Introducing a sinusoidal equation to describe lactation curves for cumulative milk yield and composition in Holstein cows

Navid Ghavi Hossein-Zadeh[1], Hassan Darmani Kuhi[1], James France[2] and Secundino López[3]

[1]Department of Animal Science, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran; [2]Department of Animal Biosciences, Centre for Nutrition Modelling, University of Guelph, Guelph ON, N1G 2W1, Canada and [3]Instituto de Ganadería de Montaña (IGM), CSIC-Universidad de León, Departamento de Producción Animal, Universidad de León, E-24007 León, Spain

Abstract

The aim of the work reported here was to investigate the appropriateness of a sinusoidal function by applying it to model the cumulative lactation curves for milk yield and composition in primiparous Holstein cows, and to compare it with three conventional growth models (linear, Richards and Morgan). Data used in this study were 911 144 test-day records for milk, fat and protein yields, which were recorded on 834 dairy herds from 2000 to 2011 by the Animal Breeding Centre and Promotion of Animal Products of Iran. Each function was fitted to the test-day production records using appropriate procedures in SAS (PROC REG for the linear model and PROC NLIN for the Richards, Morgan and sinusoidal equations) and the parameters were estimated. The models were tested for goodness of fit using adjusted coefficient of determination (R^2_{adj}), root mean square error (RMSE), Akaike's information criterion (AIC) and the Bayesian information criterion (BIC). R^2_{adj} values were generally high (>0.999), implying suitable fits to the data, and showed little differences among the models for cumulative yields. The sinusoidal equation provided the lowest values of RMSE, AIC and BIC, and therefore the best fit to the lactation curve for cumulative milk, fat and protein yields. The linear model gave the poorest fit to the cumulative lactation curve for all production traits. The current results show that classical growth functions can be fitted accurately to cumulative lactation curves for production traits, but the new sinusoidal equation introduced herein, by providing best goodness of fit, can be considered a useful alternative to conventional models in dairy research.

A major application of mathematical equations in animal science is to describe milk yield pattern over the lactation in livestock farmed for milk production (Pulina and Nudda, 2001; Thornley and France, 2007). Mathematical equations of the lactation curve provide a useful tool for basic studies intended to extend scientific understanding of the complicated physiological processes that control milk production and secretion (Dimauro *et al*., 2007). Lactation curves may be utilized by different researchers in animal science to imitate the process of lactation and to assess associations between energy and nutrient provision, the endocrine system, mammary secretory cells and environmental variables influencing the process of milk production.

Diverse growth equations may be used to model the lactation curve because the shape of the curve for cumulative milk yield is similar to that for standard growth. The cumulative milk yield curve would be influenced to a lesser degree than the typical lactation curve if there are some missing records, and it is less responsive to the presence of outliers (López *et al*., 2015). Contrary to the conventional curve, fitting the lactation curve for cumulative milk yield would only be slightly influenced by the incidence of deviations due to any disturbance. Therefore, it would be possible to describe atypical or nonstandard lactation curves when daily milk yield is fitted to days in milk (DIM). Incidence of disturbance results in conventional lactation models being unable to fit atypical curves (Rekik and Ben Gara, 2004; Macciotta *et al*., 2005; Silvestre *et al*., 2005; López *et al*., 2015). Standard growth models are suitable for fitting the cumulative milk production curve because of the monotonically increasing trajectory and can represent the underlying trend of the lactation curve in relation to the productive potential of the cow.

The broad range of mathematical models available in the literature can be considered as other possibilities in the sense that each equation fits some data sets better than others (Aziz *et al*., 2006). Currently the suitability of a sinusoidal function, as a newly applied equation, is confirmed for describing growth curves in dairy cows (Darmani *et al*., 2019a) and turkeys (Darmani *et al*., 2019b). However, there are no reports of fitting a sinusoidal function to

Table 1. Parameter estimates for the different functions

Function	Parameter	Milk production	Fat production	Protein production
Linear	k	31.64	1.041	0.964
	y_0	123.6	−2.929	−5.661
Richards	c	0.0021	0.0014	0.0016
	n	−0.8167	−0.8728	−0.8301
	y_0	4.7×10^{-16}	5.3×10^{-11}	2.0×10^{-12}
	y_f	23 004.8	1037.8	871.71
Morgan	n	1.210	1.115	1.178
	K	386.4	1341.6	972.4
	y_0	6.6×10^{-16}	8.4×10^{-16}	1.6×10^{-16}
	y_f	21 721.5	1933.9	1402.9
Sinusoidal[a]	a	10 592.0	564.59	316.02
	b	2017.4	3334.0	2010.5
	θ	6.00	6.16	5.89
	$y^\star$	2768.6	65.9	125.8
	y_f	13 360.6	630.5	441.8

[a]y_f for the sinusoidal is not an estimated parameter. It was calculated from the fitted equations as: $y_f = y^\star + a$.

the lactation curve of dairy cows. Therefore, the aim of the present work was to investigate the appropriateness of a sinusoidal function by applying such an equation to the lactation curve of Holstein cows for cumulative milk production and composition using data constructed from test-day individual cow milking records. Fitting performance is compared with three conventional growth models, viz. linear, Richards and Morgan.

Materials and methods

This study used 911 144 test-day records for milk yield, milk fat and protein production of 156 906 primiparous Iranian Holstein cows from 834 dairy herds. The dataset spanned 2000 through 2011 and is part of the data kept by the Animal Breeding Centre and Promotion of Animal Products of Iran. A detailed description of the data was reported in a previous study (Ghavi Hossein-Zadeh, 2017).

For daily milk fat and protein yields an initial calculation was made using the amount of milk produced daily and the milk composition:

Daily milk fat or protein (kg/d)

$$= \frac{\text{Daily milk fat or protein (\%)} \times \text{milk yield (kg/d)}}{100}$$

All test-day records available in the dataset for a given DIM were averaged, and then the cumulative production of milk and of milk components were calculated. Each daily milk, fat and protein yield at a given DIM is added to the sum of all the yields from calving to that particular DIM to calculate cumulative production.

In general, cumulative production shows a curvilinear pattern similar in shape to that of a typical growth curve. The curve for cumulative milk yield resembles a more linear trend (more attenuated sigmoidicity) than the standard growth curve.

The functions used to describe the lactation curves are presented in Supplementary Table S1. The linear, Richards, Morgan and sinusoidal equations (the latter three give a flexible point of inflexion) were fitted to the data to model the relationship between milk, fat or protein yield and DIM. The use of a classical growth function such as the Gompertz or logistic, both of which have a fixed point of inflexion, would force daily milk yield to peak when a fixed fraction of maximum cumulative yield (1/e for the Gompertz, 1/2 for the logistic) is attained. This limitation can be avoided using a flexible growth function.

Each function was fitted using appropriate SAS procedures (PROC REG for the linear model and PROC NLIN for the Richards, Morgan and sinusoidal equations) (SAS Institute, 2002) to estimate parameters. The NLIN procedure produces weighted or unweighted least squares estimates of the parameters of a non-linear model. To fit non-linear functions, the Gauss-Newton method was used as the iteration technique. To begin this process, the NLIN procedure first examines the starting value specifications of the parameters. If a grid of values is specified, the NLIN procedure evaluates the residual sum of squares at each combination of parameter values to determine the set of parameter values producing the lowest residual sum of squares. These values are used for the initial step of the iteration. The models were examined for goodness-of-fit (quality of prediction) using adjusted coefficient of determination (R^2_{adj}), residual standard deviation or root mean square error (RMSE), Akaike's information criterion (AIC) and the Bayesian information criterion (BIC).

RMSE is a type of generalized standard deviation and was calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\text{RSS}}{n - p - 1}}$$

where RSS is the residual sum of squares, n is the number of observations (data points) and p is the number of parameters in the model. The best model is the one with the lowest RMSE.

Table 2. Comparison of goodness-of-fit, predictive and bias performance for the linear and non-linear equations

Trait	Statistics	Function			
		Linear	Richards	Morgan	Sinusoidal
Milk yield	R^2_{adj}	0.99929	0.99997	0.99997	0.99998
	RMSE	73.59	15.73	16.09	12.28
	AIC	2589.7	1662.7	1676.6	1513.6
	BIC	2597.1	1677.6	1691.5	1528.4
	MSPE	5380	244.1	255.6	148.7
	PSB	0.01769	0.00080	0.00084	0.00049
Fat yield	R^2_{adj}	0.99983	0.99995	0.99997	1.0000
	RMSE	1.1687	0.6384	0.5285	0.0767
	AIC	95.82	−266.19	−379.91	−1542.23
	BIC	103.24	−251.37	−365.08	−1527.40
	MSPE	1.357	0.402	0.276	0.0058
	PSB	0.00408	0.00121	0.00083	0.00002
Protein yield	R^2_{adj}	0.99985	0.99998	0.99999	0.99999
	RMSE	1.0415	0.3722	0.3261	0.2044
	AIC	26.46	−591.36	−670.36	−951.86
	BIC	33.87	−576.54	−655.54	−937.03
	MSPE	1.077	0.137	0.105	0.041
	PSB	0.00389	0.00049	0.00038	0.00015

R^2_{adj}, adjusted coefficient of determination; RMSE, root mean square error; AIC, Akaike information criterion; BIC, Bayesian information criterion; MSPE, mean square prediction error; PSB, percentage squared bias.

The AIC was calculated using the following equation:

$$\text{AIC} = n \times \ln\left(\frac{\text{RSS}}{n}\right) + 2p$$

The AIC is a good statistic for comparing 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.

The BIC combines maximum likelihood (data fitting) and choice of model by penalizing the (log) maximum likelihood with a term related to model complexity as follows:

$$\text{BIC} = n \times \ln\left(\frac{\text{RSS}}{n}\right) + p \ln(n)$$

A smaller numerical value of BIC indicates a better fit when comparing models.

The percentage squared bias (PSB) calculated as proposed by Ali and Schaeffer (1987) was used to assess model accuracy, and mean square prediction error (Fahti Nasri *et al.*, 2008) to evaluate the predictive performance of the equations.

Results

Estimated parameters of the equations for cumulative milk yield and composition of Holstein cows are shown in Table 1. Also, goodness of fit statistics for the four models fitted to the cumulative curves are presented in Table 2. R^2_{adj} values were generally high, implying suitable fits to the data, and showed minor differences among the models for cumulative yields. A comparison between the equations according to the other statistical criteria (Table 2) showed differences between the models under consideration. The sinusoidal equation provided the lowest values of RMSE, AIC and BIC, and therefore the best fit to the lactation curve for cumulative milk, fat and protein yields. The linear model gave the worst fit to the cumulative lactation curve for all production traits by providing the largest values of RMSE, AIC and BIC. The sinusoidal equation also performed better than the other functions in terms of prediction precision and accuracy, showing the lowest mean square prediction error and PSB (Table 2).

Both time and cumulative yield at inflexion point for different non-linear models fitted to cumulative yields of milk, fat and protein are shown in Table 3. Based on a comparison between observed cumulative curves and their predicted counterparts (Fig. 1, and online Supplementary Figs S2 and S3), only the Richards equation provided accurate estimation of the cumulative milk yield at inflexion time. The sinusoidal equation underestimated the cumulative yield at inflexion and the Morgan equation overestimated it. For cumulative fat yield, the Richards equation provided nearly accurate estimates of cumulative yield at inflexion and sinusoidal and Morgan equations overestimated it. For cumulative protein yield, the Richards underestimated cumulative yield at inflexion and the sinusoidal and Morgan overestimated it.

Observed cumulative yields from 5 to 305 DIM are depicted in Supplementary Figure S1. As demonstrated, there was an

Table 3. Time (t^*) and cumulative yield (y^*) at inflexion point, daily milk yield at peak (m_p) and milk yield at 305 d lactation (y_{305}) for the models fitted to cumulative yields of milk, fat and protein in Holstein cows (NA = not applicable)

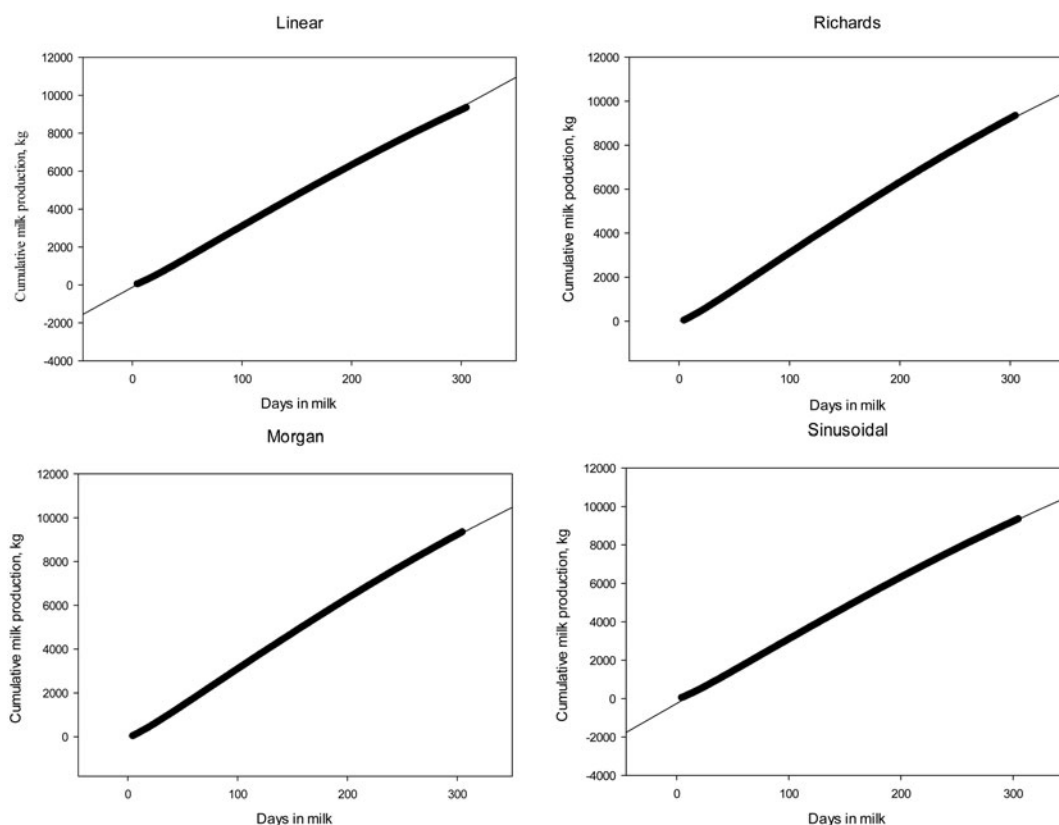| Trait	Statistics	Function			
		Linear	Richards	Morgan	Sinusoidal
Milk yield	t^* (d)	NA	96.4	95.3	91.3
	y^* (kg)	NA	2881	2891	2769
	m_p (kg/d)	31.64	33.01	33.50	32.99
	y_{305} (kg)	9526	9198	9314	9309
Fat yield	t^* (d)	NA	97.2	98.9	66.5
	y^* (kg)	NA	97.7	100.0	65.9
	y_{305} (kg)	314.5	309.2	311.0	311.2
Protein yield	t^* (d)	NA	116.4	115.7	136.0
	y^* (kg)	NA	103.0	105.8	125.8
	y_{305} (kg)	288.5	277.0	285.3	285.0

Fig. 1. Predicted lactation curves for cumulative milk yield obtained with different growth equations in Holstein cows.

increasing trend for cumulative milk yield and composition with increasing DIM. Predicted lactation curves for cumulative milk yield obtained by fitting the four growth equations are shown in Figure 1. Also, predicted lactation curves for cumulative fat yield and protein yield are shown in Supplementary Figures S2 and S3.

Discussion

Knowledge of the shape of the lactation curve is valuable in a management context, especially for decisions that are time-dependent. Modelers seek to find parametric descriptors of the shape of the lactation curve to predict characteristics including peak milk yield, time to peak, and persistency. For example,

knowing when peak milk yield will occur can assist dairy farmers or managers in planning feeding strategies to maintain peak yield for as long as possible (López *et al.*, 2015). Although several studies have compared non-linear models to fit the lactation curve for milk yield and composition, there are few reports on modeling the lactation curve for cumulative milk yield (Ghavi Hossein-Zadeh, 2014, 2017; López *et al.*, 2015) and no reports for cumulative milk composition traits. Therefore, the current study is the first to report on fitting cumulative milk fat and protein yield data, in particular assessing a sinusoidal function as an alternative to conventional models. Developing an appropriate strategy to reach a desired lactation shape through modifying the parameters of such non-linear models would be of considerable value. These parameters may act as surrogates for various genetic and environmental factors, different combinations of which can account for differences in non-linear curves.

Typically, the conventional lactation curve follows a skewed bell shape, with a quick rise after parturition to a peak a few weeks later followed by a slow decline until the end of lactation. However, several researchers have reported non-standard patterns such as continuously increasing or decreasing and reverse standard. These atypical curves are not simply fitted using conventional lactation models (Rekik and Ben Gara, 2004; Macciotta *et al.*, 2005; Silvestre *et al.*, 2005; López *et al.*, 2015). However, the consistently increasing pattern of cumulative curves does not present any difficulties in fitting sigmoidal functions to cumulative milk yield and composition traits. The current results indicate that representative non-linear growth functions, especially the sinusoidal and the Morgan equation, provide suitable fits to cumulative lactation curves not only for milk yield but also for fat and protein yields. Consistent with the results of this study, López *et al.* (2015) fitted six classical growth functions (monomolecular, Schumacher, Gompertz, logistic, Richards, and Morgan) to cumulative milk production curves of Canadian Holstein dairy cows and concluded that the fitted classical growth functions could be an alternative to conventional models when analyzing lactation data.

A cumulative lactation curve shows a monotonically increasing trend presenting a typical sigmoidal shape, with an accelerating phase related to early lactation when daily production rises, and a decreasing phase in which the slope declines continuously. Total milk, fat and protein yield over the lactation are simply obtained from this cumulative milk production curve by calculating its value at any specific DIM for each non-linear function (López *et al.*, 2015). This simple method of obtaining total production with cumulative curves differs from the method of calculating by integrating standard lactation curves to reach total milk yield. The results of the current study show that monthly test-day production records provide a suitable input for the cumulative lactation curve. This would be a benefit if a daily system of recording production data in dairy farms is not available.

Differences between the lactation curve characteristics of dairy cows are possibly responsible for the significant differences between goodness of fit of the functions. The variation in model fit in different studies on conventional curves may have resulted from the differences in mathematical form of the equations, differences in monthly production records, the breed of animals, the number of data points and the time interval between subsequent recordings of data. Compared with conventional lactation models which are sensitive to missing test-day records, especially in early DIM (Adediran *et al.*, 2007; Wasike *et al.*, 2011), fitting the non-linear growth models to cumulative

production records would be influenced to a much lesser degree by a few missing records (López *et al.*, 2015). Therefore, the smoother trend of cumulative curves for milk yield and composition is likely less susceptible to abnormal records or occurrence of outliers. Abnormal records can be a result of several problems such as measurement faults and biological disruptions in animal performance occurring when dairy cows are threatened by some severe condition which limits expression of their genetic potential, e.g. nutritional deficit, metabolic or infectious disorders (Codrea *et al.*, 2011). The identification of these bias factors and correcting them is critical and this may be achieved by analyzing daily records of production traits. However, as mentioned above, these disruptions in the lactation curve would be very problematic when working with conventional lactation models, whereas the fit of cumulative curves for milk yield and composition demonstrates the greater flexibility of these curves to varied data attributes (López *et al.*, 2015).

The current results show that growth functions can be fitted accurately to lactation curves for cumulative milk, fat and protein yields, resulting in acceptable statistical output and accurate determination of production traits. However, the sinusoidal equation introduced herein had sufficient flexibility and a variable point of inflexion and provided the best goodness of fit, so it may be considered an alternative to classical growth functions. Choosing a flexible and accurate equation to predict the lactation curve can noticeably improve the outcome of a research study and the predictive accuracy of a model guarantees its implications.

References

Adediran SA, Malau-Aduli AEO, Roche JR and Donaghy DJ (2007) Using lactation curves as a tool for breeding, nutrition and health management decisions in pasture-based dairy systems. In Fulkerson B (ed.) *Current Topics in Dairy Production*, Proceedings of the Dairy Research Foundation Symposium. vol. 12, Camden, New South Wales, Australia: The University of Sydney, pp. 74–78.

Ali TE and Schaeffer LR (1987) Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science* **67**, 637–644.

Aziz MA, Shalaby NA, El-Shafie OM, Mahdy AT and Nishida A (2006) Comparison between the shapes of lactation curve of Egyptian buffalo milk yield estimated by the incomplete gamma function and a new model. *Livestock Research for Rural Developments* **18**, Article 59. Available at http://www.lrrd.org/lrrd18/5/aziz18059.htm.

Codrea MC, Hojsgaard S and Friggens NC (2011) Differential smoothing of time-series measurements to identify disturbances in performance and quantify animal response characteristics: an example using milk yield profiles in dairy cows. *Journal of Animal Science* **89**, 3089–3098.

Darmani Kuhi H, Ghavi Hossein-Zadeh N, López S, Falahi S and France J (2019*a*) Sinusoidal function to describe the growth curve of dairy heifers. *Animal Production Science* **59**, 1039–1047.

Darmani Kuhi H, López S, France J, Mohit A, Shabanpour A, Ghavi Hossein-Zadeh N and Falahi S (2019*b*) A sinusoidal equation as an alternative to classical growth functions to describe growth profiles in turkeys. *Acta Scientiarum. Animal Sciences* **41**, e45990.

Dimauro C, Cappio-Borlino A, Macciotta NPP and Pulina G (2007) Use of a computer-aided design to develop a stress simulation model for lactating dairy sheep. *Livestock Production Science* **106**, 200–209.

Fahti Nasri N, France J, Odongo NE, López S, Bannink A and Kebreab E (2008) Modelling the lactation curve of dairy cows using the differentials of growth functions. *Journal of Agricultural Science* **146**, 633–641.

Ghavi Hossein-Zadeh N (2014) Comparison of non-linear models to describe the lactation curves of milk yield and composition in Iranian Holsteins. *Journal of Agricultural Science* **152**, 309–324.

Ghavi Hossein-Zadeh N (2017) Application of growth models to describe the lactation curves for test-day milk production in Holstein cows. *Journal of Applied Animal Research* **45**, 145–151.

López S, France J, Odongo NE, McBride RA, Kebreab E, AlZahal O, McBride BW and Dijkstra J (2015) On the analysis of Canadian Holstein dairy cow lactation curves using standard growth functions. *Journal of Dairy Science* **98**, 2701–2712.

Macciotta NPP, Vicario D and Cappio-Borlino A (2005) Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. *Journal of Dairy Science* **88**, 1178–1191.

Pulina G and Nudda A (2001) La produzione del latte. In Pulina G (ed.), *L'alimentazione degli ovini da latte*. Bologna, Italy: Avenue Media, pp. 9–31.

Rekik B and Ben Gara A (2004) Factors affecting the occurrence of atypical lactations for Holstein-Friesian cows. *Livestock Production Science* **87**, 245–250.

SAS (2002) SAS User's Guide v. 9.1: Statistics. SAS Institute, Inc., Cary, NC.

Silvestre AM, Petim-Batista F and Colaço J (2005) Genetic parameter estimates of Portuguese dairy cows for milk, fat and protein using a spline test-day model. *Journal of Dairy Science* **88**, 1225–1230.

Thornley JHM and France J (2007) *Mathematical Models in Agriculture: Quantitative Methods for the Plant, Animal and Ecological Sciences*, 2nd Edn. Wallingford, UK: CAB International.

Wasike CB, Kahi AK and Peters KJ (2011) Modelling of lactation curves of dairy cows based on monthly test day milk yield records under inconsistent milk recording scenarios. *Animal: An International Journal of Animal Bioscience* **5**, 1780–1790.