MODELLING ANIMAL SYSTEMS PAPER

Modelling the lactation curve of dairy cows using the differentials of growth functions

M. H. FATHI NASRI[1], J. FRANCE[2], N. E. ODONGO[3], S. LOPEZ[4],
A. BANNINK[5] AND E. KEBREAB[6*]

[1] *Department of Animal Science, Faculty of Agriculture, University of Birjand, Birjand, Iran*
[2] *Centre for Nutrition Modelling, Department of Animal and Poultry Science, University of Guelph, Guelph, ON, N1G 2W1, Canada*
[3] *Agriculture and Agri-Food Canada, Lethbridge Research Centre, Lethbridge, AB, T1J 4B1, Canada*
[4] *Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24007 León, Spain*
[5] *Animal Sciences Group, Wageningen University and Research Centre, Lelystad, The Netherlands*
[6] *Department of Animal Science, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada*

(*Revised MS received 28 May 2008*)

SUMMARY

Descriptions of entire lactations were investigated using six mathematical equations, comprising the differentials of four growth functions (logistic, Gompertz, Schumacher and Morgan) and two other equations (Wood and Dijkstra). The data contained monthly milk yield records from 70 first, 70 second and 75 third parity Iranian Holstein cows. Indicators of fit were model behaviour, statistical evaluation and biologically meaningful parameter estimates and lactation features. Analysis of variance with equation, parity and their interaction as factors and with cows as replicates was performed to compare goodness of fit of the equations. The interaction of equation and parity was not significant for any statistics, which showed that there was no tendency for one equation to fit a given parity better than other equations. Although model behaviour analysis showed better performance of growth functions than the Wood and Dijkstra equations in fitting the individual lactation curves, statistical evaluation revealed that there was no significant difference between the goodness of fit of the different equations. Evaluation of lactation features showed that the Dijkstra equation was able to estimate the initial milk yield and peak yield more accurately than the other equations. Overall evaluation of the different equations demonstrated the potential of the differentials of simple empirical growth functions used in the current study as equations for fitting monthly milk records of Holstein dairy cattle.

INTRODUCTION

Lactation curves are valuable tools for dairy producers for management decision-making and selection. Knowledge of the lactation curve allows prediction of total milk yield from a single or several test days early in lactation. With such a knowledge, a dairy producer can make management decisions early based on individual production (Gipson & Grossman 1990). There are various mathematical equations describing lactation curves in dairy cows, from the more empirical

* To whom all correspondence should be addressed.
Email: Ermias_Kebreab@umanitoba.ca

equations that relate input to output statistically with little consideration of the biology of lactation (e.g. Wood 1967; Rook *et al.* 1993), to the more mechanistic ones which describe the lactation curve based on the biology of lactation (e.g. Dijkstra *et al.* 1997; Pollott 2000). Although the latter may produce parameters that have a more biological interpretation, they can be too complex for routine use outside research, and some of the mechanisms represented are as yet not fully understood. Growth functions which have been used to describe growth in various animal (Lopez *et al.* 2000) and microbial (Lopez *et al.* 2004) species can also be used as simple empirical equations to describe cumulative yield of milk over lactation by

analogy with growth of body tissues and microbial populations. Differentials of these functions would give an expression for daily milk yield.

The objective of the current paper, therefore, was to compare the performance of standard growth functions (logistic, Gompertz, Schumacher and Morgan) in their differential form with the Wood equation (the most widely used lactation equation) and the Dijkstra equation (a simple mechanistic model) by fitting these equations to monthly milk records for an entire lactation from a commercial herd of Iranian Holstein cows.

MATERIALS AND METHODS

Data

For the study, data on 215 completed lactations by Iranian Holstein cows were obtained from a commercial herd belonging to Kenebist farm, Mashad, Iran. Parity is a factor that has been cited as affecting milk production and characteristics of the lactation curve (e.g. Pérochon et al. 1996). Therefore, the cows studied were of different parities (70 first, 70 second and 75 third). These animals were chosen as being representative of the herd and had undergone a minimum of 11 months of lactation and were free of any health disorders over this period. For uniformity of duration and to remove the effect of drying-off, only the first 11 months of lactation were considered. The effect of season was controlled by including only winter calving cows. Once-monthly observation of milk production was used for each lactation. The ingredients and composition of diets fed to the cows are presented in Table 1. Summary statistics of the lactational data by parity are shown in Table 2.

Equations

The equations used to describe the lactation curves are presented in Table 3. The incomplete gamma function proposed by Wood (1967), which has been used widely to study lactation curves, was selected as the simplest equation. Although this equation can give an acceptable fit to milk yield data from a given lactation, it tends to over-predict during early and late lactation and under-predict data during mid-lactation (Grossman & Koops 1988; Sherchand et al. 1995). The logistic, Gompertz, Schumacher and Morgan equations are growth functions (Thornley & France 2007). A typical lactation curve rises to a peak before falling away, which is the same trajectory mapped by the slope of a sigmoidal growth function. Therefore, growth functions written in their differential form and expressed as a function of time have potential application as lactation equations and are simple to use. The equation of Dijkstra et al. (1997) is the solution to a simple mechanistic model based

Table 1. *Diet ingredients (g/kg DM) and composition*

Item	
Ingredient	
Alfalfa hay	210
Corn silage	130
Soybean meal	119
Ground corn	125
Ground barley	133
Cottonseed whole	65
Cottonseed meal	25
Fish meal	45
Fat supplement[a]	14
Beet pulp	74
Wheat bran	48
Dicalcium phosphate	2·3
Sodium bicarbonate	5
Vitamin/trace mineral mix[b]	5
Nutrient	
Net energy for lactation, NE_L[c] (MJ/kg DM)	6·87
Acid detergent fibre	191
Neutral detergent fibre[d]	301
Non-fibre carbohydrates[e]	418
Ether extract	49
Crude protein	177
Ca	9
P	5·3

[a] Bergafat-F100 (palm oil), component (DM basis): 20 g glycerol/kg and 750 g free fatty acids/kg (myristic acid <35, palmitic acid 750–900, stearic acid 50–100, oleic acid 50–100 and linoleic acid <10 g/kg total fatty acids).
[b] Composition of trace minerals (g/kg DM) and vitamin mix: Ca, 196; P, 96; Mg, 19; Fe, 3; Na, 71; Cu, 0·3; Mn, 2; Zn, 3; Co, 0·1; I, 0·1; Se, 0·01; and Vit A, 500 000 IU/kg DM; Vit D, 100 000 IU/kg DM; Vit E, 100 IU/kg DM.
[c] Estimated from NRC (2001).
[d] N-free basis.
[e] Non-fibre carbohydrate = 1000 − (crude protein + neutral detergent fibre + ether extract + ash).

on a set of differential equations representing cell proliferation and cell death in the mammary gland and is a four-parameter algebraic equation. Although the latter is based on growth biology of the mammary gland and is more attractive as it is more realistic physiologically, the model with its four parameters is not necessarily simple enough for routine statistical use.

For all equations, expressions for initial yield, time to peak, maximum yield, total yield over a finite lactation and relative rate of decline (persistency) at the midway point of the declining phase were derived analytically (see Thornley & France, 2007, p. 620 *et seq.*, by way of example) and are presented in Table 4. These features of lactation were calculated for the different equations by using the related formulae, after parameter estimation. To determine whether an

Table 2. *Summary of the observed lactation yield data (kg) by parity*

Parity	Number of cows	y_{avg}	Minimum	Maximum	y_0	y_m	Y	Y_{305}
First	70	30·1	26·3	39·5	20·5	41·0	9780·1	8795·2
Second	70	36·8	28·8	44·5	21·0	44·5	9953·5	8944·9
Third	75	37·4	28·2	45·8	22·5	45·0	10181·0	9125·3

y_{avg}, average yield; y_0, initial yield (kg/day); y_m, peak yield (kg/day); Y, total yield; Y_{305}, 305-day yield.

Table 3. *Equations used to describe the lactation curve*

Equation	Functional form*
Wood	$y = at^b e^{-ct}$
Logistic	$y = \dfrac{abc(c-b)e^{-at}}{[b+(c-b)e^{-at}]^2}$
Gompertz	$y = ab \exp[b(1-e^{-ct})/c - ct]$
Schumacher	$y = \dfrac{ab^2 c}{(t+b)^2} \exp\left(\dfrac{bct}{t+b}\right)$
Morgan	$y = ab^c c \dfrac{t^{c-1}}{(t^c+b^c)^2}, \quad c>1$
Dijkstra	$y = a \exp[b(1-e^{-ct})/c - dt]$

* y is the milk yield (kg/day), t is the time from parturition (day), and a, b, c and d (all >0) are parameters that define the scale and shape of the lactation curve.

equation predicts initial, maximum and total yields accurately or if there is under/overprediction, the difference between predicted and observed (calculated from actual records) values were *t*-tested. The observed total yield was the yield that accumulated during the 11 months of lactation and was calculated by multiplying the milk yield observed with the number of days until next observation. The observed maximum yield was the maximum value of the 11 observations of each animal.

Statistical analyses

Each equation was fitted to monthly records of the 330-day lactation of each cow (total of 215 curves) using the PROC NLIN procedure in SAS (SAS Institute 1999) and the parameters were estimated. The models were tested for goodness of fit (quality of prediction) using a runs test, residual sum of squares (RSS), and calculation of mean square prediction error (MSPE), root of MSPE (RMSPE) expressed as a proportion of the observed mean (Theil 1966), adjusted multiple coefficient of determination (R^2_{adj}) and Akaike's information criterion (AIC). In a runs test, a run is a sequence of residuals with the same sign (positive or negative), and for this test, the average residual of replicate observations was used for each

day in milk. Clustering of residuals with the same sign and serial correlation may be indicative of inappropriate fitting of the model to experimental data or choice of model and a small number of runs of sign is obtained when the residuals are not randomly distributed, so residuals of the same sign tend to cluster together on some parts of the curve (Motulsky & Ransnas 1987). Mean square prediction errors were calculated as the sum of squared differences between observed and predicted values divided by the number of experimental observations. RMSPE was also calculated (square root of MSPE divided by the observations mean) so that MSPE could be expressed in the same units as the observed and predicted values. The MSPE was broken down into mean bias or error in central tendency (ECT); slope bias or error due to regression (ER); and random error or error due to disturbance (ED). Error in central tendency indicates how the average of predicted values deviates from the average of observed values. Error due to regression (regression bias) measures deviation of the least squares regression coefficient from 1, the value it would have been if the predictions were completely accurate. Error due to disturbance represents the variation in observed values unexplained after the mean and the regression biases have been removed. R^2_{adj} was calculated using the formula:

$$R^2_{adj} = 1 - [(n-1)/(n-p)](1-R^2),$$

where R^2 is the multiple coefficient of determination $[1-(RSS/TSS)]$, TSS is total sum of squares, n is the number of observations (data points) and p is the number of parameters in the equation. Note that R^2 was adjusted for the number of parameters in the equation to make a fair comparison between models. For simplicity, R^2_{adj} will be reported as R^2. AIC was calculated as using the equation (Burnham & Anderson 2002):

$$AIC = n \times \ln(RSS) + 2p$$

AIC is a good statistic for comparison of models of different complexity because it adjusts the RSS for number of parameters in the model. A smaller numerical value of AIC indicates a better fit when comparing models. All the above statistics were analysed for variance using the PROC GLM procedure

Table 4. *Expressions for the features of the lactation curve for each equation*

| Equation | Features of the lactation* | | | | |
	y_0	t_m	y_m	$r(t_h)$	Y
Wood	0	b/c	$a(b/c)^b e^{-b}$	$\dfrac{b}{t_h} - c$	$\dfrac{a}{c^{b+1}} \gamma(b+1, ct_f)$
Logistic	$ab(c-b)/c$	$\dfrac{1}{a} \ln\left(\dfrac{c-b}{b}\right)$	$ac/4$	$a\left[\dfrac{(c-b)e^{-at_h} - b}{(c-b)e^{-at_h} + b}\right]$	$b\left[\dfrac{c}{b+(c-b)e^{-at_f}} - 1\right]$
Gompertz	ab	$c^{-1}\ln(b/c)$	$ac\exp\left(\dfrac{b}{c} - 1\right)$	$b\exp(-ct_h) - c$	$a\{\exp[b(1-e^{-ct_f})/c] - 1\}$
Schumacher	ac	$b\left(\dfrac{bc}{2} - 1\right)$	$\dfrac{4a}{b^2 c} e^{bc-2}$	$\dfrac{b^2 c}{(t_h+b)^2} - \dfrac{2}{t_h+b}$	$a\left[\exp\left(\dfrac{bct_f}{t_f+b}\right) - 1\right]$
Morgan	0	$b\left(\dfrac{c-1}{c+1}\right)^{1/c}$	$\dfrac{a}{4bc}(c-1)^{(c-1)/c}(c+1)^{(c+1)/c}$	$\dfrac{(c-1)b^c - (c+1)t_h^c}{t_h(t_h^c + b^c)}$	$\dfrac{at_f^c}{t_f^c + b^c}$
Dijkstra	a	$c^{-1}\ln(b/d)$	$a(d/b)^{d/c}\exp[(b-d)/c]$	$b\exp(-ct_h) - d$	$a\displaystyle\int_0^{t_f} \exp[b(1-e^{-ct})/c - dt]dt$

* Definition of terms: y_0, initial yield (kg/day); t_m, time to peak yield (days); y_m, peak yield (kg/day); t_f, length of lactation (days); $t_h = (t_m + t_f)/2$; $r(t_h)$, relative rate of decline at the point halfway between peak yield and end of lactation; Y, total yield (kg); $\gamma(q, x)$, incomplete gamma function $= \int_0^x z^{q-1}e^{-z}dz$.

of SAS (SAS Institute 1999) using the following model:

$$Z_{ijk} = \mu + M_i + P_j + MP_{ij} + e_{ijk},$$

where Z_{ijk} represents the different statistics, μ is the overall mean, M_i is the fixed effect of lactation equation, P_j is the fixed effect of parity, MP_{ij} is the fixed interaction effect of lactation equation i with parity j, and e_{ijk} is the random effect ($i = 1$–6; $j = 1$–3; $k = 1$–70 for parities 1 and 2 but 1–75 for parity 3). The least square means (LSM) for each effect were compared using the Duncan test.

RESULTS AND DISCUSSION

In the present work, comparison of the equations was carried out according to three criteria: model behaviour when fitting the curves using nonlinear regression, statistical performance, and comparison of biologically meaningful parameters estimated by each equation.

Model behaviour

The relationship between predicted milk yield and days in milk (DIM) for first, second and third parity cows was investigated using six models as shown in Fig. 1. For comparison with predicted values, observed milk yield was plotted against DIM (Fig. 2) and the residuals were plotted against DIM to identify difficulties in prediction by a particular model (Fig. 3). When fitting curves by nonlinear regression using the six equations and utilizing the PROC NLIN procedure of SAS (SAS Institute 1999), non-convergence
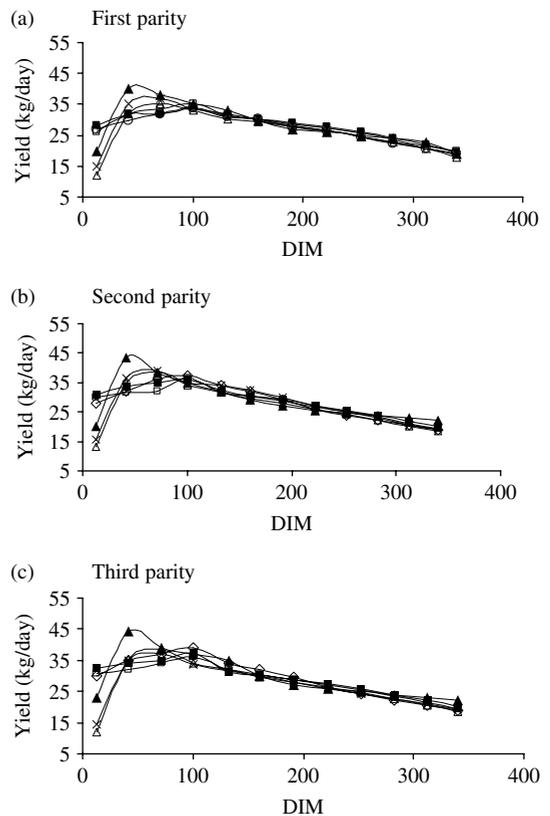

Fig. 1. Plots of predicted milk yield (kg/day) against DIM for cows of first, second and third parity using the Wood ($\triangle$), logistic ($\blacksquare$), Gompertz ($\square$), Schumacher ($\diamond$), Morgan ($\times$) and Dijkstra ($\blacktriangle$) equations.

Fig. 2. Plots of observed milk yield (kg/day) against DIM for cows of first, second and third parity.

Fig. 3. Plots of residuals against DIM for cows of first, second and third parity using Wood ($\triangle$), logistic ($\blacksquare$), Gompertz ($\square$), Schumacher ($\diamond$), Morgan ($\times$) and Dijkstra ($\blacktriangle$) equations (residuals are calculated as predicted minus observed values).
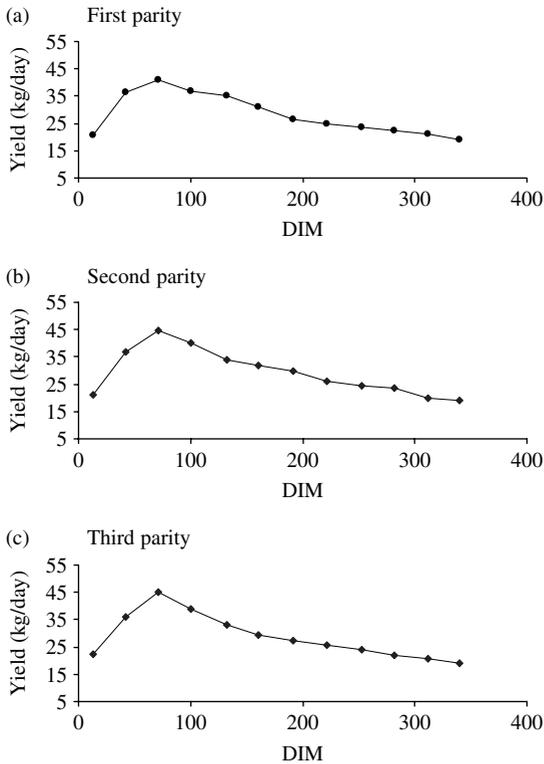
to a solution was observed with all equations, but its occurrence differed. When using growth functions, it was low (0·066, 0·08, 0·013 and 0·03 proportion of total curves for logistic, Gompertz, Schumacher and Morgan, respectively), but for the Wood and especially Dijkstra equations it was higher (0·096 and 0·305 of total curves, respectively). The range of possible starting values for parameters was selected by inspecting the final estimates for each parameter obtained with curves that were fitted without problems, after checking the uniqueness of these solutions. This method of selecting initial parameter estimates for the nonlinear modelling process was judged to be appropriate to achieve a reliable solution for most data sets. Following this approach, further sensitivity to starting values and convergence after a large number of iterations, which are considered symptomatic of an ill-conditioned (parameter values tending to 0 or to biologically unacceptable values) or inappropriate model, were observed more often for Wood and Dijkstra (0·09 and 0·11, respectively) than for the other equations. Thus, the results of model behaviour analysis showed superior ability of differentials of growth functions for fitting the lactation data used in the present study. However, it should be noted that
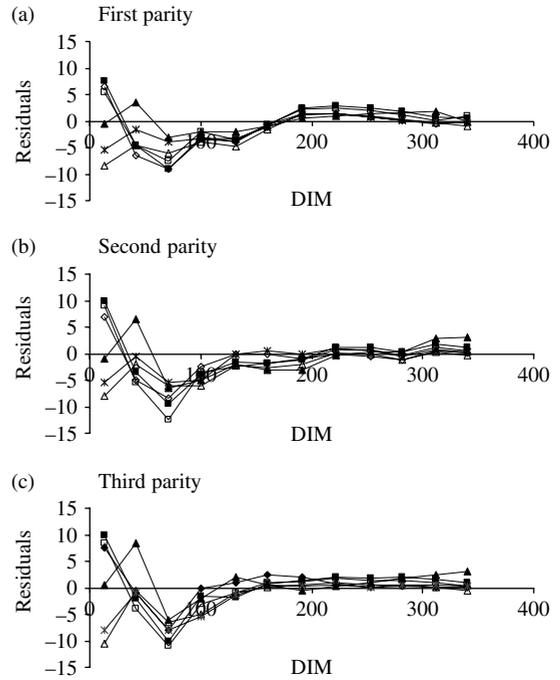
there is not a single simple method to evaluate similarities and differences between nonlinear equations and to deal with the question of which equation should be used; therefore, selection of a model to explain a particular set of data should not be based entirely on model behaviour. Statistical measures are also important (Motulsky & Ransnas 1987) and are presented in the following section.

Statistical evaluation

There are different statistical tests for ranking and evaluating models. Sometimes results from these different tests seem contradictory, so an overall assessment is needed in this situation (Fathi Nasri *et al.* 2006). The number of runs of sign of residuals, RSS, MSPE, breaking down of MSPE into ECT, ER and ED, RMSPE (as an indicator of model accuracy), R^2 and AIC, as the most widely used statistical criteria for comparing models, were calculated and the results are shown in Tables 5 and 6. Number of runs of sign of residuals from fitting the different equations (Table 5) tended to be small for the Wood, Schumacher and Morgan equations with a high percentage of curves with five or fewer runs, indicating systematic under- or over-estimation. With all the

Table 5. *Proportion of curves for each number of runs of sign of the residuals* (in the range $\leqslant 4$ to $\geqslant 8$) *observed when fitting each lactation equation* (smaller number of runs indicates systematic under- or over-fitting)

Equation	Parity	Runs of sign				
		$\leqslant 4$	5	6	7	$\geqslant 8$
Wood	1	0·42	0·42	0·17	0·00	0·00
	2	0·33	0·33	0·22	0·11	0·00
	3	0·08	0·15	0·38	0·31	0·08
Logistic	1	0·10	0·23	0·52	0·08	0·08
	2	0·00	0·20	0·30	0·40	0·10
	3	0·00	0·08	0·38	0·31	0·23
Gompertz	1	0·00	0·12	0·62	0·12	0·12
	2	0·00	0·12	0·50	0·25	0·12
	3	0·00	0·10	0·40	0·30	0·20
Schumacher	1	0·54	0·08	0·23	0·15	0·00
	2	0·47	0·20	0·27	0·07	0·00
	3	0·26	0·25	0·17	0·25	0·08
Morgan	1	0·18	0·64	0·09	0·00	0·09
	2	0·31	0·23	0·15	0·23	0·08
	3	0·27	0·27	0·27	0·00	0·20
Dijkstra	1	0·00	0·33	0·33	0·33	0·00
	2	0·00	0·14	0·43	0·29	0·14
	3	0·00	0·14	0·14	0·43	0·29

* Residuals are calculated as predicted minus observed values.

other equations, the number of runs tended to be higher as a result of random distribution of the residuals over time. The number of runs of sign tended to increase with all equations as parity increased, which means the equations were able to fit lactation records of multiparous cows better than primiparous cows. Additionally, the other statistics (apart from RSS) showed there was significant difference between parities, so the equations fitted data from multiparous cows better than primiparous cows. Differences between the characteristics of the lactation curve of primiparous and multiparous cows are likely to be responsible for the significant difference between goodness of fit of the equations for the different parities. Primiparous cows are more persistent, and their lactation curves are flatter than those of multiparous cows, sometimes with little discernible rise to a peak (Wood 1969). Scott *et al.* (1996) pointed out that lactation curves of multiparous cows are similar, except that the estimate of daily milk production is multiplied by a slightly higher factor for cows in their fourth lactation and beyond. There was no significant difference between the goodness of fit for cows of parities 2 and 3 in the present study.

After fitting each equation to raw milk yield data of cows at different parities (Fig. 1), LSM of the

statistics was calculated (Table 6). None of these statistics was able to discriminate between the goodness of fit of different equations, so despite the results of model behaviour analysis, all equations were of the same rank or quality in the fitting of the lactation data. These results are in contrast to those of Dijkstra *et al.* (1997) and Val-Arreola *et al.* (2004), who found that the Dijkstra equation fitted better than more empirical equations such as the Wood equation. One of the main differences between this and previous studies is the frequency of measurement. Both the Dijkstra *et al.* (1997) and Val-Arreola *et al.* (2004) studies had higher frequency and number of cows in the analysis, which might have helped in finding significant differences between the equations they evaluated. The results, however, are in accordance with those of Pollot (2000), who concluded (based on many dairy recording schemes which use a maximum of ten monthly test-day records per lactation) that the equations with more parameters such as Dijkstra *et al.* (1997) are over-parameterized and will not yield better fits than Wood. This is because in these situations it is common to find no recordings, or at best one recording, taken before peak yield, and this makes estimating the cell proliferation phase of the lactation and the peak yield less accurate. Both Dijkstra *et al.* (1997) and Val-Arreola *et al.* (2004) used many more observations in their studies than just once per month, so it appears that frequency of measurements determines the preferred model.

Statistical evaluation showed a significantly lower initial yield ($P < 0.001$) for the first parity than for second and third parity cows, as expected. Also, a significantly lower peak yield ($P < 0.001$), a later time to peak ($P < 0.001$) and a significantly greater persistency ($P < 0.001$) were obtained for cows in their first parity than for those in later parities, which was in agreement with Stanton *et al.* (1992) and Dijkstra *et al.* (1997). Total lactational milk yield was also significantly less ($P < 0.001$) in primiparous than in multiparous cows, as expected.

There were no significant equation × parity effects (Table 6), which shows there was no tendency for an equation to fit a particular parity better than another. This result is in contrast to Dijkstra *et al.* (1997) who pointed out that, with smooth patterns of lactation data such as primiparous records, the Wood equation will yield a better fit than the Dijkstra equation and the probable reason was that in this case the number of cells and the enzymatic activity per cell, as required parameters of their equation, cannot both be defined uniquely from lactation data owing to the timing and frequency of pre-peak observations being insufficient.

Estimates of equation parameters

The features of lactation calculated from different equations by using the formulae in Table 4 and the

Table 6. *Some features and statistics for comparing goodness of fit of the different lactations* (*standard errors are in parentheses*)

	Equation*						Parity			Equation × parity	DF†
	1	2	3	4	5	6	1	2	3		
Features‡											
y_0	0·0 (0·11)	33·3 (0·84)	31·9 (0·80)	27·3 (0·72)	0·0 (0·23)	20·7 (0·65)	15·2 (0·45)	19·5 (0·52)	19·4 (0·57)	20·5 (0·96)	184
t_m	44 (0·96)	66 (1·02)	68 (1·0)	54 (0·97)	43 (0·87)	29 (0·75)	63 (1·03)	46 (0·88)	42 (0·85)	52·4 (2·08)	170
y_m	39 (0·89)	36 (0·87)	37 (0·85)	39 (0·87)	40 (0·90)	40 (0·92)	33·6 (0·85)	39·5 (0·90)	41·8 (0·86)	39·7 (1·62)	176
$r(t_h)$	−0·0040 (0·0003)	−0·0054 (0·0004)	−0·0052 (0·0004)	−0·0037 (0·0003)	−0·0035 (0·0003)	−0·0032 (0·0003)	−0·0025 (0·0003)	−0·0044 (0·0003)	−0·0049 (0·0003)	0·0053 (0·0006)	176
$r(t_h)_{305}$	−0·0036 (0·0003)	−0·0038 (0·0003)	−0·0040 (0·0003)	−0·0039 (0·0003)	−0·0036 (0·0003)	−0·0032 (0·0003)	−0·0023 (0·0002)	−0·0040 (0·0003)	−0·0044 (0·0003)	0·0048 (0·0006)	170
Y	9420 (520·1)	10333 (590·3)	10395 (586·2)	10890 (600·1)	9678 (580·6)	10428 (598·8)	8663 (501·7)	10293 (581·5)	12331 (620·4)	10840 (1052·6)	176
Y_{305}	8916 (482·2)	8881 (480·3)	8869 (476·6)	8891 (478·3)	8950 (492·3)	9279 (502·6)	8068 (452·3)	9233 (503·4)	9323 (505·9)	9246 (982·2)	175
Statistics§											
RSS	124 (20·3)	135 (22·4)	179 (25·6)	139 (23·1)	121 (20·2)	86·7 (16·5)	166 (26·8)	105 (18·9)	119 (20·8)	150·2 (48·6)	176
MSPE	10 (1·1)	10 (1·0)	11 (0·98)	11 (1·2)	10 (1·1)	6 (0·75)	12·4 (1·6)	9·0 (0·92)	7·5 (0·81)	11·6 (2·0)	176
MSPE analysis (% MSPE)											
ECT	0·2 (0·01)	0·1 (0·01)	0·1 (0·01)	0·2 (0·01)	0·1 (0·01)	0·1 (0·01)	1·1 (0·27)	0·8 (0·21)	0·4 (0·03)	0·42 (0·07)	175
ER	1·3 (0·12)	1·2 (0·13)	1·2 (0·13)	1·4 (0·12)	1·3 (0·14)	1·2 (0·12)	2·5 (0·21)	2·2 (0·22)	1·2 (0·14)	1·9 (0·28)	175
ED	99 (10·3)	99 (10·5)	99 (10·5)	98 (10·0)	99 (10·1)	99 (10·5)	96 (9·7)	97 (9·5)	98 (9·5)	99 (18·5)	175
RMSPE	9·1 (0·65)	9·3 (0·67)	9·4 (0·64)	9·5 (0·67)	9·1 (0·67)	6·8 (0·54)	10·3 (0·71)	9·0 (0·67)	7·6 (0·56)	9·8 (1·32)	176
R^2	0·85 (0·040)	0·82 (0·042)	0·85 (0·046)	0·82 (0·041)	0·83 (0·042)	0·92 (0·051)	0·62 (0·023)	0·92 (0·055)	0·94 (0·58)	0·89 (0·096)	176
AIC	15·8 (0·27)	15·4 (0·23)	15·9 (0·29)	15·7 (0·27)	15·8 (0·31)	17·4 (0·41)	20·8 (0·56)	14·9 (0·19)	10·4 (0·12)	17·7 (0·79)	176

* Equations 1–6: Wood, logistic, Gompertz, Schumacher, Morgan and Dijkstra, respectively.
† Degrees of freedom.
‡ y_0, initial yield (kg/day); t_m, time to peak yield (days); y_m, peak yield (kg/day); $r(t_h)$ and $r(t_h)_{305}$, relative rate of decline at the point halfway between peak yield and end of lactation and between peak yield and 305 days of lactation, respectively; Y and Y_{305}, total yield and 305-day yield (kg), respectively.
§ RMSPE, square root of MSPE expressed as a percentage of the observed mean; R^2, adjusted multiple coefficient of determination; RSS, residual sum of squares; MSPE, mean square prediction error; ECT, error due to central tendency; ER, error due to deviation from regression line; ED, random error; AIC, Akaike information criteria.

Table 7. *Parameter estimates for the different lactation equations* (*standard errors are in parentheses*)

Parameter†	Equation*					
	1	2	3	4	5	6
a	20 (3·4)	0·009 (0·0013)	2368 (818·7)	454 (158·3)	16915 (1429·0)	23 (5·2)
b	0·26 (0·050)	5538 (1691·0)	0·023 (0·0033)	64 (12·4)	259 (36·1)	0·069 (0·0120)
c	0·005 (0·0007)	15054 (1140·0)	0·009 (0·0009)	0·045 (0·0155)	1·3 (0·05)	0·066 (0·0091)
d	–	–	–	–	–	0·0035 (0·00058)

* Equations 1–6: Wood, logistic, Gompertz, Schumacher, Morgan and Dijkstra, respectively.
† *a, b, c* and *d* (all >0) are parameters that define the scale and shape of the lactation curve.

parameter estimates are presented in Tables 6 and 7. Parameters *a*, *b*, *c* and *d*, which determine the scale and shape of each equation, were not compared, because their definition is different in each equation. Regarding the features of the different equations, all equations showed systematic deviation from actual milk yield in accordance with Gipson & Grossman (1989) and Vargas *et al.* (2000), especially at the beginning of lactation and peak yield. Initial milk production was forced to zero by using the Wood and Morgan equations, which is not strictly acceptable biologically and is a limitation for application of these equations, and was over-predicted by the logistic, Gompertz and Schumacher. However, the Dijkstra equation could predict it accurately. The Wood equation was one of the equations that gave lowest milk production values in early lactation, though it was expected to over-predict at this stage of lactation. All equations except Dijkstra under-predicted peak milk production (observed value taken as the highest average monthly yield and to occur at the mid-month time point), which is in accordance with the results of Scott *et al.* (1996) and Sherchand *et al.* (1995) for the Wood equation. The Dijkstra equation provided the best prediction of maximum yield, which was similar to Val-Arreola *et al.* (2004) who also found the Dijkstra equation to fit better to lactation data from Mexican dairy cows. Time to peak yield was over-predicted by the logistic, Gompertz and Schumacher equations and was under-predicted by Dijkstra, but

Wood and Morgan could predict it accurately. Despite differences between equations with respect to initial milk yield, time to peak and peak yield, estimates of total milk yield were not significantly different between equations. So, these results showed that each lactation equation (including the Wood and Dijkstra) had some disadvantages in predicting lactation curve features but all equations were able to predict total milk yield accurately, and under the conditions of the present study the tested growth functions showed potential as candidates for the lactation equation.

CONCLUSION

A mathematical function allows the lactation curve to be expressed in terms of a set of parameters that has to be estimated. Various functions have been used to study lactation in dairy cattle, with each function having advantages and disadvantages. The growth functions selected were the logistic, Gompertz, Schumacher and Morgan equations and their differentials were compared with the Dijkstra equation and a simpler equation (Wood equation). Based on criteria to measure goodness of fit, the results of this study showed that selected empirical growth functions were able to fit monthly lactation records satisfactorily. Although the Dijkstra equation had more convergence problems in fitting monthly data, the numerical tests were better for that model.

REFERENCES

BURNHAM, K. P. & ANDERSON, D. R. (2002). *Model Selection and Multimodel Inference: a Practical–theoretic Approach*, 2nd edn. Berlin, Germany: Springer-Verlag.

DIJKSTRA, J., FRANCE, J., DHANOA, M. S., MAAS, J. A., HANIGAN, M. D., ROOK, A. J. & BEEVER, D. E. (1997). A model to describe growth patterns of the mammary gland during pregnancy and lactation. *Journal of Dairy Science* **80**, 2340–2354.

FATHI NASRI, M. H., DANESH MESGARAN, M., FRANCE, J., CANT, J. P. & KEBREAB, E. (2006). Evaluation of models to describe ruminal degradation kinetics from *in situ*

ruminal incubation of whole soybeans. *Journal of Dairy Science* **89**, 3087–3095.

GIPSON, T. A. & GROSSMAN, M. (1989). Diphasic analysis of lactation curves in dairy goats. *Journal of Dairy Science* **72**, 1035–1044.

GIPSON, T. A. & GROSSMAN, M. (1990). Lactation curves in dairy goats: a review. *Small Ruminant Research* **3**, 383–396.

GROSSMAN, M. & KOOPS, W. J. (1988). Multiphasic analysis of lactation curves in dairy cattle. *Journal of Dairy Science* **71**, 1598–1608.

Lopez, S., France, J., Gerrits, W. J., Dhanoa, M. S., Humphries, D. J. & Dijkstra, J. (2000). A generalised Michaelis–Menten equation for the analysis of growth. *Journal of Animal Science* **78**, 1816–1828.

Lopez, S., Prieto, M., Dijkstra, J., Dhanoa, M. S. & France, J. (2004). Statistical evaluation of mathematical models for microbial growth. *International Journal of Food Microbiology* **96**, 289–300.

Motulsky, H. J. & Ransnas, L. A. (1987). Fitting curves to data using nonlinear regression: a practical and non-mathematical review. *FASEB Journal* **1**, 365–374.

Pérochon, L., Coulon, J. B. & Lescourret, F. (1996). Modelling lactation curves of dairy cows with emphasis on individual variability. *Animal Science* **63**, 189–200.

Pollott, G. E. (2000). A biological approach to lactation curve analysis for milk yield. *Journal of Dairy Science* **83**, 2448–2458.

Rook, A. J., France, J. & Dhanoa, M. S. (1993). On the mathematical description of lactation curves. *Journal of Agricultural Science, Cambridge* **121**, 97–102.

SAS Institute (1999). *User's Guide: Statistics, 1999. Version 8.2.* Cary, NC, USA: SAS Institute Inc.

Scott, T. A., Yandell, B., Zepeda, L., Shaver, R. D. & Smith, T. R. (1996). Use of lactation curves for analysis of milk production data. *Journal of Dairy Science* **79**, 1885–1894.

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. *Journal of Dairy Science* **78**, 2507–2513.

Stanton, T. L., Jones, L. R., Everett, R. W. & Kachman, S. D. (1992). Estimating milk, fat and protein lactation curves with a test day model. *Journal of Dairy Science* **75**, 1691–1699.

Theil, H. (1966). *Applied Economic Forecasting.* Amsterdam, The Netherlands: North-Holland Publishing Co.

Thornley, J. H. M. & France, J. (2007). *Mathematical Models in Agriculture*, 2nd edn. Wallingford, UK: CABI Publishing.

Val-Arreola, D., Kebreab, E., Dijkstra, J. & France, J. (2004). Study of the lactation curve in dairy cattle on farms in central Mexico. *Journal of Dairy Science* **87**, 3789–3799.

Vargas, B., Koops, W. J., Herrero, M. & Van Arendonk, J. A. M. (2000). Modeling extended lactations of dairy cows. *Journal of Dairy Science* **83**, 1371–1380.

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

Wood, P. D. P. (1969). Factors affecting the shape of the lactation curve in cattle. *Animal Production* **11**, 307–316.