objectives of the present study were to predict the future milk yields of each cow, to predict the persistency of production of each cow, and to access the best random regression function for predicting future milk yield using TDM.

MATERIALS AND METHODS

Data

The raw data comprised of a total of 360,945 TD milk yield records from first lactation of Holstein cows in Korea from 1997 to 2003. The recording was done once in a month and each TD yield comprised of morning and evening milk yield of each cow. The data were restricted to DIM on TD between 1 to 400 days, on cows having TD records between 5 and 14 (both included), on herd having TD records more than 10 (i.e. from each herd at least 11 cows had been recorded during each TD) and on cows having TD milk yield from 3 to 70 kg. Frequency by number of records on milk yields by herd-test-date (HTD) was shown in Figure 1. After imposing all the restrictions, the final data consist of 257,908 TD records from 28,135 cows belonging to 1,090 herds. This data set consisting of whole TD records was used to estimate the (co)variance components of random covariate coefficients using model

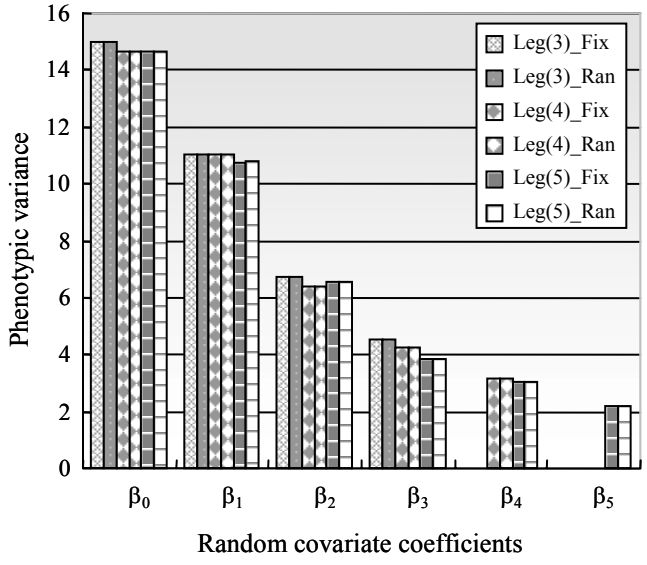

Figure 3. Comparison of variances of the random covariate coefficients for milk yield (kg^2) of a test day model using Legendre polynomials of 3rd, 4th and 5th order of fit in Holstein cows. Leg(m) with mth order of fit; Fix = Model assumed herd effect of fixed; Ran = Model assumed herd effect of random.

with Legendre polynomial of different order of fit and to estimate the milking curve of each cow. The detailed general statistics of the final data structure is given in Table 1.

For predicting the future milk yield of cows, 132,771 TD records from 28,135 cows were randomly selected from the above data by way of retaining preceding partial TD records on each cow and then future milk yields by each cow were estimated using incomplete records. The random selection of records was carried out by using function of random number generation on SAS package (SAS, 2004).

Statistical analysis

Variance component estimation : The data were classified according to the age at breeding by an interval of 3 months of age (Figure 3). For model construction, the following effects were included in the model for estimating variance component of regression coefficients after checking significances of their effects on milk yields (Table 2).

Model :

$$Y = PRY + PM + BA + TM + DIM + HERD + \sum \beta a + e$$

Where

Y = TD milk yield of a cow

PYR = parturition year (level: 7)

PM = month at parturition (level: 12)

BA = class of breeding age with an interval of every 3 month from 15-39 and above months of age (level: 9)

Table 2. Analysis of variance for milk yields on first lactation test-day records in Holstein cows ($R^2 = 0.34$)

Source	DF	Mean square	F-value
HERD	1,089	1,924.482	70.9**
PYR	6	34,611.242	1,275.09**
PM	11	2,551.568	94.00**
BA	8	7,945.884	292.73**
TM	11	4,719.500	173.87**
DIM	398	3,509.538	129.29**

** $p<0.01$.

TM = calendar month of test date (level: 12)

DIM = day in milk (level: 399)

HERD = herd (level: 1089)

Σβa: Legendre polynomial (β_0, β_1, β_2, β_3, β_4, β_5)

a = random effect of animal (which is genetic and permanent environmental effect i.e. assuming of $a \sim N(0, I\sigma_a^2)$; (level = 28,135)

e = random residual effect assuming of $e \sim N(0, I\sigma_e^2)$

Three models with 3rd, 4th and 5th order of Legendre polynomials were considered. Within each model, herd was considered as fixed effect and as random effect thus making a total of six models for analysis of covariance components. All the models were compared and the best model was chosen to predict the future milk yield of cows with criteria of prediction error variances. Computer program on which EM-REML algorithm implemented was used to obtain the (co)variance component estimates.

Model search : The shape of the lactation curve of each cow was estimated by TD random regression model with assuming as below:

$$y_{ij} = x'_{ij}\beta_k + \phi'_{ij(m)}k_{i(m)} + e_{ij}$$

Where, y_{ij} = jth test day milk yield of ith animal; x'_{ij} = incidence row vector for fixed effects of test day milk yield j in animal i; β = [*PYR:PM:BA:TM:DIM:HERD*]'; $\emptyset_{ij(m)}$ is a row vector of random regression factors of the mth order Legendre polynomial; $k_{i(m)}$ is a (m by 1) vector of individual random regression coefficients of animal i with m as the order of fit; e_{ij} = random error term. In this model, all the effects were considered as fixed except random regression factor and residual effect.

Another model with assuming herd effects of random was used as below:

$$y_{ij} = x'_{ij}\beta_k + HERD_j + \phi'_{ij(m)}k_{i(m)} + e_{ij}$$

Where β = [*PYR:PM:BA:TM:DIM*]'; $HERD_j$ = jth herd random effect.

The DIM was standardized with the range from -1 to 1.

Table 3. Variance (diagonal) and covariance (below diagonal) estimates of the random covariate coefficients for milk yields using the model with assumed 5th order of Legendre polynomials and herd effect of fixed using EM-REML algorithm in Holstein cows

	β_0	β_1	β_2	β_3	β_4	β_5
β_0	14.620					
β_1	1.828	10.770				
β_2	-2.121	0.165	6.541			
β_3	1.063	-0.668	-1.247	3.819		
β_4	-1.110	-0.664	0.240	-0.842	3.076	
β_5	0.580	-0.741	-0.561	-0.023	-0.490	2.200

β_m = parameter for ^{m}th order random regression coefficients.

Table 4. Variance (diagonal) and covariance (below diagonal) estimates of the random covariate coefficients for milk yields using the model with assumed 5th order of Legendre polynomials and herd effect of random using EM-REML algorithm in Holstein cows

	β_0	β_1	β_2	β_3	β_4	β_5
β_0	14.620					
β_1	1.830	10.780				
β_2	-2.234	0.169	6.541			
β_3	1.122	-0.670	-1.246	3.817		
β_4	-1.138	-0.662	0.236	-0.839	3.073	
β_5	0.604	-0.741	-0.559	-0.024	-0.486	2.200

Herd variance = 6.743.

β_m = parameter for ^{m}th order random regression coefficients.

Prediction of future milk yield

The second data set, with retaining preceding partial TD records on each cow, was used for future milk prediction. The prediction of the deleted records was done and the predicted milk yield was compared with the measured milk yield of each cow. For this purpose, a graph on DIM verses residual standard deviation for predicted milk yield was drawn. Graphical representation of correlation between the real and predicted milk yield was carried out. The average of all cows for predicted and observed milk yield was also plotted to see the accuracy of prediction for whole lactation.

Persistency

Persistency usually refers to the rate of decline in daily yield after the peak of lactation. The persistency of each cow was estimated as proposed by Togashi and Lin (2004) with some modification as follow:

Persistency of animal

$$i = \sum_{j=65}^{279} (\hat{a}_{280(i)} - \hat{a}_{j(i)})$$

Where, j = ^{j}th days in milk; i = ^{i}th animal

$a_{j(i)}$ was the estimated milk yield at ^{j}th days in milk of animal and given by $a_{j(i)} = k_{j(i)}.Sol_{(i)}$

Table 5. Correlations between random covariate coefficients modeled by a test day model using 3rd(LEG3), 4th(LEG4) and 5th(LEG5) order of fit Legendre polynomial, respectively, for milk yields in Holstein cows

	β_0	β_1	β_2	β_3	β_4
LEG3					
β_1	0.1548				
β_2	-0.1411	0.0533			
β_3	0.1427	0.0002	-0.1173		
LEG4					
β_1	0.1345				
β_2	-0.2385	0.0249			
β_3	0.0901	-0.0259	-0.2043		
β_4	-0.2029	-0.0916	0.0531	-0.1471	
LEG5					
β_1	0.1458				
β_2	-0.2284	0.0201			
β_3	0.1502	-0.1045	-0.2494		
β_4	-0.1698	-0.1149	0.0527	-0.2448	
β_5	0.1065	-0.1522	-0.1473	-0.0081	-0.187

$k_{j(i)}$ = vector of 5th order Legendre polynomial of ^{j}th DIM of ^{i}th animal

Sol(i) = vector of covariate estimate of Legendre polynomial on ^{i}th animal

When an average lactation curve of Holstein cows for first lactation milk yield was plotted it was observed that the peak of lactation curve was achieved on day 65 (Figure 3). This peak day was included in the formula.

RESULTS AND DISSCUSSION

Phenotypic (co)variance estimates

The phenotypic variance and covariance estimates for the random regression coefficients for milk yield modeled by TDM having 5th order of Legendre polynomials were presented in Table 3 and 4. The herd effect was considered as fixed (Table 3) and as random (Table 4). The present estimate for the intercept (β_0) was much lower than the estimate reported by Pool and Meuwissen (1999). But for the other higher order random covariate coefficients the present estimates were higher than the estimates reported by Pool and Meuwissen (1999). The (co)variance estimates from model having Legendre polynomial of 3rd and 4th order of fit were estimated (not shown). The phenotypic variances of the six models considering 3rd, 4th and 5th order of Legendre polynomial had not shown much difference but a decreasing trend was observed in the value of the estimates with an increase in number of random covariate coefficients (i.e. with the order of fit of the Legendre polynomial) in the model (Figure 3). This result was in agreement with the findings of Pool and Meuwissen (1999). The difference in the model considering herd as fixed or random effect had not shown any significant difference in estimating variances for random covariate coefficients. We

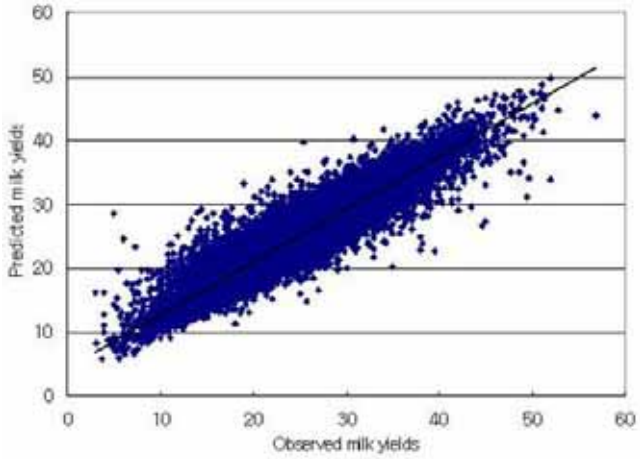

Figure 4. A plot between predicted and observed values of milk yield using 5th order of Legendre polynomials (y = 1.05674x-1.55486).

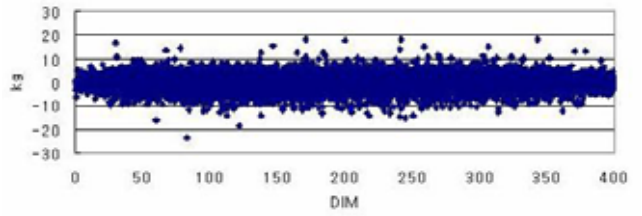

Figure 5. A plot of residuals for milk yield by DIM using 5th order Legendre polynomials in Holstein cows.

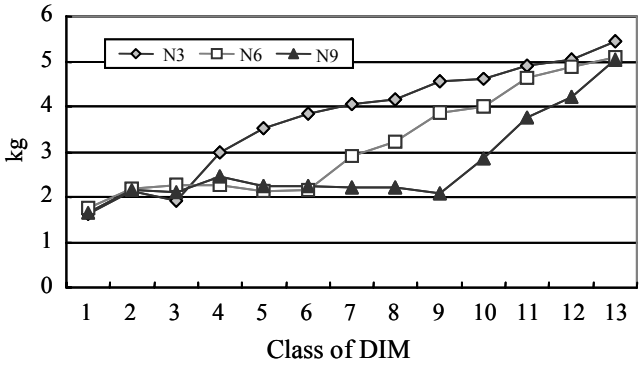

Figure 6. Residual standard deviation on monthly classes of DIM for predicted milk yields using 3, 6 and 9 TD records in Holstein cows.

used herd as fixed effect in the model for predicting future milk yield.

The correlations estimated between the random covariate coefficients for the TD model with 3rd, 4th and 5th order of Legendre polynomial function were presented in Table 5. The correlations were small in magnitude but not negligible. The present findings were similar to the previous findings of Pool and Meuwissen (1999) who reported low correlations between the random regression coefficients. There was no specific trend observed for the correlations with the increase in the order of random covariate coefficients in the model. However, Pool and Meuwissen (1999) found a tendency of stronger correlations among the coefficients with higher order of random regression coefficients in the model. The mean of all the random covariate coefficients from all the 28,135 cows were observed as zero (data not shown), which was expected.

Comparison of different models

For predicting future milk yield, a total of six models with 3rd, 4th and 5th order of Legendre polynomials were considered and with each order of Legendre polynomial two models were constructed on the basis of consideration of herd as random or fixed effect. As already mentioned above there was not much difference between the models considering herd as fixed or random on the basis of variances of the random covariate coefficients (Figure 3).

The predicted and observed values of milk yields obtained from the models with 3rd, 4th and 5th order of Legendre polynomials were plotted (Figure 4). The figures for models with 3rd and 4th order of polynomial were not shown here. The determinant of the model (R^2) increased with the higher order of Legendre polynomial in the models with assuming fixed effect of herds (Table 6). The differences between the models with different order of Legendre polynomials were also compared on the basis of residual variances. The residual variance decreased with the increase in the order of Legendre polynomial in the model (Table 6). Figure 5 showed a plot of residuals for milk yield by DIM using 5th order Legendre polynomial. On the basis of these comparisons, model with 5th order of Legendre polynomials considering herd as fixed effect was selected to predict the future milk yields. Pool and Meuwissen (1999) also suggested the suitability of model with 5th order of Legendre polynomial over model with lower or higher order of Legendre polynomial.

Prediction of future milk yield

The model equation for predicting the milk yield using

Table 6. Comparison between regression equations of actual milk yield (Y) on predicted milk yield (X) using 3 different functions for Legendre polynomial in Holstein cows

Legendre polynomials	Model equation	R^2	$\hat{\sigma}_E^2$	$\hat{\sigma}_E$
3rd order	$\hat{y}$ = 1.05487*x*-1.50367	0.896	5.064	2.250
4th order	$\hat{y}$ = 1.05591*x*-1.53217	0.905	4.420	2.102
5th order	$\hat{y}$ = 1.05674*x*-1.55486	0.912	4.116	2.029

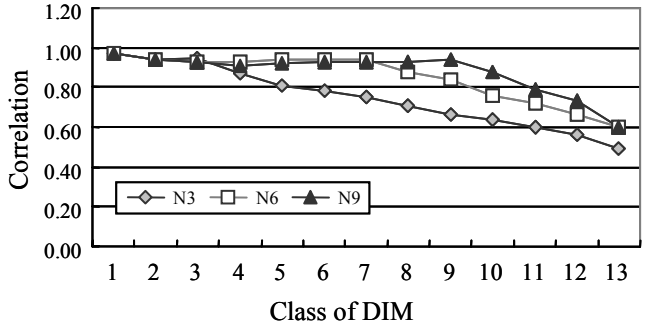

Figure 7. Plots of correlation estimates between real and predicted milk yields by monthly classes of days in milk (DIM) using 3, 6 and 9 TD records in Holstein cows.

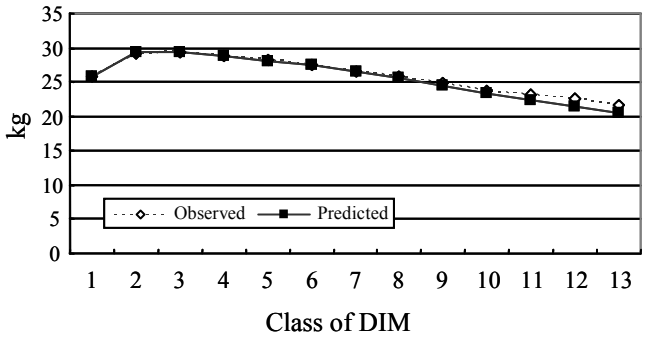

Figure 8. Plots of average predicted and observed milk yields by monthly classes of days in milk (DIM) using 3 observations for test-day milk yields on each cow in Holstein cows.

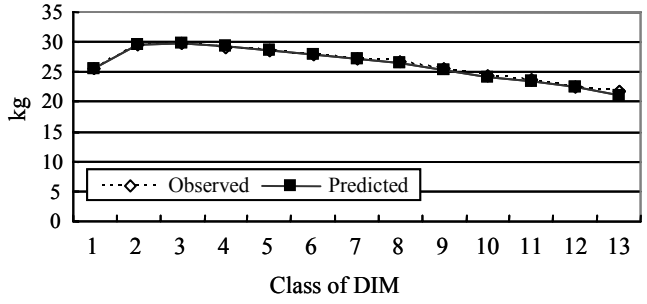

Figure 9. Plots of average predicted and observed milk yields by monthly classes of days in milk (DIM) using 9 observations for test-day milk yields on each cow in Holstein cows.

model with 5th order Legendre polynomial function was $\hat{y} = 1.05674x - 1.55486$.

For estimating the accuracy of the prediction, we compared predictions using various test-day records. Residual standard deviation on classes of DIM, which were monthly classified of DIM, for predicted milk yield considering 3, 6 and 9 TD records were presented in Figure 6. Results showed that the next four month's milk yield could be predicted with an error prediction of 4 kg. The residual standard deviation of prediction was increased up to 5 kg till the end of lactation.

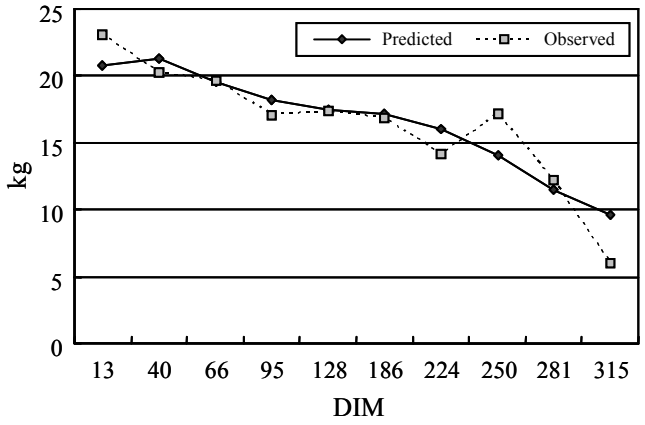

Figure 10. An example of plots for predicted and observed milk curve of a cow having low persistency (-276 kg).

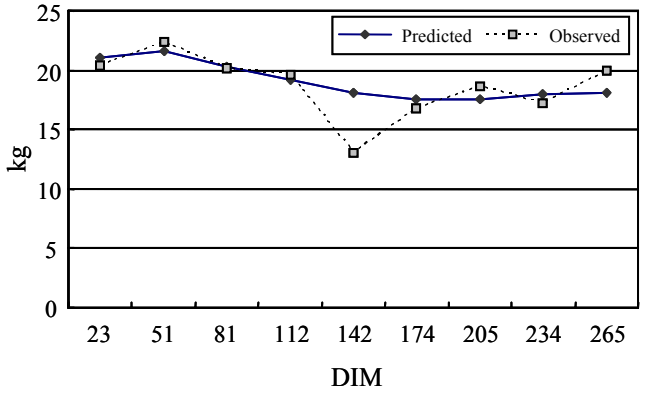

Figure 11. An example of plots for predicted and observed milk curve of a cow having high persistency (627 kg).

When DIM was plotted against the correlation between the real and predicted milk yield, almost 70% correlation between predicted and observed values was shown for following next four months (Figure 7). Even considering only 3 TD records, more than 60% of correlation was observed between the real and predicted milk yield up to 11 months of lactation. These results suggested that few TD observations could be used to predict future milk yield at an early lactation of a cow. However if the records were available on further TD, then its inclusion in the analysis would give better prediction with higher correlation.

Considering average lactation curve of all the cows using observed and predicted values (with only 3 TD records), the two curves were almost similar and a minor difference was observed from 9 month of lactation till the 13 months of lactation (Figure 8). These results suggested that the future average milk from all the cows could be accurately measured using first few months' TD records for almost whole lactation. Using 9 TD records, the two curves were found to be almost similar for whole lactation (Figure 9). Thus, the more the number of test-day records was used, the more the accuracy of prediction would be shown.

Persistency

The rate of decline in daily yield after the peak lactation is usually referred as persistency. Persistencies of all the cows were calculated and the mean was zero with standard deviation of 377 kg. The minimum persistency was -1,918 kg and the highest was 2,799 kg. For example, test-day records of milk yields on two cows with shown high and low persistency were selected and plotted by DIM (Figures 10 and 11). There was a quick decline in milk yield after achieving peak and it declines further to reach up to 13 kg at the 265 DIM on a cow having low persistency (-276 kg). Otherwise, the milk yield was maintained at 18 kg till the end of 265 DIM on a cow having high persistency (627 kg). The result showed inverse relationship between the rate of decline in milk yield and the persistency. The greater was the rate of decline, the lower would be the persistency (Togashi and Lin, 2004).

The present study dealt with the phenotypic prediction of the random covariate coefficients using 5th order of Legendre polynomial in the test day random regression model. These predictions could be used for estimating breeding value of the cows if considering pedigree information on a model (Pool and Meuwissen, 1999).

CONCLUSION

Fifth order Legendre polynomial function gave the best fit to predict future milk yield. With the restriction of 4 kg of standard deviation of error we could predict next four months milk yield of a cow. Almost 70% correlation between predicted and observed values was estimated for following next four months even though fewer TD records were used. Also, the future average milk yield of Holstein population for whole lactation can be predicted with few number of TD records with high accuracy.

ACKNOWLEDGEMENT

We thank two anonymous reviewers for helpful comments on the manuscripts. This work was supported by a grant from "Specific Joint Agricultural Research-Promoting Projects in RDA", Republic of Korea (Project No. 20050201033019). The authors are grateful to the Dairy Cattle Improvement Center for supplying the research data.

REFERENCES

Jamrozik, J., L. R. Schaeffer and J. C. M. Dekkers. 1996. Genetic evaluation of dairy cattle using test day yields and random regression model. J. Dairy Sci. 80:1217-1226.

Jones, T. 1997. Empirical Bayes prediction of 305-day milk production. J. Dairy Sci. 80:1060-1075.

Kirkpatrick, M., D. Lofsvold and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genet. 124:979-993.

Macciotta, N. P. P., D. Vicario, G. Pulina and A. Cappio-Borolino. 2002. Test day and lactation yield prediction in Italian Simmental cows by ARMA methods. J. Dairy Sci. 85:3107-3114.

Mayeres, P., J. Stoll, J. Bormann, R. Reents and N. Gengler. 2004. Prediction of daily milk, fat, and protein production by a random regression test-day model. J. Dairy Sci. 87:1925-1993.

Misztal, I. 2001. Blupf90 family of programs. http:nce.ads.uga.edu/~ignacy/ newprograms.html. Accessed May 12, 2005.

Pool, M. H. and T. H. E. Meuwissen. 1999. Prediction of daily milk yields from a limited number of test days using test day models. J. Dairy Sci. 82:1555-1564.

Schaeffer, L. R. and J. C. M. Dekkers. 1994. Random regressions in animal models for testday production in dairy cattle. Proc. of the 5th WCGALP Guelph, Canada, 18:443-446.

Shadparvar, A. A. and M. S. Yazdanshenas. 2005. Genetic parameters of milk yield and milk fat percentage test day records of Iranian Holstein cows. Asian-Aust. J. Anim. Sci. 18:1231-1236.

Togashi, K. and C. Y. Lin. 2003. Efficiency of different selection criteria for persistency and lactation milk yield. J. Dairy Sci. 87:1528-1535.