J. Dairy Sci. 89:1813–1821

The Accuracy of Seven Mathematical Functions in Modeling Dairy Cattle Lactation Curves Based on Test-Day Records From Varying Sample Schemes

A. M. Silvestre,[1] F. Petim-Batista, and J. Colaço
Departement of Animal Science–CECAV, University of Trás-os-Montes e Alto Douro, 5001-801 Vila Real, Portugal

ABSTRACT

Daily milk yield over the course of the lactation follows a curvilinear pattern, so a suitable function is required to model this curve. In this study, 7 functions (Wood, Wilmink, Ali and Schaeffer, cubic splines, and 3 Legendre polynomials) were used to model the lactation curve at the phenotypic level, using both daily observations and data from commonly used recording schemes. The number of observations per lactation varied from 4 to 11. Several criteria based on the analysis of the real error were used to compare models. The performance of models showed few discrepancies in the comparison criteria when daily or 4-weekly (with first test at days in milk 8) data by lactation were used. The performance of the Wood, Wilmink, and Ali and Schaeffer models were highly affected by the reduction of the sample dimension. The results of this work support the idea that the performance of these models depends on the sample properties but also shows considerable variation within the sampling groups.
Key words: dairy cow, lactation curve

INTRODUCTION

The modeling of lactation curves is not a new research topic. The first reference to a lactation curve model is attributed to Brody et al. (1924). Due to calculation difficulties and computer limitations, the first models of lactation curves were characterized as logarithmic transformations of exponential, polynomial, and other linear functions. At the end of the 1980s and during the 1990s, nonlinear models in the parameters were investigated (Grossman and Koops, 1988). Tozer and Huffaker (1999) characterized average lactation curves for Holstein-Friesian cows according to parity, whereas Tekerli et al. (2000) and Macciotta et al. (2005) studied the shape of lactation curve. Carta et al. (1995) studied the influence of seasonal effects on lactation curves of dairy goats, and Groenewald et al. (1995) and Hohenboken et al. (1992) investigated the suitability of mathematical models to represent the lactation curve of Merino sheep and beef cows, respectively.

Lactation curve models are an essential research tool for developing and validating mechanistic models, aimed at explaining the main features of the milk production pattern in terms of the known biology of the mammary gland during pregnancy and lactation (Neal and Thornley, 1983; Macciotta et al., 2005). The study of the mathematical properties of the lactation curve provides summarized information about dairy cattle production, which is useful in making management decisions (e.g., health monitoring, individual feeding). The models developed are also useful in the genetic analysis of test-day records to account for the effect of lactation stage (Ptak and Schaeffer, 1993) and to model the covariance between test-day records in a random regression analysis (Jamrozik and Schaeffer, 1997).

A variety of different mathematical models have been used in research on the shape of the lactation curve; this study compared the properties of 7 of these models. Wood's (1967) curve (**WOOD**) has been used by a number of researchers, most recently by Varona et al. (1998). The Wilmink (1987) model (**WIL**) was developed in the Netherlands and was the original function used to describe the shape of the lactation curve in the official program of genetic evaluation of Canada (Schaeffer et al., 2000). The logarithm-based model developed by Ali and Schaeffer (1987; **ALI**) is also an important reference that has been used in subsequent studies (Guo and Swalve, 1995; Jamrozik and Schaeffer, 1997). These 3 models have the common characteristic of being developed specifically for lactation curves. Recently, scientists have begun to apply general mathematical tools, including splines and Legendre polynomials (White et al., 1999; Macciotta et al., 2005), to lactation curve modeling.

Current technologies allow the automatic measurement and recording of milk production at every milking. In this study, daily data of 144 lactations were used to achieve 8 sampling schemes, taking into account 4 intervals from calving to first test and 2 intervals be-

Received January 2, 2005.
Accepted November 8, 2005.
[1]Corresponding author: asilvest@utad.pt

Table 1. Summary of daily data from 144 lactations for DIM interval 5 to 305

	Farm 1	Farm 2	Farm 3	Farm 4
Cows in production (June 2001)	78	51	53	30
Number of lactations	61	45	13	25
Lactations/parity (first to fourth)	27, 17, 8, 9	12, 13, 7, 13	2, 3, 1, 7	6, 6, 9, 4
Cow-days	17,561	12,442	3,734	7,402
305-d milk yield, kg	6,494 ± 1,556	8,263 ± 1,651	6,961 ± 1,031	6,506 ± 1,181
Daily milk yield, kg	22.6 ± 7.7	29.9 ± 8.4	24.2 ± 7.9	22.0 ± 7.6

tween tests. The purpose of this work was to investigate the suitability of 7 mathematical functions in modeling the lactation curve from test-day records based on these different sampling schemes.

MATERIALS AND METHODS

Data

The data were obtained from May 1999 to June 2001, from 4 dairy farms with electronic identification and automatic milking recording systems. All farms were located in the north of Portugal. The number of milking cows when the last data recording was made is an indicator of farm dimension (Table 1). On all farms, cows were housed indoors, fed complete rations, and milked twice daily.

Software used on the farms allowed the storage of data from the most recent 60 milkings (30 daily milk yields) for management purposes. However, this parameter was changed to allow the data store of 90 daily milk yields. The database used on this work had the daily milk yield of 144 complete lactations of 139 healthy cows; a total of 45,213 cow-days. Only 41,139 cow-days were considered in this study, however, because records were restricted to those between 5 and 305 DIM. Some lactations had fewer than 305 d. Of the 144 lactations, 47, 39, 25, and 33 were first, second, third, and fourth lactations, respectively. Five cows had data from both first and second lactations. Overall means (± SD) for lactation length, total milk yield, and 305-d milk yield were 317 ± 50 DIM, 7,249 ± 1,892, and 7,091 ± 1,680 kg, respectively (Table 1).

For each lactation, 8 data sets of DIM were made according to the following criteria: 4 intervals from calving to first test (8, 30, 60, and 90 d) and 2 intervals between tests (4 and 8 wk). Table 2 shows the adopted identification of the 8 sampling groups (**SG**; SG1 to SG8) and the mean number of observations per lactation, which ranged from 4 to 11. Not all sampling groups had 144 lactations because a minimum of 4 observations per lactation was required. In addition to these 8 data sets, all daily data were used to test how well the mathematical functions are able to model the lactation curves when all data are available (SG0).

Lactation Curve Models

Seven mathematical functions were applied to fit the milk yield data of individual lactations. The individual fit of lactation curves has been used in previous studies with the purpose of comparing functions or investigating the effect of environmental factors such as herd, parity, and calving season on lactation curve traits (Tekerli et al., 2000; Rekik and Gara, 2004).

The Wood Model. The gamma function described by Wood (1967) is one of the most popular models used to describe the lactation curve:

$$\mathrm{Y_t = at^b e^{-ct}}. \qquad [1]$$

For all models, Y_t is the milk yield in lactation day t. In model [1], parameter a is a scaling factor to represent yield at the beginning of lactation, and parameters b and c are factors associated with the inclining and declining slopes of the lactation curve, respectively.

The Wilmink Model. The WIL model is the following:

$$\mathrm{Y_t = a + be^{-kt} + ct}. \qquad [2]$$

According to Wilmink (1987), the parameters a, b, and c are associated with the level of production, the increase of production before the peak, and with the subsequent decrease, respectively. Parameter k is related to the time of peak lactation and usually assumes a fixed value, derived from a preliminary analysis made on average production. This implies that the model has only 3 parameters to be estimated (Wilmink, 1987; Olori et al., 1999; Schaeffer et al., 2000). In a preliminary analysis, k was estimated at 0.065.

Ali and Schaeffer Model. The ALI model was published by Ali and Schaeffer (1987) in a work where the authors studied 3 lactation curve models with the objective of computing relative efficiencies of selection to change the shape of the lactation curve. This model can be written as follows:

$$\mathrm{Y_t = a + b\gamma_t + c\gamma_t^2 + dW_t + eW_t^2} \qquad [3]$$