

Fitting milk production curves through nonlinear mixed models

Monica Piccardi¹, Raúl Macchiavelli², Ariel Capitaine Funes³, Gabriel A Bó^{4,5} and Mónica Balzarini^{1,6*}

¹ Cátedra de Estadística y Biometría de la Facultad de Ciencias Agropecuarias de la Universidad Nacional de Córdoba

² Universidad de Puerto Rico, Mayagüez

³ DAIRYTECH S.R.L., Santa Fe, Argentina

⁴ Instituto de Reproducción Animal Córdoba (IRAC), Paraje Pozo del Tigre, Zona Rural Est. General Paz, (5145) Córdoba, Argentina

⁵ Instituto de Ciencias Básicas y Aplicadas, Carrera de Medicina Veterinaria, Universidad Nacional de Villa María, Villa del Rosario, Córdoba, Argentina

⁶ Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Received 9 March 2016; accepted for publication 11 January 2017; first published online 28 March 2017

The aim of this work was to fit and compare three non-linear models (Wood, Milkbot and diphasic) to model lactation curves from two approaches: with and without cow random effect. Knowing the behaviour of lactation curves is critical for decision-making in a dairy farm. Knowledge of the model of milk production progress along each lactation is necessary not only at the mean population level (dairy farm), but also at individual level (cow-lactation). The fits were made in a group of high production and reproduction dairy farms; in first and third lactations in cool seasons. A total of 2167 complete lactations were involved, of which 984 were first-lactations and the remaining ones, third lactations (19 382 milk yield tests). PROC NL MIXED in SAS was used to make the fits and estimate the model parameters. The diphasic model resulted to be computationally complex and barely practical. Regarding the classical Wood and MilkBot models, although the information criteria suggest the selection of MilkBot, the differences in the estimation of production indicators did not show a significant improvement. The Wood model was found to be a good option for fitting the expected value of lactation curves. Furthermore, the three models fitted better when the subject (cow) random effect was considered, which is related to magnitude of production. The random effect improved the predictive potential of the models, but it did not have a significant effect on the production indicators derived from the lactation curves, such as milk yield and days in milk to peak.

Keywords: lactation curves, random effect, estimation, comparison criteria.

Knowledge of the behaviour of lactation curves is crucial for decision-making in commercial dairies (Macciotta et al. 2004, 2005). This is important not only at the population or dairy level, but also at the individual level (Dekkers et al., 1998; Vargas et al. 2000). Moreover, with models having a mechanistic interpretation it is possible to make inferences about management and physiology from parameter values through the construct of the model (Ehrlich, 2013). Ehrlich also demonstrates that lactation curve shape varies significantly between herds, suggesting that fitted parameter values could be used as a metric for monitoring herd performance, and points out that if lactation models are used to predict future daily milk production for incomplete

lactations, residuals between predicted and actual daily production can be used to quantify the response to an intervention. The estimation of lactation curves provides parameters that are critical in determining the reproductive cost per animal (De Vries, 2006). In the bio-economic model for dairy farms proposed by Cabrera & Giordano (2010), some of the inputs needed when deriving reproductive cost per animal under a given dairy production system are the parameters derived from fitting lactation curves.

Several mathematical functions have been proposed for fitting the values of daily yields according to days in milk. Those functions may be linear or nonlinear in the parameters and may use a different number of constants for a better fitting of the observed data. The incomplete gamma function proposed by Wood (1967), with three parameters, is one of the most popular and valid nonlinear models to date for describing lactation curves. Other nonlinear

*For correspondence; e-mail: mbalzari@gmail.com

models, such as the diphasic model (Grossman & Koops, 1988) and the MilkBot model proposed by Ehrlich (2011) have been widely used for fitting lactation curves. In both of these models, the number of parameters to be estimated is larger than in the Wood model. The addition of extra parameters aims at capturing certain movements of lactation curves, some of which are of genetic and others of environmental nature (Macciotta et al. 2011). A simplified 3-parameter version of the MilkBot model has been suggested for use where data do not include daily milk weights covering the first weeks of lactation. This would match the parameter count of the Woods model with little decrease in precision expected (Ehrlich, 2013). All three models are a set of functions used as a standard for modelling normal milk production and comparing changes in productivity over time in different dairy production systems. The widely used derived parameters from a lactation curve are peak yield, days in milk (DIM) to peak yield and 305-d yield.

Mostert et al. (2003) classified the productive performance of commercial dairies in two groups based on milk production: a high production group (above the national average milk yield) and a low production group, and showed how this partition affected the shape of the lactation curve. They found that 305-d yield was higher for high production dairies during the whole lactation. In addition, high production dairies showed steeper lactation curves and curves of low production dairies were flatter. As a consequence, the resulting persistence was higher in lower production dairies than in higher production ones. These results were also reported in other studies, such as in Olori & Galesloot (1999) for Holstein cows in Ireland, for the Dutch goat population (Galesloot, 2000) and for Holsteins in Hungary (Galesloot, personal communication).

Catillo et al. (2002) used lactation curves to show that calving season affects milk production. According to Tekerli et al. (2000), peak and milk yields were higher for cows that calved in fall and winter. Moreover, others studies have concluded that primiparous cows produce less milk than multiparous cows in early lactation, with lower peak yields, but are more persistent, with 305-d yields being significantly lower (Rao & Sundaresan, 1979; Singh & Shukla, 1985; Keown et al. 1986; Tekerli et al. 2000; Rekik et al. 2003; Silvestre et al. 2009). Therefore, patterns of lactation curves depend not only on time or days in milk, but also on several uncontrolled genetic and environmental factors (Nicolò et al. 2004; Dijkstra et al. 2010). Ehrlich (2013) suggested that significant differences in parameters between herds imply uncontrolled genetic and environmental factors. This fact produces high variability among lactation curves which cannot be appropriately modelled by means of classical fixed effects models with a unique random term (error term). Macciotta et al. (2005) cited the great variation that characterises parameter values estimated on individual lactation curves. Mixed models (models with fixed and random effects) (West et al. 2007) allow handling extra variability associated with factors that could affect each lactation curve differently.

In numerous works that study the fitting of lactation curves, the models are estimated assuming not only the existence of only one variance component, but also the independence of the data from consecutive productions. However, it should be noted that milk yield test records may be serially correlated because they are measurements taken over time on the same individual (longitudinal data). The correlation between those records may be high as a result not only of the animal's genetics but also of the environment such as the short-term effect of, for example, feeding (Ali & Schaeffer, 1987; Carvalheira et al. 1998; Ammon & Spilke, 2005; Macciotta et al. 2008). It can be expected that measurements that are closer together in time are more correlated than those that are more distant (positive autocorrelation). Moreover, measurements recorded on the same cow might be more correlated than those recorded on different animals because they share a common contribution from the same individual (Macciotta et al. 2011). Lack of modelling of these autocorrelation patterns in the dataset can lead to erroneous conclusions, especially when evaluating treatment differences (Nicolò et al. 2004). This difficulty can be also addressed using mixed models by modelling the covariance structure of the error terms in the model or, alternatively, by incorporating random effects associated with one (or more) of the parameters of the mean structure of the model. Therefore, mixed models are a powerful tool for analysing lactation curve data. Advances in the development of a statistical method for the treatment of normal longitudinal data (West et al. 2007; Fitzmaurice et al. 2008) have been scarcely explored in applications of milk production (Quintero et al. 2007). The nonlinear characteristic of the lactation curve models makes the mixed model estimation computationally harder than estimation of linear functions. Using contemporary statistical modelling techniques, such as nonlinear mixed models, can help improve population average estimation of lactation curves and predictions made for each cow.

The aim of this work was to compare the fit of lactation curves achieved from three nonlinear models (Wood, MilkBot, and diphasic models) using two statistical approaches: a nonlinear fixed-effect model and a nonlinear mixed model composed of the same lactation curve models carrying an additional random effect at the cow level.

Materials & methods

Data

We worked with a database that included 19 382 records obtained from monthly test data conducted for one year (2008), for a total of 2167 complete lactations, all longer than 149 d. In this dataset, there is one lactation per cow. 984 were first parity lactations and the rest were third parity lactations. We used separately these parity lactations to fit the alternative models because those two groups showed two very different shapes of lactation curve (Figs. 1 & 2). We did not work with the second parity lactation as Dematawewa et al. (2007), because as is known

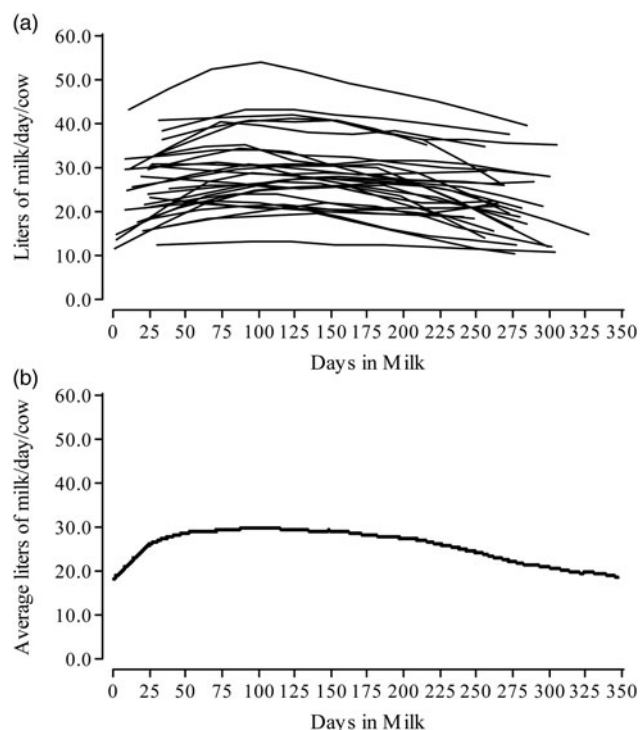


Fig. 1. Observed lactation curves from 30 randomly selected lactations (a) and raw average lactation curve observed for all lactations ($n = 984$) (b) of first lactation cows.

have an intermediate behaviour and was not of our interest. Furthermore, we worked with 2008 data because they were highly depurated. The dairies involved are located in the centre and south of Santa Fe and Córdoba provinces in Argentina. The lactations selected were associated with dairies with high reproductive performance (average annual pregnancy $\geq 16\%$) and production efficiency (total liters produced per lactation $\geq 7 \cdot 200$ l). The lactations included in this study were those that started between March and August (fall + winter), which are considered to be the cool seasons.

Statistical analyses

All three non-linear functions, Wood, MilkBot, and diphasic, were fitted to each of the recorded lactations as fixed effect models.

The model of the incomplete gamma function (Wood, 1967) (1) was

$$Y(t) = at^b \exp(-ct) \quad (1)$$

where $Y(t)$ is milk yield on day t of lactation, a is a parameter representing the production at the beginning of lactation; b and c are parameters associated with the increase and decrease of the slopes of the lactation curve, respectively. Typical lactation curves have positive b and c values. Once the fixed parameters were estimated, peak yield

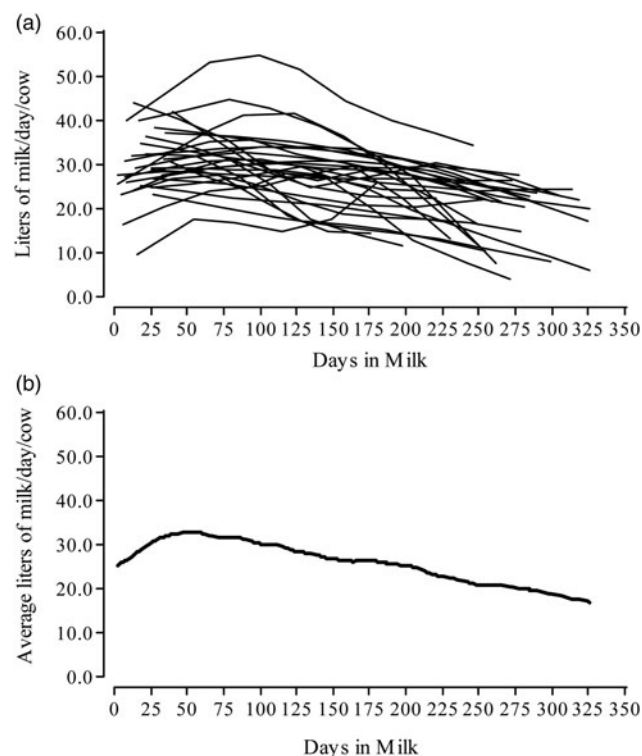


Fig. 2. Observed lactation curves from 30 randomly selected lactations (a) and raw average lactation curve observed for all lactations ($n = 1183$) (b) for third lactation cows.

(Y_{\max}) (2) and days in milk at the peak yield DIM_{peak} (3) were estimated using the following formulas:

$$Y_{\max} = a \left(\frac{b}{c} \right)^b \exp^{-b} \quad (2)$$

$$\text{DIM}_{\text{peak}} = \frac{b}{c} \quad (3)$$

The MilkBot function (Ehrlich, 2011) (4) for each lactation for milk yield on day t of lactation Y (t):

$$Y(t) = a \left(1 - \frac{\exp^{(c-t)/b}}{2} \right) \exp^{-dt} \quad (4)$$

This function consists of four parameters: parameter a called *scale*, which is a multiplier that determines the total amount of milk yield and is expressed in l/d, it must be a positive number; parameter b , *ramp*, controls the rate of increase in milk production in early lactation and is expressed in days; parameter c , *offset*, represents the offset in time between calving and the time when the highest rate of increase in milk production occurs, and is expressed in days; finally, parameter d , or *decay*, controls the loss of productive capacity and is expressed as day^{-1} . With these parameters, peak yield (Y_{\max}) (5) and days in milk at peak DIM_{peak} (6) were estimated using the following formulas:

$$Y_{\max} = a_1 \left(1 - \frac{bd}{1 + bd} \right) \exp^{-d(c - b \log(2bd/(1+bd)))} \quad (5)$$

$$\text{DIM}_{\text{peak}} = -b \log\left(\frac{2db}{db+1}\right) + c \quad (6)$$

The third function adjusted in this work was the diphasic function (Grossman & Koops, 1988) (7), which is based on the sum of two logistic functions, defined as:

$$Y(t) = a_1 b_1 [1 - \tanh^2(b_1(t - c_1))] + a_2 b_2 [1 - \tanh^2(b_2(t - c_2))] \quad (7)$$

where $Y(t)$ is milk yield of day t (t are the days in milk; DIM), \tanh is the hyperbolic tangent, for the first phase, a_1 and a_2 are half the asymptote of the total production (L) in phase one and two, respectively; b_1 and b_2 are the production rate relative to a_1 (d^{-1}) or a_2 of phase one and two, respectively. Finally, c_1 and c_2 are the peak time (d) of phase one and two, respectively. Liters associated with peak production of phase i were calculated as the product of parameters a and b of a same phase.

After fitting the three nonlinear models as fixed effects models, we fit the same models as mixed models with an extra random component at the parameter related with the magnitude of milk production. Therefore, for the Wood and MilkBot model was the parameter a , and for the diphasic model was choose the parameter a_1 . This random component is assumed normally distributed with zero mean and variance σ_a^2 , which should be interpreted as an unobservable variable that represents a random deviation of the coefficient a of the i th lactation from the population parameter. Such random deviation is assumed to be independent of the error term. Fixed and Mixed Nonlinear Models were fitted with PROC NLMIXED in SAS (SAS, 2008).

Comparison criteria

The significance of the variance component associated with the additional random effect was evaluated by comparing, for the three functions (Wood, Milkbot, and diphasic), the mixed model with the analogous fixed effects model, using a likelihood ratio test (LRT). The LRT statistic was calculated by computing the maximum log-likelihood of the reference model (with a higher number of parameters, the mixed model) and the maximum log-likelihood of the simpler model (Fixed effects model). Twice the difference of log-likelihoods is usually compared with a chi-square distribution with degrees of freedom equal to the difference between the numbers of parameters estimated by the two models, except when comparison includes a mixed model with an extra random effect and a fixed model (Molenberghs & Verbeke, 2007). In this case, which was our particular case, a chi-square distribution with 0.5 degrees of freedom is used to evaluate the approximate significance of the log-likelihood differences (Molenberghs & Verbeke, 2007). This test was performed with restricted maximum likelihood (REML) estimators (Searle et al. 1992). Then, if the test is significant, the correct model is the most complex (mixed model); otherwise, the reduced model (fixed model) is the appropriate one.

The Akaike Information Criterion (AIC; Sakamoto et al. 1986) and Bayesian Information Criterion (BIC; Schwarz, 1978) were also used as criteria for model selection. These indices can be used to rank different models and choose the one of lowest value.

The Durbin-Watson (DW) test (Durbin, 1970) was used to evaluate the hypothesis of independence between the errors of each fitted model. This test assumes that the errors are normally distributed with zero mean and equal variance. The null hypothesis for this test is that the errors are independent, against the alternative hypothesis that the errors are correlated with a first-order autoregressive structure. The test statistic is $d = 2(1 - r)$, where r is the sample autocorrelation of the model residuals. If the errors are independent, d will be close to 2. By contrast, if the errors are strongly autocorrelated, d will be far from 2.

Finally, to test the predictive value of any of the models we used the cross-validation method. We randomly chose 75% of the data to be in a calibration set, and the remainder 25% of the data was the validation set. The parameters were estimated from the calibration set using all the studies models and validated on the validation set. The models were fitted to individual monthly test data yields within each parity class and parameters were estimated by nonlinear regression using the NLMIXED procedure. The root mean square error (RMSE) was used for comparing models within fixed and random models; thus, those with the lowest values are the ones of best fit (Kobayashi & Salam, 2000). In order to compare two models, we used the ratio between RMSE's, selecting the model with the lowest RMSE as referent. This ratio was used as measurement of the relative efficient (RE) of a given model with regard to the reference model.

Results & discussion

Observed lactation curves

In the first lactation, peak yield (27 l) occurred between 78–86 DIM. In the third lactation, peak yield (33 l) occurred between 57–64 DIM (Tables 1 & 2). Consequently, in agreement with other authors (Rekik et al. 2003; Macciotta et al. 2004; Dematawewa et al. 2007; Silvestre et al. 2009; Cole et al. 2011), third lactation cows (mature cows) reached peak yield earlier and with higher milk yield than first-lactation cows. The rate at which production declines after reaching the peak was also higher in third lactation than in first lactation cows.

Lactation curves observed from 30 randomly selected lactations and average lactation curve observed for all first lactations ($n = 984$) are presented in Fig. 1. Lactation curves observed in 30 randomly selected lactations and average lactation curve observed for all lactations ($n = 1183$) for third lactation cows are presented in Fig. 2. In both figures, the variability between animals of the same group becomes apparent.

Table 1. Parameter estimates and goodness of fit criteria for three models of lactation curves, as fixed effects model and mixed model with random effect associated at the parameter a for the Wood and MilkBot model and at the parameter a_1 for the Diphasic[†] model. First lactation cows calved in the cool season

Parameter	Fixed effects models			Mixed models		
	Wood	Milkbot	Diphasic [†]	Wood	Milkbot	Diphasic
a	13·689 (0·210) [‡]	32·684 (0·242)	$a_1 = 2502·780$ (0·593) $a_2 = 5700·670$ (0·111)	12·947 (0·132)	32·530 (0·198)	$a_1 = 2053·930$ (76·929) $a_2 = 4002·510$ (13·429)
b	0·199 (0·004)	34·390 (1·664)	$b_1 = 0·006$ (1E-4) $b_2 = 0·003$ (1E-5)	0·212 (0·002)	33·850 (0·968)	$b_1 = 0·006$ na $b_2 = 0·004$ (1E-5)
c	-0·002 (3·5E-5)	-0·686 (0·667)	$c_1 = 79·491$ (2·974) $c_2 = 259·050$ (9·948)	-0·002 (2·2E-5)	0·974 (0·357)	$c_1 = 105·060$ (6·259) $c_2 = 228·540$ (0·510)
d		0·001 (3E-5)			0·001 (1·8E-5)	
S_u^2				2·832 (1·028)	11·145 (1·036)	1408·105 (1·025)
Milk yield [§]	27·090 (0·058)	27·562 (0·066)	16·095 (0·482) 18·247 (0·312)	26·951 (0·106)	27·438 (0·108)	16·980 (0·083) 16·135 (0·604)
DIM [¶]	83·184 (0·851)	78·767 (1·484)	79·491 (2·975) 259·05 (9·948)	86·232 (0·496)	79·908 (0·901)	105·060 (6·259) 228·540 (0·510)
Model fit criteria						
RE	1	1·003	1·008	1	1·001	1·025
AIC	199 813	199 787	62 584	178 664	178 684	56 445
BIC	199 846	199 829	62 644	178 694	178 720	56 484
DW	0·590	0·598	0·452	1·557	1·579	1·517
-2 log L.	199 805 [1]	199 777 [2]	62 581 [3]	178 654 [4]	178 672 [5]	56 429 [6]
LRT				[1 vs. 4]: 21 151 $P < 0·0001$	[2 vs. 5]: 21 105 $P < 0·0001$	[3 vs. 6]: 6·152 $P < 0·0001$

RE, Relative efficient; AIC, Akaike information criterion; BIC, Bayesian information criterion; DW, Durbin-Watson statistic; -2log L., -2 log(likelihood); LRT, likelihood ratio test; na, not available

[†]For the diphasic model, each parameter was estimated for the two phases of the function separately, a : a_1 and a_2 ; b : b_1 and b_2 ; c : c_1 and c_2

[‡]Standard error between parentheses

[§]Milk yield in the Wood model = $a(b/c)b \times e^{-b}$; in the MilkBot = $(a) \times ((1 - e^{-(c - (c - b \times \text{Log}((2 \times b \times d)/(1 + b \times d))))/b/2}) \times e^{-d \times (c - b \times \text{Log}((2 \times b \times d)/(1 + b \times d)))))$; and in the Diphasic model = ab

[¶]Days in milk to Peak in the Wood model = b/c ; in the MilkBot model = $b \times (\text{Log}((2 \times d \times b)/(d \times b + 1))) + c$; and in the Diphasic model = c_1 and c_2 for each phase.

Table 2. Parameter estimates and goodness of fit criteria for three models of lactation curves as fixed effects model and mixed model with random effect associated with parameter a for the Wood and MilkBot model and at the parameter a_1 for the Diphasic[†] model. Third lactating cows calved in the cool season

Parameter	Fixed effects models			Mixed models		
	Wood	Milkbot	Diphasic ²	Wood	Milkbot	Diphasic
a	17.511 (0.231) [‡]	43.986 (0.380)	$a_1 = 2.511,540$ (0.063) $a_2 = 5.703,100$ (0.353)	16.666 (0.166)	45.864 (0.369)	$a_1 = 2.502.110$ na $a_2 = 5.699.140$ na
b	0.212 (0.004)	36.228 (1.725)	$b_1 = 0.006$ (2E-4) $b_2 = 0.003$ (5E-5)	0.227 (0.003)	42.759 (1.449)	$b_1 = 0.006$ na $b_2 = 0.003$ na
c	-0.004 (3.6E-5)	-0.653 (0.532)	$c_1 = 58.961$ (3.412) $c_2 = 159.500$ (10.794)	-0.004 (2.5E-5)	0.505 (0.373)	$c_1 = 75.668$ na $c_2 = 119.280$ na
d		0.003 (3.7E-4)			0.003 (3E-5)	
S_u^2				3.158 (1.016)	8.680 (1015)	1334.084 na
Milk yield [§]	33.443 (0.067)	33.836 (0.088)	16.052 (0.548) 18.030 (0.297)	33.489 (0.105)	33.733 (0.105)	15.828 (0.072) 16.797 (1E-5)
DIM to peak [¶]	57.267 (0.569)	61.430 (0.887)	58.704 (3.427) 159.37 (10.652)	58.822 (0.385)	64.831 (0.588)	75.668 na 119.280 na
Model fit criteria						
RE	1.004	1	1.015	1	1.016	1.056
AIC	230.501	230.606	73.537	215.174	215.442	68.052
BIC	230.535	230.648	73.588	215.204	215.478	68.092
DW	0.745	0.758	0.657	1.463	1.482	1.520
-2 log L.	230.493 [1]	230.596 [2]	73.523 [3]	215.164 [4]	215.430 [5]	68.036 [6]
LRT				[1 vs. 4]: 21.151 $P < 0.0001$	[2 vs. 5]: 21.105 $P < 0.0001$	[3 vs. 6]: 5487 $P < 0.0001$

RE, Relative efficient; AIC, Akaike information criterion; BIC, Bayesian information criterion; DW, Durbin-Watson statistic; -2log L., -2 log(likelihood); LRT, likelihood ratio test; na, not available

[†]For the diphasic model, each parameter was estimated for the two phases of the function separately, a : a_1 and a_2 ; b : b_1 and b_2 ; c : c_1 and c_2

[‡]Standard error between parentheses

[§]Milk yield in the Wood model = $a(b/c)b \times e^{-b}$; in the MilkBot = $(a) \times ((1 - e((c - b \times \text{Log}((2 \times b \times d)/(1 + b \times d))))/b)/2) \times e(-d \times (c - b \times \text{Log}((2 \times b \times d)/(1 + b \times d))))$; and in the Diphasic model = ab

[¶]Days in milk to Peak in the Wood model = b/c ; in the MilkBot model = $b \times (\text{Log}((2 \times d \times b)/(d \times b + 1))) + c$; and in the Diphasic model = c_1 and c_2 for each phase

Fitted lactation curves

Estimates of the model parameters for fitting the lactation curves for the Wood (1967), MilkBot (Ehrlich, 2011), and diphasic (Grossman & Koops, 1988) models for first lactation calving are shown in Table 1. According to the goodness of fit criteria used, AIC, BIC, and LRT, the diphasic model had the best fit for first lactations. However, it had the worst relative efficiency (RE) values. In terms of goodness of fit of the observed data, a low RMSE might be also a consequence of an over-parameterised model (Grossman & Koops, 1988; 2003; Dematawewa et al. 2007). Over-parameterised models fit the available data well, but may fail to make good predictions with new data. Over-parameterisation also produces estimation problems because of parameter correlations (singularity or near singularity in the Hessian matrix), and standard errors of estimation could become not available (Tables 1 & 2; Dematawewa et al. 2007).

For the Wood and MilkBot classical models (fixed effects), while majority of the information criteria suggest selecting MilkBot, there are no differences between the RE criteria. Therefore, following the principle of parsimony, selecting Wood model with three parameters would be a good option. According to Dematawewa et al. (2007), the Wood model compared with other models not included in this study, presented the lowest RMSE values. Since Wood and MilkBot model use the same exponential decay function, they would be expected to function very similarly in the extended lactations which were the focus of that study.

The comparison of each of the fixed effects models with their respective fits with mixed models with an additional random effect showed an improvement of fit according to AIC and BIC. Both approaches (fixed and mixed models) exhibit a statistically significant difference according to the LRT. Moreover, as shown by the DW test, the inclusion of the random effect at the cow level reduced the correlation between repeated test day records on the same lactations and, in this case, on the same cow. DW statistic values increased, approaching the expected value of 2 for white noise (uncorrelated data), and even the predictive ability of the models improved significantly under the mixed model framework (Dematawewa et al. 2007).

The estimate of the variance of the random effect was relatively high in the MilkBot and diphasic models, highlighting the high variability between individual lactation curves and the need to adjust estimates/forecasts to the different characteristics of different lactations, even within a single group. This was not reflected in the Wood model. However, according to the comparison criteria, predictions are improved by adding a random effect to the Wood model.

The additional random effect produced a better fit from a statistical standpoint, increasing the predictive potential (Calegario & Matri, 2005), but did not affect the estimates of dairy parameters derived from lactation curves, such as DIM to peak and peak yield. The derived parameter that was most sensitive to statistical modelling was DIM to

peak, especially in the MilkBot model. However, using MilkBot or Wood models generated greater differences than the addition or not of a random effect to model correlations between test day records. Nonetheless, these random terms are timely because as concluded by Lindstrom & Bates (1990), incorporating extra random deviations gives flexibility to the model, eliminating the need to adjust different functional forms for individual curves of the same population. It should be noted that the interpretation of the derived parameters in the fixed and mixed models differ slightly: while in the fixed model they represent averages through the population of dairies considered, in the mixed model they represent averages for 'typical' cows, i.e., cows whose random effects are average.

Estimates of the parameters for fits made to lactation curves with the Wood (1967), MilkBot (Ehrlich, 2011) and diphasic (Grossman & Koops, 1988) models for third lactation cows calved in cool seasons are shown in Table 2. Although the parameter values are higher in third lactation cows, except for DIM to peak, the same, but more pronounced, behaviour of residuals is observed when comparing with first lactations cows at the model level. A correct fit of diphasic models with random cow effects was not accomplished because of correlated parameter estimates.

Moreover, curves for lactations that started in hot seasons (September to February) were also fitted and the effect of heat stress was reflected in the indicator peak yield in both lactations studied (first and third lactation), resulting in a reduction of between 2 to 3 l at peak yield. The indicator DIM to peak was also affected, with a reduction of 2 and 15 DIM until peak yield occurs in the first and in the third lactations, respectively (data not shown). The effect of heat stress on milk production was previously described (Tekerli et al., 2000; Catillo et al. 2002; Rekik et al. 2003; Silvestre et al. 2009). At the model level, similar behaviours to those observed in both lactations started in the cool season were found (data not shown).

Conclusions

Fitted lactation curves are useful to estimate dairy productive indicators. The three nonlinear models for lactation curves had a better fit when an additional random effect was included to model the variability between lactations and data correlation from successive milk yield tests. However, the inclusion of random effects did not significantly affect production parameters derived from the curve, such as milk yield or days in milk to peak. In the dairy farms studied, peak yield occurred at about 80 DIM, with an average peak production of 27 l. We recommend adding a random cow effect when the fitted lactation curve is used for making predictions. The diphasic model was computationally complex because of its over-parameterisation, and therefore it proved to be impractical for fitting lactation curves such as those in the pasture-based production systems. Moreover, since the differences in

goodness of fit between Wood and MilkBot models were not important, the most parsimonious model (Wood) was selected to derive the productive indicators.

We thank the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) for financial support.

References

- Ali TE & Schaeffer LR** 1987 Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science* **67** 637–644
- Ammon C & Spilke J** 2005 Comparison of fixed- and random-regression models using different functional approaches of lactation curves for milk yield forecasts EFITA/WCCA. In *Joint Congress on IT in Agriculture*, Vila Real, Portugal, pp. 630–635
- Cabrera VE & Giordano J** 2010 Economic decision making for reproduction. In *2010 Dairy Cattle Reproduction Conference*, St Paul, MN
- Calegario N & Mastri R** 2005 Estimativa do crescimento de povoalimentos de Eucalyptus baseada na teoria dos modelos não lineares em multinível del efecto misto. [Stands of Eucalyptus growth estimate based on the theory of nonlinear models in multilevel of the mixed effect]. *Ciencia Forestal, Santa María* **15** 285–292
- Carvalho JGV, Blake RW, Pollak EJ, Quaas RL & Duran-Castro CV** 1998 Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in United States Holstein herds. *Journal of Dairy Science* **81** 2738–2751
- Catillo G, Macciotta NPP, Carretta A & Cappio-Borlino A** 2002 Effects of age and calving season on lactation curves of milk production traits in Italian water buffaloes. *Journal of Dairy Science* **85** 1298–1306
- Cole JB, De Vries A & Null DJ** 2011 Short communication: best prediction of 305-day lactation yields with regional and seasonal effects. *Journal of Dairy Science* **94** 1601–1604
- Dekkers JCM, en Hag JHT & Weersink A** 1998 Economic aspects of persistency of lactation in dairy cattle. *Livestock Production Science* **53** 237–252
- Dematawewa CMB, Pearson RE & VanRaden PM** 2007 Modeling extended lactations of Holsteins. *Journal of Dairy Science* **90** 3924–3936
- De Vries A** 2006 Economic value of pregnancy in dairy cattle. *Journal of Dairy Science* **89** 3876–3885
- Dijkstra J, Lopez S, Bannink A, Dhanoa MS, Kebreab E, Odongo NE, Fathi Nasri MH, Behera UK, Hernandez Ferrer D & France J** 2010 Evaluation of a mechanistic lactation model using cow, goat and sheep data. *Journal of Agriculture Science* **148** 249–262
- Durbin J** 1970 An alternative to the bounds test for testing for serial correlation in least-squares regression. *Econometrica* **38** 422–429
- Ehrlich JL** 2011 Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. *The Bovine Practitioner* **45** 88–95
- Ehrlich JL** 2013 Quantifying inter-group variability in lactation curve shape and magnitude with the MilkBot® lactation model. *PeerJ* **1** e54 <http://dx.doi.org/10.7717/peerj.54>
- Fitzmaurice G, Davidian M, Verbeke G & Molenberghs G** 2008 *Longitudinal Data Analysis. Chapman and Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press: New York
- Galesloot JB** 2000 Development of lactation curves and projection factors for 305-day yields for the Dutch goat population as a base for management indicators. In *Proc. 32nd Biennial Session Int. Comm. Anim. Recording (ICAR)*, Bled, Slovenia
- Grossman M & Koops WJ** 1988 Multiphasic analysis of lactation curves in dairy cattle. *Journal of Dairy Science* **71** 1598–1608
- Grossman M & Koops WJ** 2003 Modeling extended lactation curves of dairy cattle: a biological basis for the multiphasic approach. *Journal of Dairy Science* **86** 988–998
- Keown JF, Everett RW, Empet NB & Wadell LH** 1986 Lactation curves. *Journal of Dairy Science* **69** 769–781
- Kobayashi K & Salam MU** 2000 Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal* **92** 345–352
- Lindstrom MJ & Bates DM** 1990 Nonlinear mixed effects models for repeated measures data. *Biometrics* **46** 673–687
- Macciotta NP, Vicario D, Di Mauro C & Cappio-Borlino A** 2004 A multivariate approach to modeling shapes of individual lactation curves in cattle. *Journal of Dairy Science* **87** 1092–1098
- Macciotta NPP, Vicario D & 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
- Macciotta NPP, Mele M, Conte G, Serra A, Cassandro M, Dal Zotto R, Cappio-Borlino A, Pagnacco G & Secchiari P** 2008 Association between a polymorphism at the stearoyl CoA desaturase locus and milk production traits in Italian Holsteins. *Journal of Dairy Science* **91** 3184–3189
- Macciotta NPP, Dimauro C, Rattu S, Steri R & Pulina G** 2011 The mathematical description of lactation curves in dairy cattle. *Italian Journal of Animal Science* **11** 213–223
- Molenberghs G & Verbeke G** 2007 *Likelihood Ratio, Score, and Wald Test in a constrained 10 Parameter Space*. Taylor & Francis, Ltd. on behalf of the American Statistical Association. *The American Statistician* **61**, No. 1
- Mostert BE, Theron HE & Kanfer FHJ** 2003 Derivation of standard lactation curves for South African dairy cows. *South African Journal of Animal Science* **33** 70–77
- Nicolò PP, Macciotta NPP, Cappio-Borlino A & Pulina G** 2004 *Growth and Lactation Curves. Chapter 6. Genetic Analysis of Complex Traits Using SAS by Arnold Saxton*. Cary, NC, USA: SAS Institute Inc., ISBN 1-59047-507-0
- Olori VE & Galesloot JB** 1999 Projection of records and calculation of 305-day yields for dairy cattle in the Republic of Ireland. IP/99-2174/PG/NK
- Quintero Vélez J, Serna Gallo J & Cerón-Muñoz M** 2007 Modelos mixtos no lineales en curvas de lactancia de búfalas en un sistema de producción orgánica en el Magdalena Medio Antioqueño (Colombia). *Livestock Research for Rural Development* **19** Article #52. Retrieved March 10, 2017, from <http://www.lrrd.org/lrrd19/4/quin19052.htm>
- Rao MK & Sundaresan D** 1979 Influence of environment and heredity on the shape of lactation curves in Sahiwal cows. *Journal Agricultural Science. Cambridge* **92** 393–401
- Rekik B, Gara AB, Hamouda MB & Hammami H** 2003 Fitting lactation curves of dairy cattle in different types of herds in Tunisia. *Livestock Production Science* **83** 309–315
- Sakamoto Y, Ishiguro M, Kitagawa G** 1986 *Akaike Information Criterion Statistics*. Dordrecht: Reidel
- SAS Institute** 2008 *SAS/STAT Software for Windows 9.2*. Cary, NC: SAS Institute Inc
- Schwarz G** 1978 Estimating the dimension of a model. *Annals of Statistics* **6** 461–464
- Searle SR, Casella G & McCulloch CE** 1992 *Variance Components*. New York: John Wiley and Sons
- Silvestre AM, Martins AM, Santos VA, Ginja MM & Colaço JA** 2009 Lactation curves for milk, fat and protein in dairy cows: a full approach. *Livestock Science* **122** 308–313
- Singh J & Shukla KP** 1985 Factors effecting persistency of milk production in Gir cattle. *Indian Veterinary Journal* **62** 888–894
- Tekerli M, Akinci Z, Dogan I & Ackan A** 2000 Factors affecting the shape of lactation curves of Holstein cows from the Balikesir province of Turkey. *Journal Dairy Science* **83** 1381–1386
- Vargas B, Koops WJ, Herrero M & Van Arendonk JAM** 2000 Modeling extended lactations of dairy cows. *Journal Dairy Science* **83** 1371–1380
- West BT, Welch KB & Gatecki AT** 2007 *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton, USA: Chapman and Hall/CRC
- Wood PDP** 1967 Algebraic model of the lactation curve in cattle. *Nature (London)* **216** 164–165