Prediction of 305 Days Milk Production from Early Records in Dairy Cattle Using an Empirical Bayes Method

J. A. C. Pereira*, M. Suzuki and K. Hagiya[1]

Obihiro University of Agriculture and Veterinary Medicine, Lab. of Genetics, Obihiro, Hokkaido 080-8555, Japan

ABSTRACT : A prediction of 305 d milk production from early records using an empirical Bayes method (EBM) was performed. The EBM was compared with the best predicted estimation (BPE), test interval method (TIM), and the linearized Wood's model (LWM). Daily milk yields were obtained from 606 first lactation Japanese Holstein cows in three herds. From each file of 305 daily records, 10 random test day records with an interval of approximately one month were taken. The accuracies of these methods were compared using the absolute difference (AD) and the standard deviation (SD) of the differences between the actual and the estimated 305 d milk production. The results showed that in the early stage of the lactation, EBM was superior in obtaining the prediction with high accuracy. When all the herds were analyzed jointly, the AD during the first 5 test day records were on average 373, 590, 917 and 1,042 kg for EBM, BPE, TIM, and LWM, respectively. Corresponding SD for EBM, BPE, TIM, and LWM were on average 488, 733, 747 and 1,605 kg. When the herds were analyzed separately, the EBM predictions retained high accuracy. When more information on the actual lactation was added to the prediction, TIM and LWM gradually achieved better accuracies. Finally, in the last period of the lactation, the accuracy of both of the methods exceeded EBM and BPM. The AD for the last 2 samples analyzing all the herds jointly were on average 141, 142, 164, and 214 kg for LWM, TIM, EBM, and BPE, respectively. In the current practices of collecting monthly records, early prediction of future milk production may be more accurate using EBM. Alternatively, if enough information of the actual lactation is accumulated, TIM may obtain better accuracy in the latter stage of lactation. *(Asian-Aust. J. Anim. Sci. 2001. Vol 14, No. 11 : 1511-1515)*

Key Words : Dairy Cattle, Daily Milk Recording, Best Predicted Estimation, Test Interval Method, Wood's Model

INTRODUCTION

Accurate measurement of milk production per cow in the lactation period is of a major interest in dairy genetic improvement programs. Using these measures, a cumulative 305 d milk production is calculated and used not only in genetic evaluations, but also in herd management as a decision tool (e.g., culling or feeding management). However, a problem arises every time a genetic evaluation is performed. At that time, a considerable number of lactations are still in progress and cannot be included in the genetic analysis. In such cases, a prediction of future production is required. The current methodology relies on extension and adjustment factors to predict future production. In order to calculate accurate adjustment factors, a great quantity of data must be stored in advance. This point becomes a restriction in countries where availability of data is limited. In these conditions, the derivation of extension factors is difficult to perform. Some studies (Conglenton and Everett, 1980; Jones, 1997) reported that the empirical Bayes method (EBM) seems to give predictions that appear to be of good quality, even when relatively little information from herd-mates is available. Therefore, a prediction of future milk production can be performed without the construction of factors and extension tables or the accumulation of large quantities of data as is required by the current methodology. The objectives of this study were to predict future cumulative 305 d milk production from early records of lactations in progress using the EBM, and to compare the predictions of EBM with the test interval method (TIM), the linearized Wood's model (LWM), and the best predicted estimation (BPE).

* Address reprint request to J. A. C. Pereira. Tel: +81-155-49-5115, Fax: +81-155-49-5414, E-mail: antonio@obihiro.ac.jp

[1] Iwate University, The United Graduate School of Agricultural Sciences, Morioka, Iwate 020-8550, Japan.

Received February 27, 2001; Accepted June 18, 2001

MATERIALS AND METHODS

Data

Daily milk yields from 606 first lactation Japanese Holstein cows recorded in three herds were used. The records were collected during different seasons on a twice-daily basis in two different areas. In addition, different numbers of cows by herds were available. The number of cows was 85 in herd one, 20 in herd two and 501 in herd three. For analysis, only lactations that had a complete record of daily milk yields during the period of 305 days of milk production were included. In order to simulate the milk sampling of the current recording system only monthly records of the lactation were needed. Therefore, from each file containing 305 daily records, 10 records of milk production (test day records) with an interval of approximately one month were taken randomly. This operation was repeated 10 times in all of the lactation files,

giving a final number of 6,060 first lactation records.

Empirical Bayes method

Empirical Bayes uses techniques and results of the Bayes approach. Therefore, a prior distribution must be estimated using the observed marginal distribution of the recorded data (Maritz and Lwin, 1989). In dairy cattle, a prior distribution can be calculated because previous data are readily available from records of historical lactations collected by dairy cattle associations or by the farmers in previous years. Historical lactation is defined as a complete lactation period of a cow recorded in previous years containing at least 10 test day records with an interval of one month approximately. The EBM can be adapted to different mathematical functions in order to predict future milk production. In this study, a Wood's function (Wood, 1967) was applied using the Bayesian derivation of the Kalman filter technique (Harrison and Stevens, 1976) described by Goodwall and Sprevack (1985). The technique employs parameters calculated from previous data (historical lactations). Therefore, modifications in the parameters had been accomplished. The method is based on the state-space formulation of the linear model, which in its general form may be written as follows:

$$Y_t = Z_t\theta_t + \nu_t$$
$$\theta_t = F_t\theta_{t-1} + \eta_t \qquad (1)$$

where Y_t is a vector of observations at time t, θ_t is the state of the system and contains the vector of unknown parameters at time t, Z_t and F_t are known transformation matrices of suitable dimensions which might be known functions of time, and ν_t and η_t are error terms which are assumed mutually independent and normally distributed with covariance matrices and zero means. The following subsections describe the steps required by EBM in order to perform the predictions of future production.

Step 1. Historical database : The historical database consists of a list of positive parameters ($\hat{a}_k$, $\hat{b}_k$, and $\hat{c}_k$) estimated using a LWM from the recorded milk yields of historical lactations (10 test day records from calving to 305 days in milking). The parameters must be divided among comparable cows (similar conditions of management, same area, same parity, specific season of parity or same level of production). Because only first lactation records were analyzed in this study, the parameters were divided only by herds without the explicit use of other effects such as season of parity, age at parity, etc. When all the herds were analyzed jointly, the parameters were not divided. In addition, a pooled estimate of variance from all the regressions was computed using the following formula:

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^{n}(m_k - 3)\hat{\sigma}_k^{\,2}}{\sum_{k=1}^{n}(m_k - 3)} \qquad (2)$$

where m_k is the number of milk yields available in the record of the historical lactation k, the number three refers to the degrees of freedom required to calculate the regressions using the LWM, $\hat{\sigma}^2{}_k$ is the residual mean square error of the regressions, a measure of the degree to which the milk yields in the past lactation k deviates from an ideal LWM that uses 10 test day records in order to estimate the parameters $\hat{a}_k$, $\hat{b}_k$, and $\hat{c}_k$, and n is the number of regressions included in the historical database.

Finally, a variance-covariance matrix of the parameters included in the historical database needs to be calculated. This matrix is named $\hat{G}_0$, and the element (4, 4) of the matrix $\hat{G}_0$ was set to its estimated asymptotic value $\hat{G}_0(4,4) = \hat{\sigma}^2/(1-\hat{\alpha}^2)$. The parameter α is calculated such that $|\alpha| < 1$; in this case, the value of 0.07033 was used.

Step 2. Selecting initial parameters : Information of lactations in progress must be compared to the historical database searching for historical lactations that most resemble the current lactation. Jones (1997) called this the deviance of historical lactation k from the observed yields of the lactation in progress, and is defined as:

$$D_k = \sum_{k=1}^{m}[\ln Y_{(x_i)} - \ln W(x_i;\hat{a}_k,\hat{b}_k,\hat{c}_k)]^2 \qquad (3)$$

where $\ln Y_{(x_i)}$ is the weight in kilograms at x_i days in milking of the lactation in progress, $\ln W(x_i;\hat{a}_k,\hat{b}_k,\hat{c}_k)$ is estimated milk weight in kilograms at the same x_i days in milking according to the historical parameters $\hat{a}_k$, $\hat{b}_k$, and $\hat{c}_k$, and m is the number of test day records in the lactation in progress. The deviance is zero if the observed milk yields of the lactation in progress happen to coincide with the milk yields of historical lactation k. The weight for a historical lactation k is computed from D_k as:

$$W_k = e^{\frac{-D_k}{2\hat{\sigma}^2}} \qquad (4)$$

The weight is inversely related to the deviance, which results in low weights being assigned to parameters that are dissimilar to the record of the lactation in progress. The

result is a vector named $\hat{\theta}_0$ that contains the parameters of the historical database that most resembles the lactation in progress.

Step 3. Calculating updated parameters : The possibility of using prior information accelerates the rate of the convergence of the estimation. Therefore, after setting the vector $\hat{\theta}_o$ to the better initial parameters selected from the historical database, and setting the matrix $\hat{G}_0$ to the sample variance-covariance matrix of the parameters from the same historical database, the recursive procedure described by Goodwall and Sprevack (1985) is applied in order to estimate the updated parameters $\hat{a}$, $\hat{b}$, and $\hat{c}$ for the lactation in progress. Using the updated parameters, future production for the current lactation can be predicted from early lactation stages (even from the first test day record). The early prediction can be obtained because the method uses prior information as starting values. The recursive procedure, which gives the estimates of the mean of $\hat{\theta}_t$ and of its variance-covariance matrix, can be summarized as follows:

The state predictor is:

$$\hat{\theta}_t^{(p)} = F_{t-1}\hat{\theta}_{t-1} \quad (5)$$

The predicted variance-covariance matrix of the error term is:

$$P_t = F_{t-1}G_{t-1}F'_{t-1} + W_{t-1} \quad (6)$$

Thereafter that the filter gain is calculated as follows:

$$K_t = P_t Z'_t \left[Z_t P_t Z'_t + V_t\right]^{-1} \quad (7)$$

The estimate of the state at time t becomes:

$$\hat{\theta}_t = \hat{\theta}_t^{(p)} + K_t[Y_t - Z_t\hat{\theta}_t^{(p)}] \quad (8)$$

The variance-covariance matrix becomes:

$$G_t = \left[I - K_t Z_t\right]P_t \quad (9)$$

where I is the identity matrix of suitable dimensions.

To express the above procedure for the lactation curve in terms of the state-space representation, the following formula is given:

$$\theta_t = \begin{bmatrix} \ln a \\ b \\ c \\ \ln \varepsilon_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \alpha \end{bmatrix} \begin{bmatrix} \ln a \\ b \\ c \\ \ln \varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \varepsilon'_t \end{bmatrix} \quad (10)$$

$$Y_t = [1 \ \ln t \quad -t \quad 1]\,\theta_t \quad (11)$$

where $\ln \varepsilon_t$ is a modification introduced by Goodwall and Sprevack (1985) to LWM in which $\ln \varepsilon_t$ is assumed to follow a first-order autoregressive model, ε'_t is a sequence of independently distributed normal random errors with mean zero and variance $\hat{\sigma}^2$, and α was defined in step one.

Current methods

The future cumulative prediction of the lactation obtained by EBM was compared with predictions estimated by three current methodologies. The LWM is an algebraic model for the lactation curve in dairy cattle reported by Wood (1967), the TIM is officially used in many countries, and the BPE (VanRadem, 1997) which applies best prediction to test day records; this method uses previously established correlations between individual test days and includes an inversion of a matrix for each lactation record (Norman et al., 1999). In this study, TIM was used to estimate the cumulative yields up to the last test day, and then the projection factors described by Miller et al. (1972) as method P were applied to predict the 305 d milk production.

Statistical analysis

Several Fortran programs were developed in order to estimate the prediction of 305 days production from early records using the current methodology and the EBM. The availability of real milk production over the entire lactation period due to the daily milk recording, permitted comparison of the methods by calculating the absolute difference (AD) between the real and the predicted 305 d milk production. In addition, a standard deviation (SD) of the differences was also calculated. Descriptive statistics of data were performed using the Means procedure of SAS (SAS System for Linear models, Third edition, SAS Institute, Cary, NC, USA).

RESULTS AND DISCUSSION

The results showed that in the early stage of the lactation, EBM was superior in obtaining the prediction with high accuracy. When all the herds were analyzed jointly (figure 1), the AD between the real and the predicted 305 d milk production during the first five test day records were on average, 373, 590, 917 and 1,042 kg for EBM, BPE, TIM, and LWM, respectively. The SD of the differences between real and predicted 305 d milk production (table 1) obtained by the EBM were smaller than the differences obtained by the other methods. Corresponding SD for EBM, BPE, TIM, and LWM during the first five test days were on average, 488, 733, 747 and 1,605 kg. When the herds were analyzed separately, the results showed that except for small variations the patterns observed in all data sets were equal. Figures 2 to 4 show that in early lactation stages (first to fifth test day record), the EBM predictions retained its high accuracy. To compare

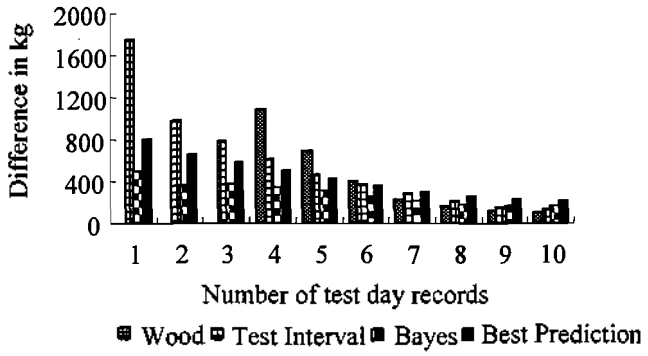

Figure 1. Absolute difference between real and predicted 305 d milk production by four methods in all the herds

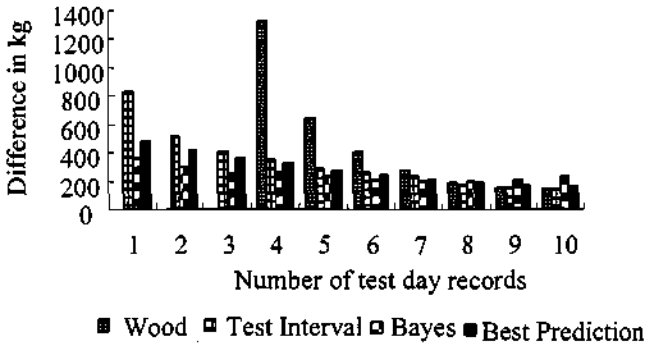

Figure 2. Absolute difference between real and predicted 305 d milk production by four methods in the first herd

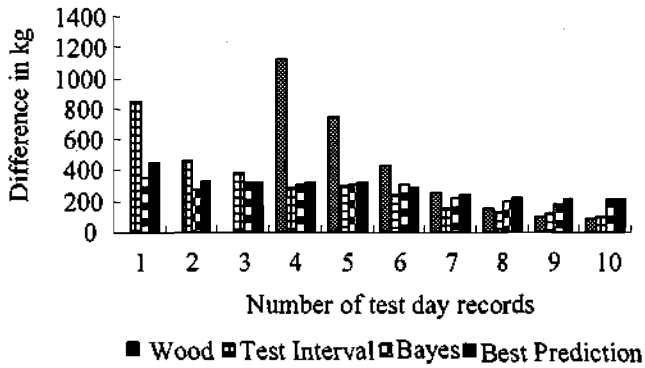

Figure 3. Absolute difference between real and predicted 305 d milk production by four methods in the second herd

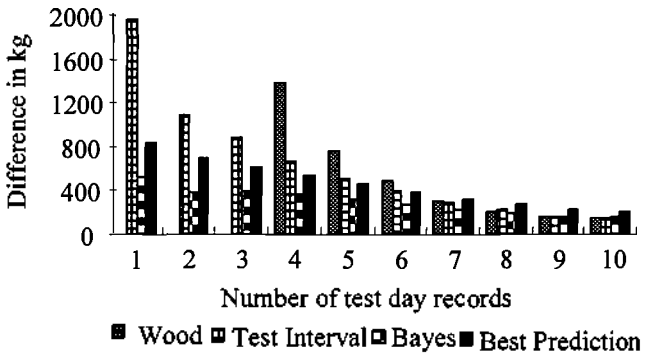

Figure 4. Absolute difference between real and predicted 305 d milk production by four methods in the third herd

the methods during the first five test day records the average of the AD was calculated. The average of the AD for the EBM predictions were 278, 325 and 392 kg in herds one, two, and three, respectively. During the same period, the BPE predictions were on average, 476, 344 and 619 kg in herd one, two, and three, respectively. The average of the AD for the TIM predictions were 474, 453 and 1,011 kg in herds one, two and three, respectively. LWM was unable to predict future production unless at least three test day records of the lactation in progress were accumulated. The results of the LWM agree with Cobby and Le Du (1978), those researchers found that LWM was unable to predict future production accurately when only small amounts of information of the current lactation were available.

When the whole lactation period was analyzed the accuracy of the EBM showed small variation from early stages through the last period of the lactation. The AD of the EBM throughout the 10 test day records never attained differences greater than 600 kg (figures 1 to 4). Moreover, the accuracy of EBM increased when more information of the current lactation was added to the prediction process. The increment of the accuracy of the EBM predictions stopped in two of the herds (herds one and two) in the last period of the lactation, and a decrease in the EBM predictions was observed. The BPE showed similar patterns to the EBM through the whole lactation period. In contrast, LWM and TIM showed more variability throughout the lactation period, but in general, the accuracy of both methods increased when more information about the lactation in progress was accumulated. In the last period of the lactation, the accuracy of both those methodologies exceeded the EBM and BPE predictions. The SD in the last test day record were 219, 258, 163 and 174 kg for EBM, BPE, TIM, and LWM, respectively.

In conclusion, the EBM was successful in attaining high accuracy in the prediction of 305 d milk production from early lactation records and through the whole lactation period. To achieve the predictions this method did not require accumulation of large quantities of data. Moreover, the EBM obtained high accuracy in its predictions even when the analysis did not use explicitly the effects of season at calving or of different areas. In the current practices of collecting monthly records, early prediction of future milk production may be more accurate using the EBM. Alternatively, if enough information of the lactation in progress is accumulated, TIM or LWM may obtain better

Table 1. Standard deviation (SD) of differences between real and estimated 305 days milk production in all the herds

Test day records	Wood model	Bayes method	Test interval	Best predicted
	SD (kg)	SD (kg)	SD (kg)	SD (kg)
1	-	631	1,091	994
2	-	471	824	813
3	-	483	728	710
4	2,218	455	614	620
5	992	402	484	529
6	614	347	407	451
7	396	286	310	379
8	265	245	227	326
9	194	219	179	279
10	174	218	163	258

accuracy in the latter stage of lactation. In countries where previous information of past lactations is limited, EBM might be useful for predicting future milk production even in herds with reduced number of cows. However, further studies about causes of the fall of accuracy in the last period of the lactation when using the EBM should be investigated. In addition, if the EBM is to be used in genetic evaluations, the influence of areas, seasons, ages, and parities in the division of the historical database according to comparable cows should be carefully investigated.

REFERENCES

Cobby, J. M. and Y. L. P. Le Du. 1978. On fitting curves to lactation data. Anim. Prod. 26:127-133.

Conglenton jr, W. R. and R. W. Everett. 1980. Application of the incomplete gamma function to predict cumulative milk production. J. Dairy Sci. 63:109-119.

Goodwall, E. A. and D. A. Sprevack. 1985. A Bayesian estimation of the lactation curve of dairy cow. Anim. Prod. 40:189-193.

Harrison, P. J. and C. F. Stevens. 1976. Bayesian Forecasting. J. Royal Statistics Soc. 38:205-248.

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

Maritz, J. S. and T. L. Lwin. 1989. Empirical Bayes methods. 2nd. Ed. Chapman and Hall, New York.

Miller, R. H., R. E. Pearson, M. H. Fohrman and M. E. Creegan. 1972. Methods of projecting complete lactation production from part-lactation yield. J. Dairy Sci. 55:1602-1606.

Norman, H. D., P. M. VanRadem, J. R. Wright and J. S. Clay. 1999. Comparison of test interval and best prediction methods for estimation of lactation from monthly, a.m.-p.m., and trimonthly testing. J. Dairy Sci. 82:438-444.

SAS Institute Inc. 1991. SAS System for linear models. 3rd. Ed. SAS Institute Inc., Cary, North Carolina.

VanRadem, P. M. 1997. Lactation yields and accuracies computed from test day yields and (co)variances by best prediction. J. Dairy Sci. 80:3015-3022.

Wood, P. D. P. 1967. Algebraic model of the lactation curve in cattle. Nature (Lond.). 216:164-165.