



Modelling of individual lactation curves of Tunisian Holstein-Friesian cows for milk yield, fat, and protein contents using parametric, orthogonal and spline models

M. Bouallegue^{1,5}, R. Steri², N. M'hamdi³ and M. Ben Hamouda⁴

¹ High Agronomic Institute of Chott Mariem, BP. 47, Chott Meriem Sousse, Tunisia

² Centro di Ricerca per la Produzione delle Carni e il Miglioramento Genetico, Rome, Italy

³ National Institute of Agronomy of Tunis, 43 Av. Ch. Nicole, 1082 Tunis Mahragène, Tunisia

⁴ Institution of Agricultural Research and Higher Education, Ministry of Agriculture, 30, Rue Alain Savary, Tunisia

KEY WORDS: Tunisian Holstein-Friesian, lactation curve, test day, milk yield, mathematical models

Received: 17 September 2013

Revised: 17 February 2015

Accepted: 20 March 2015

⁵ Corresponding author: e-mail: mahdibouallegue@yahoo.fr

ABSTRACT. Different equations were employed to directly model individual lactation curves of Tunisian Holstein-Friesian cows using 260 241 test day records for milk yield, fat, and protein contents. Eleven mathematical models were compared. Parametric curves (Legendre polynomials, Ali and Schaeffer (AS), Wood, and Wilmink models) and regression splines were tested. Goodness of fit was assessed by considering the adjusted *R*-square ranked according to five classes, statistical criteria, and residuals analysis. Regression splines (quadratic and cubic spline models with three knots) showed better fitting performances and greater flexibility for all milk traits. The sixth-order Legendre orthogonal polynomial model and the Ali and Schaeffer model also gave the best fit for milk traits compared with the three-parameter models (Wood and Wilmink) and with the lower-order polynomial models, but the AS model was less correlated for fat content ($R = 0.87$), gave a moderate correlation ($R = 0.90$) for protein content, and had less similarity between the observed and estimated lactation curves. The performance of Legendre orthogonal polynomials and quadratic splines was strongly affected by the models' order and the number of knots.

Introduction

The accurate description of lactation curves is an interesting topic for breeding and management purposes in dairy farming. Lactation curves are also an efficient tool for the genetic improvement of milk production traits. In the last two decades, a great deal of effort has been devoted to investigating the direct use of test day (TD) yields in the genetic evaluation of dairy cattle (Otwinowska-Mindur et al., 2013) and genetic merit based on test day models was proposed (Ptak and Schaeffer, 1993). Several methodologies have already been

presented for genetic evaluation of production traits in dairy cattle based on test day records (Jaffrezic et al., 2002). Currently, lactation curve fit is used for modelling the random part of the lactation curve in the random regression models, the first functions applied were the Ali and Schaeffer and the Wilmink models (Bohmanova et al., 2008). More recently, Legendre polynomials (Misztal, 2006) and spline models (Meyer, 2005) have been used as the basic function in many analyses by random regression models.

The use of individual curves to select the most accurate model to fit the lactation curve is more

efficient than the use of average curves. Mean lactation curves are usually estimated with a large number of records and are characterized by rather regular patterns (Macciotta et al., 2010). Moreover, individual patterns are of interest for the estimation of the genetic merit of animals (Macciotta et al., 2005) *via* random regression test day models.

This study examined the properties of eleven selected models. The most popular one was first proposed by Wood (1967) to describe the lactation curve (WD). Wilmink (1987) proposed a combination of an exponential and linear term. The Wilmink model (WIL) can be reduced to a three-parameter model by setting the k exponent to a suitable fixed value. In 1987, Ali and Schaeffer suggested a polynomial regression model (AS) to fit the lactation curve; this model is also an important reference that has been used in numerous studies (Silvestre et al., 2006). Legendre orthogonal polynomials (LEG; Kirkpatrick et al., 1990) have been used to model a variety of curves, and some authors have reported their advantages in comparison with several parametric models (Silvestre et al., 2006; Macciotta et al., 2010; Steri et al., 2012). Splines are a type of segmented regression in which the curve is divided into several sections, joined at points called knots, each fitted with different polynomials; splines have been reported to provide extra flexibility in fitting lactation curves (White et al., 1999; Macciotta et al., 2010).

In order to use lactation curve functions in genetic evaluation, identification of the best fitting function is a prerequisite. The purpose of this study was to investigate the suitability of eleven mathematical models for describing the individual lactation curves of milk yield and fat and protein contents from test-day records of Tunisian Holstein-Friesian cows.

Material and methods

Data were supplied by the National Centre for Genetic Improvement (CNAG: Sidi Thabet, Tunis). A sample of 260 241 test days of 5649 cows, recorded in 188 herds from 1994 to 2002 was used. Lactation records with less than 10 consecutive testdays were not included. Likewise, test day yields recorded up to 5 days following calving and very low milk yield (< 3 kg) were omitted. The lactation period was limited to 5–305 days in milk (DIM) after calving. Table 1 displays the structure of the data (average test day milk yield (MY), and percentages of fat (FP) and protein (PP) in milk composition with parity for first, second, third, and greater.

Table 1. Description of data used in this study based on mean and standard deviations for milk yield, fat and protein percentage with parity

Traits	N	Parity	Mean	Min	Max
Milk, kg	82.499	1	18.59 (6.63)	3	63.8
	61.739	2	20.42 (7.67)	3	61.20
	116.003	≥3	20.53 (8.03)	3	59.80
Fat, %	82.499	1	3.40 (0.86)	0.10	9.80
	61.739	2	3.46 (0.90)	0.10	9.80
	116.003	≥3	3.48 (0.89)	0.10	9.50
Protein,%	82.499	1	3.06 (0.41)	0.10	8.40
	61.739	2	3.16 (0.42)	0.10	8
	116.003	≥3	3.12 (0.41)	0.40	8

N – number of test day records, () – standard deviation

Eleven mathematical models (Table 2) available in the literature were selected to fit the individual lactation curves of milk yield. The function of time P_i (Table 2) was calculated using values published by Schaeffer (2004). Quadratic splines were used in this study with one (QSL1), two (QSL2), and three knots (QSL3), the number and locations of knots are very important in fitting splines. Actually, few procedures are available for determining them. In this study, knot positions were estimated by considering them as additional independent variables in a non-linear estimation procedure (Macciotta et al., 2010) and the position of knots N_j was placed at 50, 100, and 200 DIM for cubic splines and at 40, 160, and 200 DIM for quadratic splines. The WD, WIL, and AS models were fitted by the

Table 2. Lactation curve models applied to fit milk yield, fat and protein contents

Model	Function	Reference
WD	$Y_i = a t^b e^{-ct}$	Wood (1967)
WIL	$Y_i = a + b e^{-kt} + ct$	Wilmink (1987)
LEG3	$Y_i = a_0 P_0 + a_1 P_1 + a_2 P_2 + a_3 P_3$	Kirkpatrick et al. (1990)
LEG4	$Y_i = a_0 P_0 + a_1 P_1 + a_2 P_2 + a_3 P_3 + a_4 P_4$	
LEG5	$Y_i = a_0 P_0 + a_1 P_1 + a_2 P_2 + a_3 P_3 + a_4 P_4 + a_5 P_5$	
LEG6	$Y_i = a_0 P_0 + a_1 P_1 + a_2 P_2 + a_3 P_3 + a_4 P_4 + a_5 P_5 + a_6 P_6$	
CSPL	$Y_i = a + b_1 t + b_2 t^2 + b_3 t^3 + \sum c_j (t - N_j)^3$	White et al. (1999)
QSL	$Y_i = a + b_1 t + b_2 t^2 + \sum c_j (t - N_j)^3$	
AS	$Y_i = a + bX + cX^2 + f \log(1/X) + e(\log(1/X))^2$	Ali and Schaeffer (1987)

N_j – position of the knots, $X = t/\theta$ (θ – lactation length). For WIL model $K = 0.65$ for milk; yield and $k = 0.10$ for fat and protein contents. AS – Ali and Schaeffer; LEG3 – third-order Legendre polynomials; LEG4 – fourth-order Legendre polynomials, LEG5 – fifth-order Legendre polynomials, LEG6 – sixth-order Legendre polynomials, WD – wood, WIL – Wilmink, CSPL – cubic spline, QSL – quadratic spline; P_i – function of time with:
 $P_0 = 0.70$, $P_1 = 1.22 \beta$, $P_2 = -0.79 \beta + 2.37 \beta^2$, $P_3 = -2.80 \beta + 4.67 \beta^3$,
 $P_4 = 0.79 \beta - 7.95 \beta^2 + 9.28 \beta^4$, $P_5 = 4.39 \beta - 20.52 \beta^3 + 18.46 \beta^5$
 β – standardized unit of time ranging from -1 to +1 with $\beta = 2 - 1$,
 $\beta = 2 \left(\frac{t - t_{\min}}{t_{\max} - t_{\min}} \right) - 1$, where: t_{\min} – the earliest day in milk and t_{\max} – the latest test day

Levenberg-Marquardt's iterative method in a non-linear regression procedure of SAS (2001); linear regression was used to fit Legendre polynomials (LEG), cubic spline (CSPL) and quadratic spline (QSL). For the WIL model, the parameter k is connected to the time of lactation peak and usually assumes a fixed value derived from a preliminary analysis made on average production (Wilmink, 1987). Goodness of fit was assessed by considering various criteria. The adjusted R square (R^2_{adj}) measures how successful the fit is in explaining the variation of the data. The error in absolute terms (RES) is computed from the difference between the predicted and observed yields per test day and refers to the capability of a model to estimate the daily yield on a specific test day (Guo and Swalve, 1995). The correlation between the measured and estimated test day yield (R) enumerates the degree of association between observed and estimated values (Guo and Swalve, 1995). The root means square error (RMSE) is a frequently used measure of the difference between values predicted by a model and the values actually observed. The standard error of the individual predicted values (STDIN) specifies a variable that contains the standard error of the individual predicted yield. The quotient (Q) between the sum of square errors and the observed sum of squares highlights the similarity between the measured lactation curve and the estimated lactation curve (Ali and Schaeffer, 1987). The analysis of residuals is a common technique in the evaluation and comparison of models, in which a residual for a given record of daily yield is the difference between the observed value and the value predicted by the regression equation (Silvestre et al., 2006); the residual distributions were obtained for each model and plotted against days in milk using the SAS GPLOT procedure (SAS, 2001).

Results

Table 3 shows the percentage of individual lactation curves within each class of adjusted R-squared (C1 to C5) obtained by fitting the different models for all milk traits. All models fit better MY and PP than FP and a difference in the proportion of individual curves was observed between models and traits. The three-parameter models (WIL and WD) and the fourth parameter model (LEG3) presented the lowest percent of curves having an $R^2_{adj} > 0.80$ for MY, FP, and PP. In particular, the fit of the LEG3 model was very poor for FP and had only about 8% of individual curves with an R^2_{adj} higher than 0.80. The fits of models characterized by a high number of parameters show the greatest proportion of individual lactations with

$R^2_{adj} > 0.80$. An increase in the number of curves showing an $R^2_{adj} > 0.80$ was observed with the Legendre orthogonal polynomial and quadratic spline models as the order or the number of knots increased. For example, the percentage of C5 class increased from about 29% (LEG3) to 47% (LEG6) and an increase of 10% was observed for QSL between one and three knots.

Table 4 shows selection criteria for the various models fitted to individual lactation curves for MY, FP, and PP. For the comparison criterion RMSE, AS ranked as the best model to fit MY by obtaining the smallest values, followed by the QSL3, CSPL, and LEG6 models. In contrast to this, the three-parameter models gave a higher magnitude of the error illustrated by large values of RMSE (2.80–2.89). All models predicted the average milk yield with RESmargin of < 2 kg. The AS model predicted yield with an absolute error higher than LEG6 and QSL3. In addition, the QSL3 presented the lowest values of STDIN. For FP and PP results, AS and LEG5 produced similar values of RMSE and STDIN, but LEG5 was superior in terms of the difference between observed and predicted values (RES). However, on the basis of this criterion, QSL3 and LEG6 showed better fitting performances for FP and PP. Moreover, the WD model was very poor, especially for PP. The Q parameter showed a significant difference between models for all milk traits and the lowest value was obtained with LEG6, followed by QSL3. The correlation between observed and estimated yield (R) ranged from 0.94 to 0.97, 0.75 to 0.92, and 0.80 to 0.94 for MY, FP, and PP, respectively. The correlation values (R) for milk yield were high for all models ($R > 0.94$). Similar results were obtained by Otwinowska-Mindur et al. (2013) for Polish Holstein-Friesian cows. Estimated yields (expressed as a percentage for fat and protein) were better correlated for MY and PP than for FP. For MY the AS, LEG6, and QSL3 models showed better performance than the other models that quantified a higher degree of association between observed and estimated yield. In this direction, the AS model was less correlated, especially for FP ($R = 0.87$) and gave a moderate correlation ($R = 0.90$) for PP.

Average residual distributions were obtained for each model and plotted against days in milk in the range from 5 to 305 DIM, to determine the pattern of bias in predicting milk yields or contents. Selected residual distributions are presented in Figure 1 as an example of models giving the worst fit (example chosen for the WD model) and for QSL3 and LEG6, chosen among models giving the best performance by considering statistical criterion results. For MY, the distribution of residual for model WD (Figure 1A) fluctuated throughout the

Table 3. Relative frequencies of fit for individual curves for milk yield, fat and protein percentage among five classes of adjusted R^2

R^2_{adj} class	Lactation curve models										
	AS ¹	LEG3 ²	LEG4 ³	LEG5 ⁴	LEG6 ⁵	WD ⁶	WIL ⁷	CSPL ⁸	QSL1	QSL2	QSL3
Milk yield											
C ₁	14.84	16.24	14.12	12.96	13.99	18.86	17.79	12.10	13.68	12.12	12.96
C ₂	7.85	11.51	9.44	7.92	6.87	12.40	11.57	7.92	9.23	7.80	6.87
C ₃	12.74	16.88	15.42	13.44	11.23	17.82	17.15	13.27	14.80	13.21	11.72
C ₄	22.81	26.57	25.77	24.08	20.97	25.60	25.68	24.05	25.88	23.92	21.33
C ₅	41.76	28.80	35.25	41.60	46.94	25.32	27.81	42.66	36.41	42.95	47.12
Fat content, %											
C ₁	15.51	30.26	23.56	18.12	13.84	30.48	29.65	18.39	23.38	18.32	14.19
C ₂	17.29	25.93	23.44	19.97	16.75	26.18	24.89	19.87	23.80	19.99	16.64
C ₃	17.92	20.70	21.74	21.14	18.67	19.94	19.26	21.10	21.34	21.29	18.64
C ₄	17.94	14.85	18.39	20.29	21.37	14.13	14.69	20.83	18.59	20.54	21.27
C ₅	31.34	8.26	12.87	20.48	29.37	9.27	11.51	19.81	12.89	19.86	29.26
Protein content, %											
C ₁	10.92	22.56	17.21	13.14	10.38	24.18	23.50	12.95	17.58	13.22	10.78
C ₂	13.53	23.27	19.20	16.51	13.21	23.42	21.91	16.15	19.45	16.24	13.06
C ₃	16.56	21.68	21.55	19.33	17.16	20.41	20.25	19.82	21.55	19.80	16.53
C ₄	19.70	19.14	21.71	23.19	22.38	17.90	17.75	23.59	21.61	23.43	22.66
C ₅	39.29	13.35	20.33	27.83	36.87	14.09	16.59	27.49	19.81	27.31	36.95

C₁: $R^2_{adj} < 0.20$; C₂: $0.20 \leq R^2_{adj} < 0.40$; C₃: $0.40 \leq R^2_{adj} < 0.60$; C₄: $0.60 \leq R^2_{adj} < 0.80$; C₅: $R^2_{adj} \geq 0.80$; ¹⁻⁸see Table 2; QSL1 – quadratic spline with one knot, QSL2 – quadratic spline with two knots, QSL3 – quadratic spline with three knots

Table 4. Selection criteria for comparison of lactation models fitted with test-day records for milk yield, fat and protein content

Goodness of fit	Lactation curve models										
	AS ¹	LEG3 ²	LEG4 ³	LEG5 ⁴	LEG6 ⁵	WD ⁶	WIL ⁷	CSPL ⁸	QSL1 ⁹	QSL2 ¹⁰	QSL3 ¹¹
Milk yield											
RMSE	2.49 (1.34)	2.89 (1.33)	2.75 (1.33)	2.66 (1.38)	2.63 (1.45)	2.80 (1.30)	2.83 (1.31)	2.61 (1.34)	2.71 (1.31)	2.62 (1.34)	2.60 (1.44)
RES	1.26 (1.35)	1.88 (1.71)	1.59 (1.52)	1.31 (1.35)	1.03 (1.18)	1.83 (1.70)	1.80 (1.76)	1.26 (1.38)	1.52 (1.57)	1.26 (1.38)	1.02 (1.21)
STDIN	3.10 (1.71)	3.29 (1.56)	3.22 (1.65)	3.15 (1.77)	3.07 (1.95)	3.35 (26.69)	3.26 (1.55)	3.08 (1.72)	3.15 (1.62)	3.08 (1.72)	3.06 (1.93)
Q	1.65	3.13	2.22	1.57	0.94	3.55	3.52	1.71	2.48	1.71	1.12
R	0.97	0.94	0.95	0.95	0.97	0.94	0.94	0.96	0.95	0.96	0.97
Fat content											
RMSE	0.63 (0.35)	0.68 (0.32)	0.66 (0.34)	0.63 (0.36)	0.60 (0.39)	0.67 (0.31)	0.67 (0.32)	0.64 (0.37)	0.66 (0.34)	0.64 (0.37)	0.61 (0.39)
RES	0.30 (0.32)	0.42 (0.38)	0.35 (0.35)	0.28 (0.31)	0.21 (0.27)	0.42 (0.39)	0.41 (0.41)	0.28 (0.33)	0.34 (0.37)	0.28 (0.33)	0.21 (0.28)
STDIN	0.80 (0.46)	0.79 (0.38)	0.80 (0.41)	0.80 (0.46)	0.79 (0.51)	0.79 (0.38)	0.79 (0.39)	0.80 (0.46)	0.80 (0.42)	0.80 (0.46)	0.79 (0.51)
Q	3.22	7.08	5.01	3.34	1.97	7.15	7.06	3.39	5.40	3.38	2.18
R	0.87	0.76	0.82	0.82	0.92	0.75	0.75	0.87	0.81	0.87	0.91
Protein content											
RMSE	0.26 (0.16)	0.28 (0.15)	0.27 (0.16)	0.26 (0.16)	0.24 (0.17)	0.67 (0.31)	0.28 (0.14)	0.26 (0.16)	0.27 (0.15)	0.26 (0.16)	0.25 (0.18)
RES	0.12 (0.14)	0.17 (0.17)	0.14 (0.15)	0.11 (0.13)	0.08 (0.11)	0.42 (0.39)	0.17 (0.18)	0.11 (0.14)	0.14 (0.16)	0.11 (0.14)	0.08 (0.12)
STDIN	0.33 (0.21)	0.34 (0.17)	0.33 (0.19)	0.33 (0.21)	0.32 (0.23)	0.79 (0.38)	0.34 (0.18)	0.33 (0.21)	0.33 (0.19)	0.33 (0.21)	0.32 (0.23)
Q	0.68	1.88	1.16	0.48	0.23	1.86	1.63	0.74	1.50	0.86	0.42
R	0.90	0.80	0.85	0.85	0.94	0.80	0.80	0.89	0.85	0.89	0.93

¹⁻¹¹see Tables 2 and 3; RMSE – the root means square error, RES – the error in absolute term (defined as the absolute values of the difference between the predicted and observed test-day yields), STDIN – the standard error of the individual predicted values, Q – the quotient between the sum of square errors and the observed sum of squares, R – the correlation between the real and estimated values

lactation between positive and negative values, it was negative at the beginning of lactation and from about 270 DIM until the end of lactation. A high positive residual variance was observed between 30 and 70 DIM and also from 160 to 260 DIM. The residual pattern obtained for the WD model for MY is related to that reported by Olori et al. (1999). For FP and PP, a positive residual was observed during the initial phase; then residuals were distributed between negative and positive values on the horizontal axis throughout the rest of lactation, with a higher and more fluctuating magnitude for FP than for PP (Figure 1B and 1C). Residuals plots for other models were not included in this paper, but a similar pattern was observed with LEG3 and LEG4 and comparable results were obtained with the LEG5, WD, and WIL models for all milk traits, but a small difference was detected for the WIL model, with an accurate estimation of only the initial milk constituent (FP and PP).

Average values of residual along the lactation for QSL3 model tended to remain close to zero with a lower magnitude for MY and FP, and were very close to the horizontal axis, especially for PP and MY. Lactation curves were fitted adequately by QSL3, particularly in the first lactation stage and around peak (from DIM 5 to 50 DIM). QSL3 was slightly better in the second part of the lactation trajectory, but estimation improved at the end from about 280 DIM for all traits. Similar trends in the residual distribution were obtained with the CSPL and AS models for all milk traits. Although the performance of LEG6 was among the best according to the statistical criteria results, the patterns found in Figure 1A for MY suggest that the distribution of the errors for LEG6 was not random during lactation and was only close to zero at the end of lactation. While for FP and PP a similar pattern was observed for models QSL3 and LEG6.

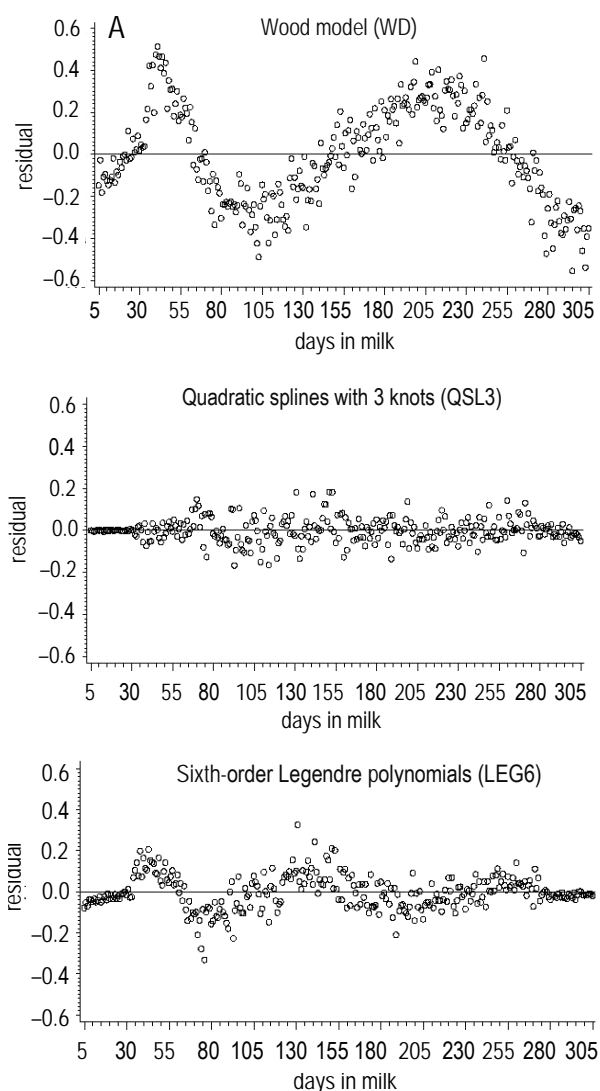


Figure 1A. Distribution of milk average error (residuals) for WD, QSL3 and LEG6 models for milk yield

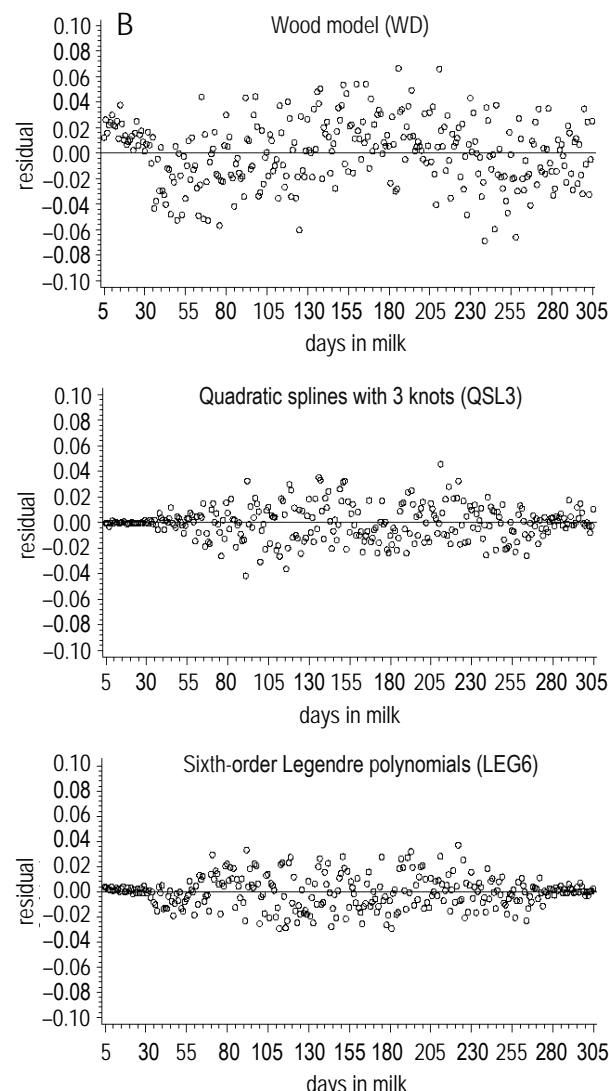


Figure 1B. Distribution of milk average error (residuals) for WD, QSL3 and LEG6 models for fat content

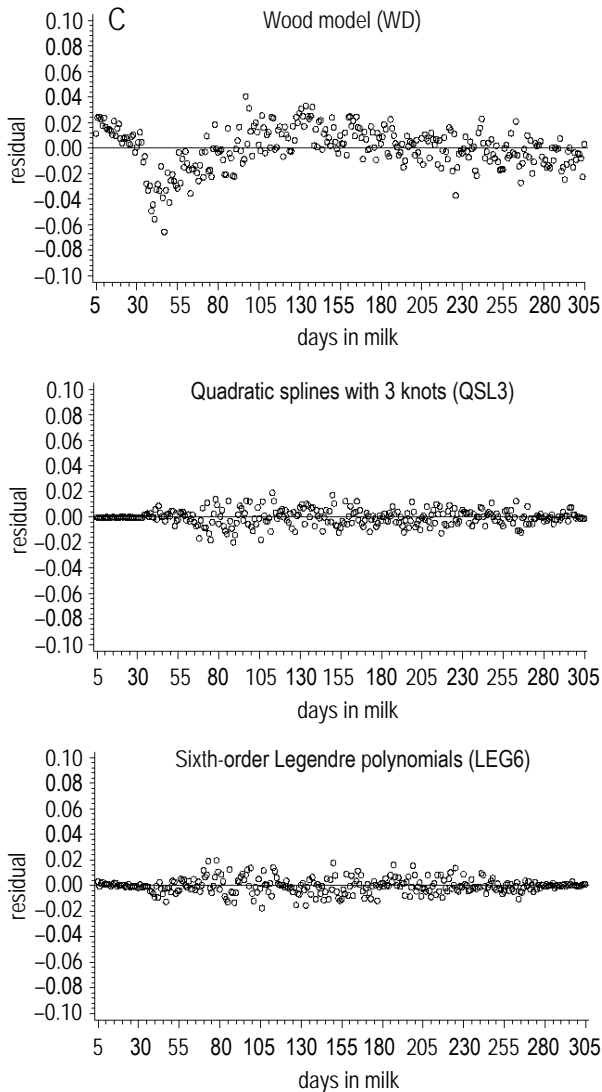


Figure 1C. Distribution of milk average error (residuals) for WD, QSL3 and LEG6 models for protein content

Discussion

The coefficient of determination adjusted by the number of parameters (R^2_{adj}) is equal to the percentage of the variance of daily yield explained by the model. Not all models had the same efficiency in the fitting of the MY and its compositions, the difference in proportions between the number of curves with $R^2_{adj} > 0.80$ was the highest between MY and FP and ranged from 10.42 points (AS) to 23.52 points (QSL3), similar results were noted by Silvestre et al. (2009). Poor fitting performance of individual lactation curves for FP was indicated by Quinn et al., 2006; Silvestre et al., 2009; Steri et al., 2012. In addition, the large relative frequencies of MY curves showing an $R^2_{adj} > 0.80$ are in agreement with previous results reported in dairy cattle (Macciotta et al., 2005; Silvestre et al., 2009; Macciotta et al., 2010; Steri et al., 2012). The R^2_{adj} classification

revealed that models with a higher number of parameters (AS, LEG5 and LEG6) and spline models (CSPL, QSL2 and QSL3) showed better fitting performances for MY. However, AS and LEG5 models had very similar performances for MY, in agreement with Steri et al. (2012). The relative frequencies of R^2_{adj} more than 0.80 for MY were the highest for LEG6 and QSL3. Macciotta et al. (2010) reported an increase in the number of curves showing an $R^2_{adj} > 0.80$ with the quadratic and cubic splines in comparison with AS and fourth-order Legendre orthogonal polynomials for MY. For FP and PP, the AS, LEG6, and QSL3 models gave the best fit to individual lactation curves. Moreover, the LEG6 and QSL3 models presented the same relative frequencies of R^2_{adj} more than 0.80. Steri et al. (2012) illustrated that QSL3 and CSPL showed the best fit for all milk traits, based on the percentage of individual curves having an R^2_{adj} greater than 0.70.

Generally, for all milk traits the goodness of fit increased with the number of function parameters describing the curve, in agreement with earlier suggestions (Kirkpatrick et al., 1990; Meyer, 2005). The high-order orthogonal Legendre polynomial (LEG6) and the quadratic spline with three knots (QSL3) presented similar performance with respect to the RES, STDIN and R parameters for MY. LEG6, recording the lowest Q parameter, indicated a closer similarity between the observed and estimated values. For FP and PP, LEG6 and QSL3 presented exactly the same values of RES and STDIN and almost similar values of the RMSE and R parameters, and presented the same result as MY with respect to the Q parameter. For MY, a general reduction of performance of fit was found with LEG6 in relation to the residual distribution for MY, because the prediction was underestimated at the beginning of lactation and overestimated around the peak yield. Pool et al. (2000) noted similar results in which variance predictions at the extremes of the lactation trajectory were still overestimated, and residuals showed a systematic pattern over the lactation period when the genetic and permanent environmental components were both modelled by a polynomial regression. The QSL3 and the AS models showed a random distribution of residuals for all milk traits. However, Druet et al. (2003) noted that the Legendre polynomials had an increased bias at the beginning and the end of lactation and showed waves in the middle of the lactation. The statistical parameter results obtained in this study for individual lactations indicated that the best performance obtained by a spline model with three knots is close to that of a high-order polynomial function. This may be explained by the suitability

of spline models. However, splines play a role similar to higher order polynomials and spline curves are constructed from pieces of low-degree polynomials, joined at knots.

Non-parametric models (QSL3 and CSPL) showed better fitting performance compared with parametric models (WD and WIL), particularly in the first part of lactation and around the peak, WD compared with QSL3 (Figure 1), which is in agreement with some reports (Silvestre et al., 2006; Macciotta et al., 2010; Steri et al., 2012). This problem was also observed with polynomial models only for MY (Figure 1A) but with a lower magnitude. The difficulty for the parametric function to adequately model the peak yield can be explained by the high degree of correlation between its parameters. This is not the case for a Legendre polynomial function because of the orthogonality propriety of polynomials (Steri et al., 2012). The former models (parametric models) were unable to properly model the peak of the lactation curve. Poor fits noted for this part reached around 70 DIM (around peak yield) and were reported in several studies using parametric models (Olori et al., 1999; Silvestre et al., 2006; Macciotta et al., 2011). We clearly observed this disadvantage in this work with the WD, WIL, and orthogonal polynomials.

Variation in individual fit of the various models was observed among animals and among models. The variation in fit among animals, illustrated mainly by the Q values (Silvestre et al., 2006), suggests that the suitability of the considered model depends on the mathematical form of the model and on individual trends in TD milk yield and its composition, which vary between animal and lactation (as an indication $Q_{WD} = 7.25$; $Q_{LEG6} = 1.97$ and $Q_{WD} = 1.68$; $Q_{LEG6} = 0.23$ respectively for fat and protein contents). These results are similar to previous findings by Silvestre et al. (2006). In addition, Olori et al. (1999) used weekly averages of daily milk yield of 488 first-lactation cows recorded in a single herd to investigate the fit of five models and concluded that there was little difference between the models in fitted average herd lactation and that the variation in fit was almost entirely between cows. Guo and Swalve (1995) evaluated the goodness of fit of 16 lactation curve models for milk yield using data collected from three farms and reported that differences between cows is of great importance. For fat and protein contents, Quinn et al. (2006) concluded that the investigated models were similar in term of goodness of fit using the whole data set.

The results of the present study showed that the most adequate functions for modelling the individual lactation curves for all milk traits in the Tunisian population of Holstein-Friesians were

LEG 6, AS, QSL3, and CSPL, with an advantage for the quadratic spline with three knots. The advantage of the spline can be explained by its flexibility. However, regression splines offer the possibility of using a number and placement of knots along the lactation trajectory (White et al., 1999; Meyer, 2005). In addition, regression splines have advantages for modelling lactation curves, such as limited number of required parameters, good flexibility, smoothness and limited sensitivity to data (Druet et al., 2003).

Conclusions

The present study used test day data to fit individual lactation curves. Eleven mathematical models were compared for accuracy of modelling lactation curves for milk yield and contents. The results indicate that higher-order Legendre orthogonal polynomials (LEG6), cubic and quadratic splines with three knots are more flexible and suitable to fit individual lactation curves for milk yield and its composition in the Tunisian population of Holstein-Friesians, although the goodness of fit shows a certain superiority of the quadratic splines with 3 knots (QSL3) model. The Ali and Schaeffer model also presents a respectable suitability, especially compared with parametric models and low-order polynomial models.

Acknowledgements

The National Centre for Genetic Improvement of Tunisia (CNAG: Sidi Thabet, Tunis) is greatly acknowledged for providing the data used in this study.

References

- Ali T.E., Schaeffer L.R., 1987. Accounting for covariances among test day milk yields in dairy cows. *Can. J. Anim. Sci.* 67, 637–644
- Bohmanova J., Miglior F., Jamrozik J., Misztal I., Sullivan P.G., 2008. Comparison of random regression models with Legendre polynomials and linear splines for production traits and somatic cell score of Canadian Holstein cows. *J. Dairy Sci.* 91, 3627–3638
- Druet T., Jaffrezic F., Boichard D., Ducrocq V., 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy Sci.* 86, 2480–2490
- Guo Z., Swalve H.H., 1995. Modelling of lactation curve as a sub-model in the evaluation of test day records. In: *Proceedings of the Open Session of the Interbull Annual Meeting, Prague (Czech Republic)*. Department of Animal Breeding and Genetics, SLU, Uppsala (Sweden). *Interbull Bulletin* 11, 4–7
- Jaffrezic F., White I.M.S., Thompson R., Visscher P.M., 2002. Contrasting models for lactation curve analysis. *J. Dairy Sci.* 85, 968–975

- Jamrozik J., Kistemaker G.J., Dekkers J.C.M., Schaeffer L.R., 1997. Comparison of possible covariates for use in a random regression model for analyses of test day yields. *J. Dairy Sci.* 80, 2550–2556
- Kirkpatrick M., Lofsvold D., Bulmer M., 1990. Analysis of the inheritance, selection and evaluation of growth trajectories. *Genetics* 124, 979–993
- Macciotta N.P.P., Dimauro C., Rassa S.P.G., Steri R., Pullna G., 2011. The mathematical description of lactation curves in dairy cattle. *Ital. J. Anim. Sci.* 10, 213–223
- Macciotta N.P.P., Miglior F., Dimauro C., Schaeffer L.R., 2010. Comparison of parametric, orthogonal, and spline functions to model individual lactation curves for milk yield in Canadian Holsteins. *Ital. J. Anim. Sci.* 9, 460–464
- Macciotta N.P.P., Vicaro D., Cappio-Borlino A., 2005. Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. *J. Dairy Sci.* 88, 1178–1191
- Meyer K., 2005. Random regression analysis using B-splines to model growth of Australian Angus cattle. *Genet. Sel. Evol.* 37, 473–500
- Misztal I., 2006. Properties of random regression models using linear splines. *J. Anim. Breed. Genet.* 123, 74–80
- Olori V.E., Brotherston S., Hill W.G., Mcguirk B.J., 1999. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livest. Prod. Sci.* 58, 55–63
- Otwinowska-Mindur A., Ptak E., Jagusiak W., Satola A., 2013. Modeling lactation curves of Polish Holstein-Friesian cows. Part I: The accuracy of five lactation curve models. *J. Anim. Feed Sci.* 22, 19–25
- Pool M.H., Janss L.L., Meuwissen T. H., 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. *J. Dairy Sci.* 83, 2640–2649
- Ptak E., Schaeffer L.R., 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci.* 34, 23–34
- Quinn N., Killen L., Buckley F., 2006. Modeling fat and protein concentration curves for Irish dairy cows. *Irish J. Agr. Food Res.* 45, 13–23
- SAS, 2001. SAS® 9.1 Qualification Tools User's Guide. SAS Institute Inc. Cary, NC
- Schaeffer L.R., 2004. Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86, 35–45
- Silvestre A.M.D., Martins A.M., Santos V.A., Ginja M.M, Colaço J.A., 2009. Lactation curves for milk, fat and protein in dairy cows: A full approach. *Livest. Prod. Sci.* 122, 308–313
- Silvestre A.M.D., Petim-Batista F., Colaço J., 2006. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test day records from varying sample schemes. *J. Dairy Sci.* 75, 1813–1821
- Steri R., Dimauro C., Canavesi E.F., Nicolazzi E.L, Macciotta N.P.P., 2012. Analysis of lactation shapes in extended lactations. *Animal* 6(10), 1572–1582
- White I.M.S., Thompson R., Brotherstone S., 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* 82, 632–638
- Wilmink J.B.M., 1987. Adjustment of test day milk fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16, 335–348
- Wood P.D.P., 1967. Algebraic model of the lactation curve in cattle. *Nature* 216, 164–165