Listy Biometryczne - Biometrical Letters
Vol. 36(1999), No. 2, 169-178

Evaluation of selected mathematical models of lactation curves

Wojciech Piotr Perz and Zbigniew Sobek

Department of Genetics and Animal Breeding, Poznań Agricultural University, Wołyńska 33, 60-637 Poznań, Poland

SUMMARY

An evaluation of 21 lactation curve models, concerning accuracy of 100-day milk yield estimation in cows based on several (4-7) test yields – a problem occurring in breeding practice – was made. Stability of the models (the lowest possible estimation error variance) was considered, instead of the most frequently used determination coefficient criterion. The results confirm the assumption adopted in this approach that simple models, although they do not exactly render the real course of lactation curve, are less vulnerable to the fluctuations appearing in the data. Thus, they give more stable prediction results, at the same time not needing intensive computations necessary to obtain parameters of complex models.

KEY WORDS: lactation curve, milk yield, milk yield evaluation.

1. Introduction

Advantages resulting from the knowledge of the course of lactation curves for milk industry are given by Weigel et al. (1991). Such information can be used to supplement incomplete lactations with the aim to employ them for genetic evaluation of cattle, determination of their rations or monitoring of the share of a given animal in the overall productivity of a herd. Furthermore, the lactation curve is also used for economical assessment in various management schemes. The knowledge of the lactation curve can also be useful in disease diagnostics in cases of mastitis or ketosis (Detilleux et al., 1994). A recently proposed model of the curve of lactation yield is a multiphase logistic function.

The area under the curve portrays the milk yield and each point of the curve designates daily yield (DY). Milk yield for a given period (in our case for 100 days) can be calculated in two ways, depending on which of them is easier from the point of view of calculation: either by adding up daily yields for consecutive days of a given time interval or by calculating the definite Riemann integral of the function describing the curve.

The objective of this study was to find a function which would allow to estimate the yield of true lactation most precisely.

2. Material and methods

Error variances of estimations of 100-day milk yields of cows calculated with the help of models represented here by equations (1)–(22) were compared:

$$y = a + bx \tag{1}$$

$$y = a + bx + cx^2 \tag{2}$$

$$y = a + bx + cx^2 + dx^3 \tag{3}$$

$$y = a + bx + cx^2 + dx^3 + ex^4 \tag{4}$$

$$y = a + bx + cx^2 + dx^3 + ex^4 + fx^5 \tag{5}$$

$$y = a + bx + cx^2 + dx^3 + ex^4 + fx^5 + gx^6 \tag{6}$$

$$y = \beta_0 x^{\beta_1} \tag{7}$$

$$y = \beta_0 e^{\beta_1 x} \tag{8}$$

$$w(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)(x - x_1) + ... + c_n(x - x_0)(x - x_1)...(x - x_{n-1}) \tag{9}$$

where $w(x)$ is the value of the n-th degree polynomial for any value of x (in our case, it is the yield of a cow for any day of lactation in the 100-day period), $x_0, x_1, ..., x_{n-1}$ – represent days of test milking, $c_0, c_1, ...c_n$ – are coefficients of the Newton interpolating polynomial which can be calculated using condition (10),

$$w(x_i) = y_i, \quad i = 0, 1, ..., n, \tag{10}$$

where $n + 1$ is the number of test milkings,

$$y = a\exp(bx) - a\exp(cx) \tag{11}$$

$$y = a\exp(bx - cx^2) \tag{12}$$

$$y = x/(a + bx + cx^2) \tag{13}$$

$$y = ax^b\exp(cx) \tag{14}$$

$$y = a - bx + c\ln(x) \tag{15}$$

$$y = a + bx + cx^2 + d\ln(x) \tag{16}$$
$$y = ax^{mc}\exp(-cx) \tag{17}$$
$$y = ax^b/\cosh(cx) \tag{18}$$
$$y = a[1 - \exp(bx)]/\cosh(cx) \tag{19}$$
$$y = a\arctan(bx)/\cosh(cx) \tag{20}$$
$$y = a\arctan(bx)\exp(-cx) \tag{21}$$
$$y = a_1\{1 - \tanh^2[b_1(x^k - c_1)]\} + a_2\{1 - \tanh^2[b_2(x - c_2)]\}. \tag{22}$$

Coefficients of equations (1) to (8) were determined using the least squares method (Martin, 1976). The problem of finding values of coefficients of the interpolating polynomial presented in the form (9) is reduced to the calculation of difference quotients of consecutive orders (Bronsztejn and Siemiendiajew, 1998) which is much simpler than solving the polynomial regression equations for functions (1) – (8) containing moreover values such as $\Sigma x_i y_i$ (Martin, 1976). Coefficients of equations (11) – (22) were determined with the assistance of the shareware statistical package Lisp-Stat which uses an iterative approach. The analysis of the curves described by equations (11) – (22) can be found in the paper by Sherchand et al. (1995). These curves have been widely used in disease diagnostics or in animal nutrition.

Equation parameters were determined using 4, 5, 6 or 7 daily yields selected from the 100-day lactation period from the day of calving and treated as test milkings.

Models were evaluated on the basis of empirical results of weight measurements of milk obtained every day from 101 cows of pure Holstein-Friesian (h-f) breed or hybrids with a high h-f upgrading. These cows were randomly selected from two herds comprising about 1000 animals from the Experimental Station of the Institute of Animal Science in Pawłowice near Leszno. The first group of 29 experimental animals comprised cows which began their lactations between January and April (winter season 1995). The second group of 31 cows began their lactations in the period from September to October (summer season 1996), while the third group consisted of 41 cows which calved in the period from October to December 1996.

Researchers who undertake investigations on the problem of the choice of the most suitable mathematical model of the lactation curve usually adopt the determination coefficient R^2 as the evaluation criterion. Such an approach favours models which are capable of adjusting to data in the most flexible manner. In breeding practice the most important consideration is the question how accurately will a given function with coefficients determined from only a few data (test milkings) allow researchers to calculate milk yield in the entire 100-day period. The shape of the curve (graph of the function) and the faithfulness with which it depicts changes in the daily milk yield in time are less significant. That is why the authors of this paper evaluated models from the point of view of the accuracy of approximation of the 100-day yield.

The authors of the study had at their disposal data concerning daily milk production during the first 100 days of lactation, so the true 100-day yield was known. The breeding situation in which only a small number of daily yields is known was simulated by selecting, from among 100 data for each cow, only a few observations in defined days of lactation corresponding to 4, 5, 6 or 7 test milkings (Perz, 1998). The models presented in the above mentioned study were analysed from the viewpoint of selection of sampling days of test milkings for which the accuracy of milk yields assessment was the highest. The author employed the method of counts (sets of days of test milk sampling which meet the criterion of accuracy of yield assessment are registered in the counts table as good). Taking into account results of these investigations, three sets of days of test milk samplings for 4-7 milkings in the analysed period were selected. In comparison with studies on the choice of test milking days by Perz and Sobek (1997), some modifications were made here. Here is the list of the selected days:

4a	–	13, 35, 60, 84	6a	–	9, 27, 41, 61, 76, 91
4b	–	13, 38, 63, 88	6b	–	9, 25, 41, 57, 73, 89
4c	–	20, 45, 70, 95	6c	–	13, 29, 45, 61, 77, 93
5a	–	10, 28, 48, 70, 90	7a	–	8, 22, 35, 51, 64, 79, 92
5a	–	10, 30, 50, 70, 90	7b	–	8, 22, 36, 50, 64, 78, 92
5c	–	16, 36, 56, 76, 96	7c	–	12,26, 40, 54, 68, 82, 96

Yields were obtained by adding up function values in consecutive days of the 100-day interval. The calculated values were compared with true yields. The assumed criterion of the quality assessment of the model was the value of the variance of errors of cow yields estimation with the assistance of the given function (1) – (22).

Each of the tested functions was fitted separately to data of individual cows and the estimated 100-day yield was calculated followed by the assessment of the relative error (in comparison with the true value). Finally, error variances for all cows within one set of days of a simulated milk test samplings were calculated. Additionally, mean variances for the same number of test milkings but for different sets of days were computed. The authors decided to take into consideration the error variance rather than the mean error because the model whose predictions are burdened with a considerable but consistent error (small variance) is better than the one in which errors exhibit a considerable scatter, because the former can be fixed by the introduction of an appropriate correction.

3. Results and discussion

Variances of the estimation errors for 100-day milk yields of cows calculated with the assistance of models represented by equations (1) – (22) as well as for the sets of sampling days of test milkings quoted in the previous section were compared. Values of error variances are presented in Figures 1 – 4. Mean variances for the three groups of the examined cows are tabulated. Numbers given in the x axis correspond to designations given to equations describing functions used in this paper. Results for the functions from the study by Sherchand et al. (1995) are given first, then, for comparative purposes, results for functions (1) – (8) estimated by Perz (1998).

The above figures do not show results obtained from Newton interpolating polynomial. Results obtained from the estimation of the milk yield of cows with the assistance of Newton interpolating polynomial for n days are consistent with the results for the $n-1$ degree polynomial which were obtained using the method of the least squares (the sum of squares of differences $\Sigma(y_{est} - y_{obs})^2$ is the smallest and equals zero if the polynomial passes through the required points; therefore, the task can be reduced to the problem of interpolation); Perz and Sobek (1997).

Table 1 presents the ranking of compared models on the basis of the mean variance of the cow's milk yield estimation error as the criterion of the model quality.

Table 1. Model rankings according to the growing variance of the estimation error

	Variance of the estimation error for 4 to 7 test milkings							
No.	Function	4 test milkings	Function	5 test milkings	Function	6 test milkings	Function	7 test milkings
1	1	7.04	1	4.30	1	4.77	1	2.50
2	11	7.14	2	4.39	11	4.99	8	2.52
3	21	7.16	8	4.42	2	5.07	2	2.55
4	8	7.38	12	4.50	3	5.14	3	2.68
5	2	7.69	21	4.58	12	5.33	11	2.71
6	12	8.21	11	4.89	8	5.43	12	2.73
7	7	8.96	7	5.05	14	5.89	21	3.13
8	3	9.02	14	5.18	4	6.91	4	3.14
9	14	17.32	3	5.25	15	7.35	14	3.19
10	15	17.51	15	6.27	7	7.47	15	3.68
11	20	23.59	4	9.38	13	7.62	7	4.38
12	17	26.33	16	16.69	21	7.98	5	5.02
13	13	27.26	20	18.21	5	11.22	13	5.42
14	19	30.82	13	22.84	16	11.79	16	6.06
15	18	87.30	19	23.13	20	19.02	20	14.40
16	22	138.33	17	23.18	19	24.83	17	20.96
17			18	48.99	17	30.27	6	26.90
18			22	126.43	18	118.98	18	57.12
19					22	123.22	22	127.38

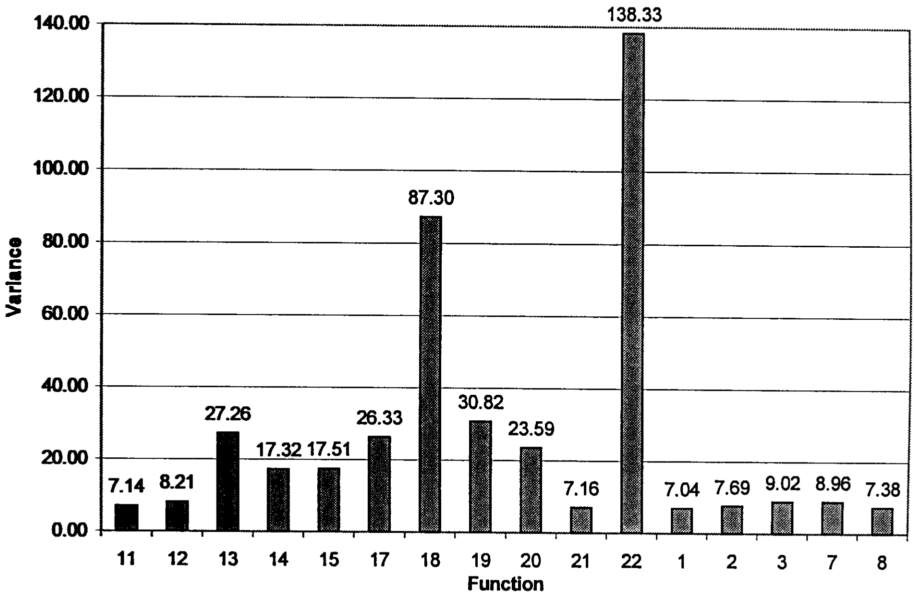

Figure 1. Milk yield estimation error variance (4 test milkings)

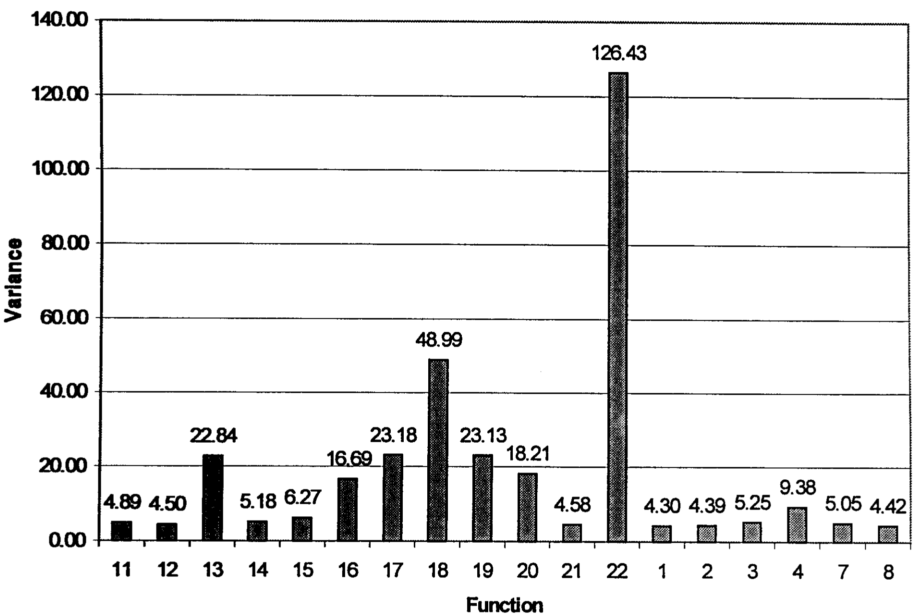

Figure 2. Milk yield estimation error variance (5 test milkings)

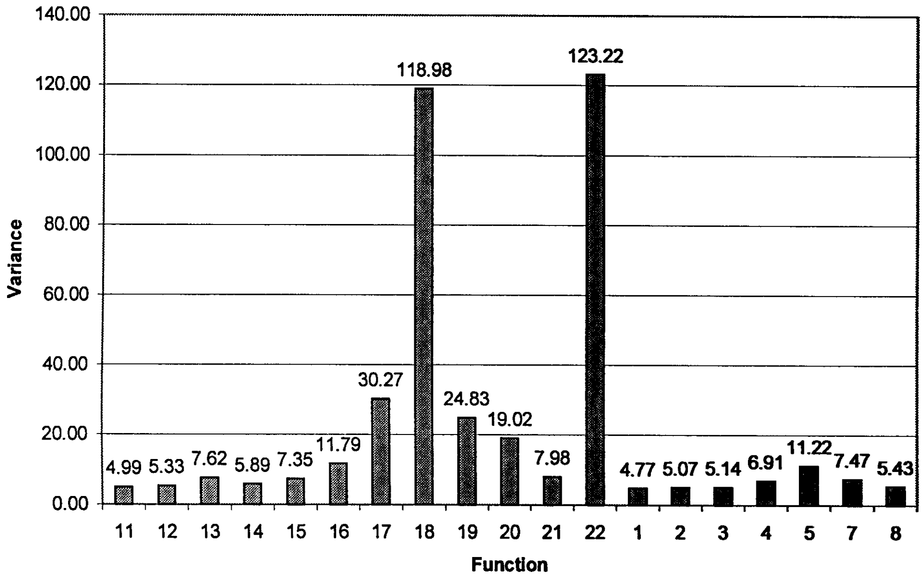

Figure 3. Milk yield estimation error variance (6 test milkings)

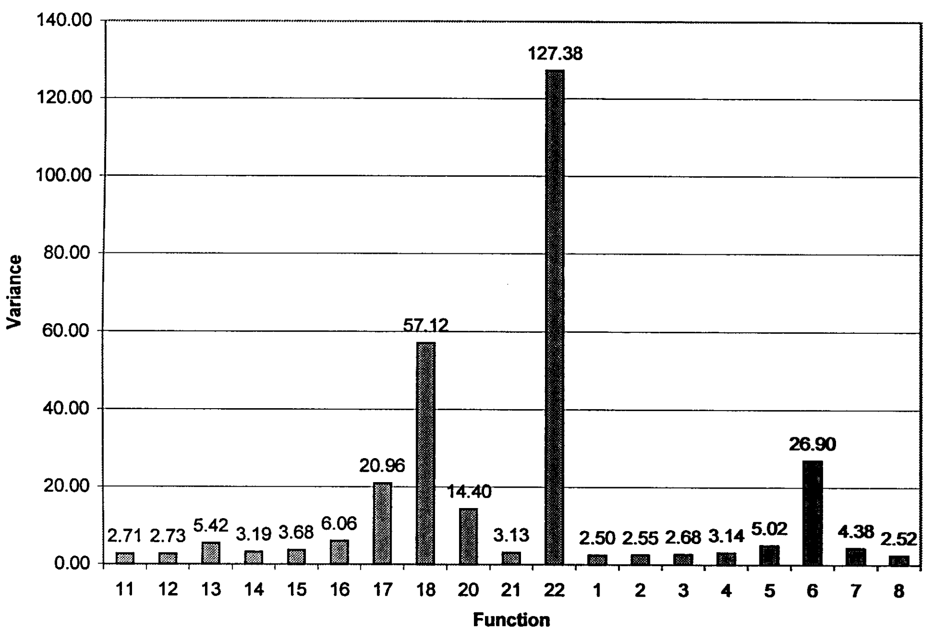

Figure 4. Milk yield estimation error variance (7 test milkings)

It can be noticed that not all the functions occur in each figure. It was not possible, using the Lisp-Stat package, to find coefficients of equation (16) for 4 milkings and of equation (19) for 7 milkings (it was impossible to select initial values in such a way as to obtain convergence of the iterative process).

The smallest error variance for 4 test milkings was found for the following functions: linear (1), (11) and (21) and exponential (8).

For 5 milkings, the concerned functions include: linear (1), square (2) and exponential (8) and (12).

For 6 milkings: linear (1), (11), square (2) and third degree (3) functions.

For 7 milkings: linear (1), exponential (8), square (2) and third degree (3) functions.

The above mentioned functions are characterised not only by a low value of the total mean of variance, but also by a small scatter of partial means for different sets of test milkings which confirms the existence of an advantageous and small susceptibility of these functions to the choice of the milking day. However, it is worth noting that there are also unfavourable coincidences of groups of cows and sets of milking days for which the variance reaches considerably higher values in comparison with others (for example, function number 18).

Function positions in the ranking alter with the change in the number of test milkings. Despite this, generally speaking, the same functions remain at the top, while slight variance differences between them can be incidental (associated, for example, with the choice of specific sets of days).

A significant improvement of results was obtained by the increase in the number of milkings to seven. On the other hand, no dramatic variance differences between the estimations from five and six milkings are observed; in fact, results from six milkings are even slightly worse than for five test milkings.

The original Wood's curve (14) (Wood, 1967) occupies the following positions in the ranking: ninth for 4 and 7 milkings, eighth – for 5 and seventh – for 6 milkings. The most complicated models have relatively worst results of estimations. This fact may be caused by an insufficient number of measurement points.

The analysis of results from Figures 1 – 4 was used to evaluate which of the investigated models are best for the examined groups of cows. It was concluded that the linear function (1) is the best model for the analysed groups followed by the square function (2) for groups I and III and exponential function (8) for group II. The third position for group I is occupied by the difference of two exponential functions (11), for group II – by the square function (2) and for group III – by the exponential function (8).

4. Conclusions

As mentioned earlier when comparing models represented by equations (1) – (9) with those discussed in papers by Sherchand et al. (1995) and Schaeffer and Jamrozik (1997a,b) [represented by equations (11) - (22)], we employed the criterion of the estimation error variance (error was computed in relation to the true 100-day milk yield of a cow). On the basis of results from Table 1, it can be concluded that the simplest models yield either better or at least compatible estimation results with those from the quoted publications.

Functions which match well the shape of the true curve courses do not always (especially at small number of test milkings) guarantee high estimation accuracy since they are susceptible to fluctuations in daily yields. Simple models appear to be more useful for the estimation of milk yield.

References

Bronsztejn J.N., Siemiendiajew K.A. (1998). *Matematyka. Poradnik encyklopedyczny.* Wyd. XV. Wydawnictwo Naukowe PWN, Warszawa.

Detilleux J.C., Grohn Y.T., Quaas R.L. (1994). Effects of clinical ketosis on test day milk yields in Finnish Ayrshire cattle. *J. Dairy Sci.* **77**, 3316-3323.

Martin F.F. (1976). *Wstęp do modelowania cyfrowego.* PWN, Warszawa.

Perz W., Sobek Z. (1997). Ocena wybranych metod statystycznych do estymacji wydajności mlecznej krów na podstawie próbnych udojów. IV Krajowa Konferencja "Komputerowe wspomaganie badań naukowych", Wrocław - Świeradów Zdrój, 75-78.

Perz W. (1998). Ocena różnych metod aproksymacyjnych do szacowania 100-dniowej wydajności mlecznej krów na podstawie próbnych udojów. Praca doktorska AR Poznań.

Schaeffer L., Jamrozik J. (1997a). Random regression models for dairy cattle test day yields. Random regression in animal breeding. Course Notes, CGIL Guelph, June 25-June 28, 1-31.

Schaeffer L., Jamrozik J. (1997b). Multiple-trait prediction of lactation yields for dairy cows. *J. Dairy Sci.* **79**, 2044-2055.

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

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

Received 15 May 1999; revised 20 September 1999

Ocena wybranych modeli matematycznych dla krzywych laktacji

STRESZCZENIE

Dokonano oceny 21 modeli krzywej laktacji pod względem dokładności szacowania za ich pomocą 100-dniowej wydajności mlecznej krów na podstawie kilku (4-7) udojów próbnych. Problem ten występuje w praktyce hodowlanej. Brano pod uwagę stabilność modelu (możliwie najniższa wariancja błędu estymacji) zamiast stosowanego najczęściej jako kryterium współczynnika determinacji. Wyniki potwierdzają leżące u podstaw przyjętego podejścia przypuszczenie, że proste modele, choć nie oddają wiernie rzeczywistego przebiegu krzywej laktacji (mniej elastycznie dopasowują się do danych), są mniej wrażliwe na pojawiające się w danych fluktuacje. Dają więc stabilniejsze wyniki szacowania, nie wymagając przy tym stosunkowo intensywnych obliczeń, koniecznych do wyznaczenia parametrów funkcji złożonych modeli.

SŁOWA KLUCZOWE: krzywa laktacji, wydajność mleczna, szacowanie wydajności mlecznej.