Kasetsart J. (Nat. Sci.) 42 : 246 - 255 (2008)

Comparison of Five Mathematical Functions for Prediction of Monthly Yield in Thai Multibreed Dairy Cattle Population

Warangkana Kitpipit[1,2], Panwadee Sopannarath[1],
Sayan Buaban[3] and Sornthep Tumwasorn[1*]

ABSTRACT

Monthly yields of 32,583 records from 3,523 first lactation cows, collected by the bureau of Biotechnology in Livestock Production, Department of Livestock Development (DLD) were used in this study. These dataset were obtained between 1993 and 2003 in 777 farms throughout the country. The objective was to evaluate five mathematical models for describing the phenotypic lactation curve of dairy cattle in Thai multibreed population. In the previous analysis, monthly milk yields were adjusted by the variation of herd-year-season of calving, Holstein Friesian (HF) fractions, calving age and first test-day after calving. The adjusted monthly milk yields were used to fit five mathematical functions. Five functions, i.e. Ali and Schaeffer polynomial (A&S), Schaeffer and Dekkers (S&D) logarithm, Wilmink exponential (WIL), Wood incomplete gamma (WD) and Legendre polynomials (LEG) functions were used in this study. For the whole data analysis, all functions showed few differences in comparison criteria. A&S fitted the actual yield better than other functions according to the adjusted coefficient of determination (R^2_{adj}=0.39) and mean square error (MSE=6.84). For individual lactation analysis, A&S provided the smaller value of residual errors including means of absolute error (RES), means of mean square error (MMSE) and means of square prediction error (MSPE) than other functions. Ranking all functions from the best to the worst were A&S, WD, S&D, WIL and LEG, respectively. The results indicated that A&S function could be fitted well for the whole and individual data. Therefore, A&S could be applied to describe the phenotypic lactation curve in this dairy population.

Key words: dairy cattle, lactation curve, mathematical functions, monthly milk yield

INTRODUCTION

Lactation curve is an efficient tool for selection criteria and management purpose in dairy farming because the variation of lactation curve can influence dairy farm profitability. Accurate description of the lactation curve is important to activities such as conducting feeding trials with lactating cattle, estimating total yield from incomplete records, and forecasting herd performance on a whole herd or individual cow basis (Val-Arreola *et al.,* 2004).

The analysis of relationship between mathematical function properties and lactation patterns has been focused mainly on the evaluation of fitting milking performances. A variety of

---

[1] Department of Animal Science, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand.
[2] Faculty of Animal Science and Agricultural Technology, Silapakorn University, Phetchaburi 76120, Thailand.
[3] Bureau of Biotechnology in Livestock Production, Department of Livestock Development, Bangkok 12000, Thailand.
[*] Corresponding author, e-mail: sornthep@ksc.th.com

Received date : 09/10/07 Accepted date : 31/01/08

different functions has been used to predict the lactation pattern in several researches (Sherchand *et al.,* 1995; Olori *et al*., 1999; Vargas *et al*., 2000; Macciotta *et al*., 2005). Wood's (1967) function has been widely used by a number of researches such as Olori *et al*., 1999; Tekerli *et al.* (2000); Macciotta *et al*., (2005). The Wilnink (1987) function was developed in The Netherlands and was the original function used to describe the lactation curve in the official program of genetic evaluation in Canada (Schaeffer *et al*., 2000). The logarithm function was developed by Ali and Schaeffer (1987) and has been used in random regression test day model later study (Jamrozik and Schaeffer, 1997). Moreover, Schaeffer and Dekker (1994) developed a function to describe lactation curve based on logarithm function and it was applied to genetic evaluation. These functions have been used to fit the lactation curve. Currently, Legendre polynomial has been applied to fit the lactation curve (Macciotta *et al*., 2005).

Numerous studies have investigated the mathematical functions to describe milking ability in many tropical countries such as USA, Canada, Italy, England, Portugal and France (Ali and Schaeffer, 1987; Olori *et al*., 1999; Druet *et al*., 2003; Macciotta *et al*., 2005; Silvestre *et al*., 2006). However, there are problems in directly applying those functions in dairy cattle raised in Thailand due to differences in climate (hot and humid) and varying in individual genetic difference. In Thailand, Holstein Friesian and Zebu breed, such as Brahman or Native cattle, are major breeds for dairy population. According to the characteristic of the dairy population, purebred and crossbred are equal in being chosen the parent of the new generation. Such dairy population is, thereby known as multibreed (Koonawootrittriron *et al.,* 2002). Hence, fitting the mathematical function to describe the lactation pattern under these conditions (multibreed and tropical environment) would need lot more information than those with single breed population raised under the temperate environments. Thus, the objective of this study was to compare the goodness of fit of five mathematical functions used for describing lactation curve in Thai crossbreed dairy cattle population.

MATERIALS AND METHODS

Data source and structure

Monthly test-day (TD) yields of 32,583 records from 3,523 of first lactating dairy cows were collected by the bureau of Biotechnology in Livestock Production, Department of Livestock Development (DLD), Thailand. Data were collected from 777 farms located in different parts of the country between 1993 and 2003. All farmers were collaborated under the Master Bull Project. The dataset were stored in the database containing information on herd (i.e. farm identification number), animals (i.e. identification number, breed fractions of HF), calving date, day in milk (DIM) and TD yields. The milking program measured milk yield once a month of each cow after calving until drying off. Cows were milked twice a day under the same nutritional (concentrate and roughage) and management system in this herd.

In this study, calving age was restricted between 18 and 52 months. The interval from calving to first test date was restricted from 5 to 35 days. Calving seasons were classified as summer (March to June), rainy (July to October), and winter (November to February). Contemporary groups were defined as herd-year-season of calving (HYS). Breed groups were classified by considering the fraction of Holstein Friesian (HF) gene (0.0 to 1.0HF). Editing was on lactation length more than 240 days and the number of TD per cow more than 6 records. Data were adjusted for the fixed effects of herd-year-season of calving, the fraction of Holstein gene, calving age and first test day prior to analysis. The data structure is shown in Table 1.

Table 1 The general descriptive data structure for analysis.

Item	
Number of cows	3,523
Number of herds	777
Number of test-day records	32,583
Average test-day yield (kg) ± S.E.	13.09 ± 4.45
Average age at calving (mo) ± S.E.	33.00 ± 6.31
Average lactation length (days) ± S.E.	352.86 ± 83.68

Mathematical functions and statistical analysis

Five mathematical functions were used to evaluate the lactation curve by the Marquardt method using PROC NLIN of SAS program (SAS, 1990) and estimated all lactation curves parameters. These functions were described by the following formulas:

Function 1: Ali and Schaeffer polynomial function (A&S) (Ali and Schaeffer, 1987)

$y^*_t = a + b\times(t/305) + c\times(t/305)^2 + d\times[\ln(305/t)] + k\times[\ln(305/t)]^2 + e_t$

Function 2: Schaeffer and Dekkers logarithm function (S&D) (Schaeffer and Dekkers, 1994)

$y^*_t = a + b\times\ln(305/t) + c\times t + e_t$

Function 3: Wilmink exponential function (WIL) (Wilmink, 1987)

$y^*_t = a + b\times exp^{-0.05t} + c\times t + e_t$

Function 4: Wood incomplete gamma function (WD) (Wood, 1967)

$y^*_t = at^b\, exp^{-ct} + e_t$

Function 5: Legendre polynomials function (LEG) (Gengler *et al.*, 1999)

$y^*_t = a\times L1 + b\times L2 + c\times L3 + e_t$

where y^*_t is adjusted test-day milk yield on DIM t, a, b, c, d and k are parameters to be estimated. $L1$ is 1, $L2$ is $\sqrt{3}L$, $L3=\sqrt{5}/4\times(3L^2-1)$; $L=(-1)+2\times(t-1)/(305-1)$ and e_t is residual error.

Comparison criteria for functions

All functions were fitted to overall and individual records to consider the residual errors.

The main criteria were evaluated the goodness of fit of all functions following these criteria (Neter *et al.*, 1985):

1). Mean square error, (MSE);

$MSE = SSE/(n-p-1)$

where SSE = sum squares error, n = number of observations, and p = number of parameters in each function.

2). Adjusted coefficient of determination, (R^2_{adj});

$$R^2_{adj} = 1 - \frac{SSE/(n-p-1)}{SST/(n-1)}$$

where SSE = sum squares error, SST = sum squares total, n = number of observations, and p = number of parameters in each function.

3). Mean absolute error, (RES);

$$RES = \frac{1}{n}\sum_{t=1}^{n}\left|y^*{}_t - \hat{y}^*{}_{ti}\right|$$

where n = number of observation, y^*_t = actual milk yield and $\hat{y}^*_i$ = predicted milk yield

4). Mean of mean square error, (MMSE);

$$MMSE = \frac{1}{N}\times(MSE)$$

where N = number of animal, MSE = mean square error

5). Mean square prediction error, (MSPE);

$$MSPE = \frac{1}{n}\sum_{t=1}^{n}\left(y^*{}_t - \hat{y}^*{}_t\right)$$

where n = number of observation, = actual milk yield and = predicted milk yield

Criteria 1) and 2) were considered the goodness of fit for fitting five functions to overall data, while 3), 4) and 5) were considered the goodness of fit for individual lactation. The mathematical function with the best fit according to previous criteria was selected to fit the lactation curve of dairy cattle. The small values of residuals were considered to be the best fit function. For the reason, that less residual variation remained in the data. In addition, the predicted and actual curve obtained from five functions and residual distributions in each function were plotted using Microsoft Excel 2000 (Dodge and Stinson, 1999).

RESULTS AND DISCUSSION

Prediction of average lactation curve

Five mathematical functions were fitted with 32,583 monthly milk records from 3,523 first lactating cows. The estimated coefficients and statistical criteria from all functions used overall data analysis are represented in Table 2.

Figure 1 represents the average actual and predicted lactation curves which were analyzed from overall data set. The results showed that the actual and predicted curves were almost similar for A&S, WD, WIL and S&D functions while a minor difference was observed from LEG which was slightly overestimated in the early and the end of lactation period.

The results found that estimated R^2_{adj} ranged from 0.38 to 0.39. Given the results from the analyses, all functions were accepted LEG provided identically values of R^2_{adj} (0.38). It seemed to be little difference among functions. In addition, the values of MSE ranged from 6.84 to 6.92. A&S provided more accurate than other functions which considered from the higher and lower values for R^2_{adj} and MSE. In contrast LEG fitted overall data with the lower accuracy. Ranking functions ability to describe lactation curves from the best to the worst were A&S, WD, WIL, S&D and LEG, respectively. Hence, A&S performed better biologically and statistically, as it was more closely similar the underlying process in this population. These results were different from Mekvilai (1998) who reported that Wood's gamma function (R^2_{adj}=0.22) was the most suitable function for describing the lactation curve of dairy cows under the Government Dairy Promotion Project in Thailand. However, it was not possible to directly compare the results generated in this study with other studies, such as Tozer and Huffaker (1999); Vargas *et al.* (2000); Silvestre *et al.* (2006), due to the principle of population structure, data recording

Table 2 Estimated parameters of average lactation curve, adjusted coefficient of determination (R^2_{adj}) and mean square error (MSE) for different functions.

Function[1/]	Predicted functional forms	R^2_{adj}	MSE
A&S	$\hat{y}^*_t$ = 29.303 –28.810× (t/305)+9.048× $(t/305)^2$– 6.277× ln(305/t)+ 0.720× $\ln(305/t)^2$	0.39	6.84
WD	$\hat{y}^*_t$ = 12.933× $t^{0.086}$ $\exp^{-0.003\times t}$	0.39	6.85
WIL	$\hat{y}^*_t$ = 17.366– 2.901× $\exp^{-0.05\times t}$ – 0.027× t	0.39	6.86
S&D	$\hat{y}^*_t$ = 18.694 – 0.806× ln(305/t) – 0.032× t	0.39	6.87
LEG	$\hat{y}^*_t$ = 13.070 – 2.165×L2 – 0.140×L3	0.38	6.92

[1/] WD=Wood incomplete gamma function, A&S=Ali and Schaeffer polynomial function, S&D=Schaeffer and Dekkers logarithm function, LEG=Legendre polynomial function and WIL=Wilmink exponential function

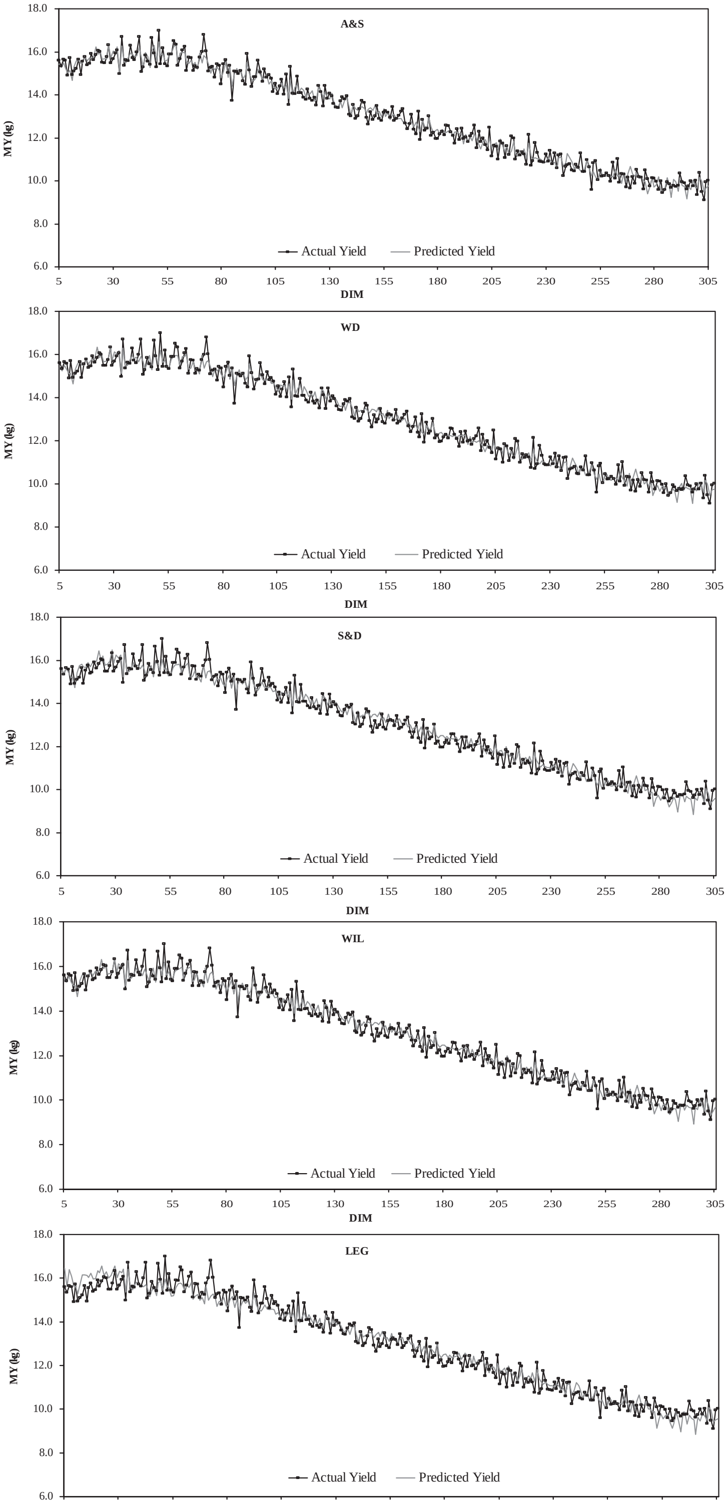

Figure 1 The actual and predicted lactation curves fitted form five functions.

interval and different breed of cows used in the particular studied.

Prediction of individual lactation curve

Five functions were fitted to individual lactation curve to determine the goodness of fit. The goodness of fit based on RES, MMSE and MSPE were used to select the most fitted function represented in Table 3. The results found that least square means of RES, MMSE and MSPE had significant difference from other functions ($P<0.05$). A&S and LEG provided smaller and higher value of residual errors respectively. Ranking functions ability to describe lactation curves from the best to the worst were A&S, WD, S&D, WIL and LEG respectively. All residual error measurements from S&D and WIL had no significant difference ($P>0.05$). It suggested that the S&D and WIL functions could predict milk yield very closely because both functions are applied from the same original function (Schaeffer and Dekker; 1994). For this reason, it indicated that both functions might be used to describe the lactation curve in the same manner.

It could be observed that A&S function fitted the individual lactation better than other functions. While, WD function performed more accurate in fitting individual lactation than S&D, WIL and LEG, respective. It suggested that A&S could be used to describe the lactation curve in dairy cattle which were raised under the same conditions and composed with many breed structure. Also for these three statistical criteria, some differences between curves were marginal. Considering of goodness of fit, it is not trivial to conclude which of these curves is the best and different ranking might be obtained across sample or studied. Moreover, a compromise, often based on subjective grounds, must be accepted between goodness of fit and other properties such as flexibility, robustness and computational considerations. This explained the large variation in models across studies and countries (Druet *et al.*, 2003).

However, these results were different those of from Koonawootrittriron *et al.* (2001) which showed that the second-degree polynomial model was the best fit to predict 305-d milk yield using monthly test-day data in a Holstein Friesian-Red Sindhi herd in the Northeast of Thailand. The difference may be caused from the data structure of population. Furthermore, Rowlands *et al.* (1982) compared four types of mathematical functions used to describe the lactation curve in British Friesian and reported that WD function fitted the weekly yield slightly better on average than others. The differences may be caused from different genetic factors, management strategies and time interval in that particular population

Table 3 Comparison of the goodness of fit of mathematical functions to describe the lactation curve based on means of absolute error (RES), means of mean square error (MMSE) and means of square prediction error (MSPE).

Functions[1/]	Goodness of fit measurements		
	RES	MMSE	MSPE
A&S	0.566^{a}	2.323^{a}	0.801^{a}
WD	0.846^{b}	2.641^{b}	1.497^{b}
S&D	0.867^{c}	2.753^{bc}	1.561^{c}
WIL	0.860^{c}	2.820^{c}	1.598^{c}
LEG	0.910^{d}	2.951^{c}	1.680^{d}

[a,b,c,d] least square means within the column with different superscripts differ ($P < 0.05$)

[1/] WD=Wood incomplete gamma function, A&S=Ali and Schaeffer polynomial function, S&D=Schaeffer and Dekkers logarithm function, LEG= Legendre polynomial function and WIL=Wilmink exponential function

(Olori *et al*., 1999; Tozer and Huffaker, 1999; Tekerli *et al*., 2000; Rekik *et al*., 2003).

Figure 2 shows the average residuals distribution (differences between actual and predicted milk yield) in the range from DIM 5 to 305. It was possible to see that the pattern of residual distribution from A&S function was closed to zero (actual minus predicted yield is not

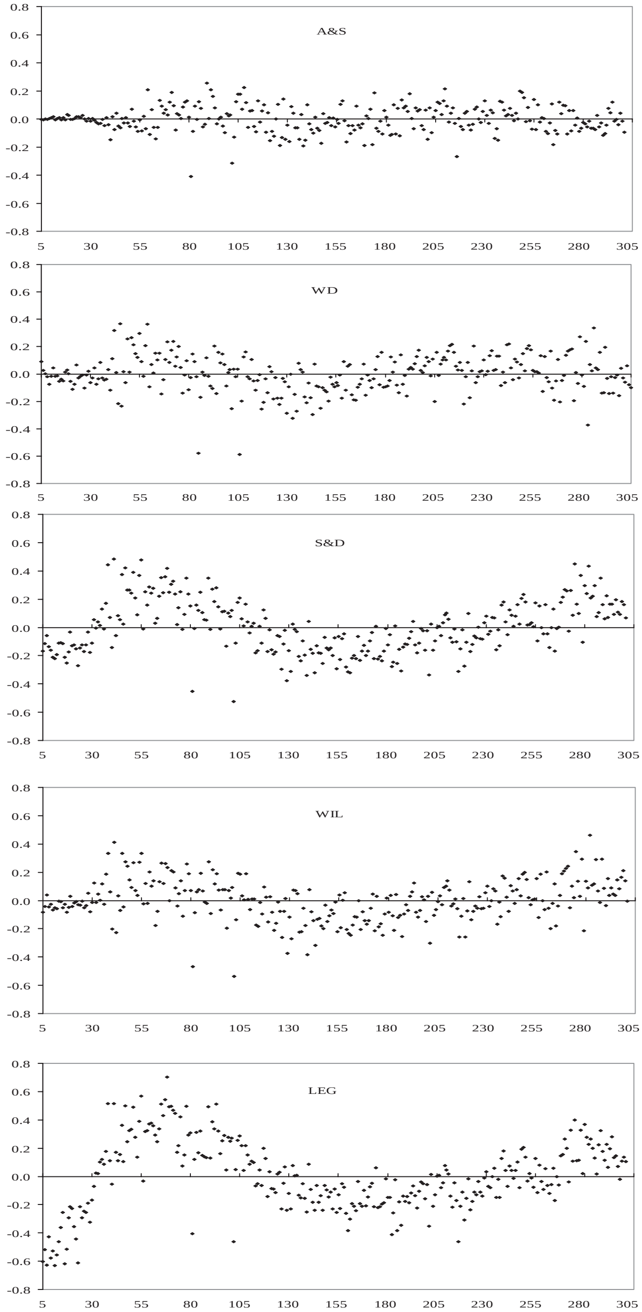

Figure 2 Distribution of average residuals (kg/d) for DIM 5 to 305 for Ali and Schaeffer (A&S), Wood (WD), Schaeffer and Dekkers (S&D), Wilmink (WIL) and Legendre polynomial (LEG) functions.

more different), especially in early of lactation. It performed consistently through the lactation, indicating more satisfactory description of the lactation curve by this function. Moreover, the residuals distribution from WD, S&D and WIL fluctuated throughout the lactation between positive and negative value. The distribution pattern from these functions little underestimated after calving between DIM 5 and 35; extremely overestimated around peak yield and the end of lactation. On the contrary, LEG function provided more variation of residual distribution than the other functions, especially after calving until peak period. These results were also related to the goodness of fit in Table 3. Perhaps the genetic difference caused to the dissimilarity of curve, especially in the hot and humid condition. It could imply that the early lactation fitted quite well with polynomial form. Thus, A&S function may be suitable describing the milking ability of dairy cows in this population.

These results obtained were similar to those of Olori *et al.* (1999) who found that A&S and WIL provided residuals error smaller than WD function. However, Scott *et al.* (1996) found that WD function overestimated milk yield after calving up to 10 weeks and underestimated milk yield form 10 to 20 weeks of lactation. The difference of fluctuation throughout the lactation might depend on the peak yield, breed compositions, stage of lactation and seasonal effects (Cobby and Le Du, 1978; Rowlands *et al.*, 1982; Scott *et al.* 1996; Koonawootrittriron *et al.* 2001). The results of this research support the idea that the performance of each function depends on the sample quality but also shows considerable variation within the sampling groups. In this study, A&S function could be selected to describe the lactation pattern and applied for genetic evaluation using random regression test-day model.

CONCLUSION

For whole lactation analysis, A&S was the best suitable to describe the lactation curve which provided the high value of R^2_{adj} (0.39) and the less value of MSE (6.84). For individual lactation, A&S could describe the individual lactation curve better than others. It could be concluded that for the whole and individual lactation curve analysis the results were consistent in the same way. Ranking all functions from the best to worst were A&S, WD, S&D and WIL and LEG, respectively. The ability of function fitting depended on the data collecting interval; furthermore, the variation between cows and general form of fitting function should be considered within sampling schemes. Thus, A&S might be the most suitable describing the characteristic of lactation pattern in Thai multibreed dairy cattle under tropical condition.

ACKNOWLEDGEMENTS

The authors are grateful for the Ph.D grant from the Commission on Higher Education, the research funded by the graduate school and thank the bureau of Biotechnology in Livestock Production, Department of Livestock Development (DLD) for providing the valuable data. Finally, the authors would like to specially thank Dr. Monchai Duangjinda for his guidance.

LITERATURE CITEDS

Ali T.E. and L.R. Schaeffer. 1987. Accounting for covariance among test day milk yields in dairy cows. **Can. J. Anim. Sci**. 67: 637–644.

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

Dodge M. and C. Stinson. 1999. **Running Microsoft Excel 2000**. Microsoft Press, Washington, 994 p.

Druet T., F. Jaffrâzic, D. Boichard and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. **J. Dairy. Sci**. 86: 2480-2490.

Gengler, N., A.Tijani, G.R Wiggans, C.P. Van Tassell and J.C. Philpot. 1999. Estimation of (co)variance of test day yields for first lactation Holsteins in United States. **J. Dairy. Sci**. 80: 2340–2354.

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.

Koonawootrittriron, S., M.A. Elzo, S. Tumwasorn and W. Sintala. 2001. Prediction of 100-d and 305-d milk yields in a multibreed dairy herd in Thailand using monthly test-day records. **Thai J. Agri. Sci**. 34: 163-174.

Koonawootrittriron, S., M.A. Elzo, S. Tumwasorn and K. Nithichai. 2002. Estimation of covariance components and prediction of additive genetic effects for the first lactation 305-d milk and fat yields in a Thai multibreed dairy population. **Thai J. Agri. Sci**. 35: 245-258.

Macciotta, N.P.P., D. Vicario and A. Cappio-Borlino. 2005. Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. **J. Dairy. Sci**. 88: 1178-1191.

Mekvilai., A. 1998. **Lactation Curve of the First Lactation Period of Holstein Friesian Crossbred in Thailand.** M.S. thesis, Chulalongkorn University, Bangkok.

Neter, J., W. Wasserman and M.H. Kutner. 1985. **Methods Applied Statistical**. 2nd ed. Richard D. Irwin, Inc., Homewood, IL. 682 p.

Olori, V.E., S. Brotherston, W.G. Hill and B.J. McGuirk. 1999. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. **Livest. Prod. Sci**. 58: 55-63.

Rekik, B., A.B. Gara, M.B. Hamouda and H. Hammami. 2003. Fitting lactation curves of dairy cattle in different types of herds in Tunisia. **Livest. Prod. Sci**. 83: 309-315.

Rowlands, G.J., S. Lucey and A.M. Russel. 1982. A comparison of different models of the lactation curve in dairy cattle. **Anim. Prod**. 35: 135-144.

SAS., 1990. **SAS/STAT User's Guide**. 4th ed. SAS Institute Inc., Cary, NC.

Schaeffer, L.R. and J.C.M. Dekkers. 1994. Random regression in animal models for test day production in dairy cattle, p 433. *In* **Proc. 5th World Congr.** Genet. Appl. Livest. Prod., Guelph.

Schaeffer, L.R., J. Jamrozik, G.J. Kistemaker and B.J. Van Doormaal. 2000. Experience with a test-day model. **J. Dairy Sci**. 83: 1135–1144.

Sherchand, L., R.W. McNew, D.W. Kellogg and Z.B. Johnson. 1995. Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. **J. Dairy. Sci**. 78: 2507-2513.

Scott, T.A., B. Yandell, L. Zepeda, R.D.Shaver and T.R. Smith. 1996. Use of lactation curves for analysis of milk production data. **J. Dairy. Sci**. 79: 1885-1894.

Silvestre, A.M., F. Petim-Batista and J. Colaco. 2006. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. **J. Dairy. Sci**. 89: 1813-1821.

Tekerli, M., Z. Akinci, I. Dogan and A. Akcan. 2000. Factors affecting the shape of lactation curves of Holstein cows from the Balikesir province of Turkey. **J. Dairy. Sci**. 83: 1381-1386.

Tozer, P.R. and R.G. Huffaker. 1999. Mathematical equation to describe lactation curves for Holstein-Friesian cows in New South Wales. **Aust. J. Agric. Res**. 50: 431-440.

Val-Arreola, D., E. Kebreab, J. Dijkstra and J. France. 2004. Study of the lactation curve in dairy cattle on farms in central Mexico. **J. Dairy. Sci**. 87: 3789-3799.

Vargas, B., W.J. Koops, H. Herrero and J.A.M. Van Arendonk. 2000. Modelling extended lactations of dairy cows. **J. Dairy. Sci**. 83: 1371-1380.

Wilmink, J.B.M. 1987. Adjustment of test day milk, fat and protein yield for age, season and stage of lactation. **Livest. Prod. Sci**. 16: 335-348.

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