J. Dairy Sci. 90:3924–3936
doi:10.3168/jds.2006-790

Modeling Extended Lactations of Holsteins

C. M. B. Dematawewa,*[1] R. E. Pearson,* and P. M. VanRaden†
*Department of Dairy Science, Virginia Polytechnic Institute and State University, Blacksburg 24060
†Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

Modeling extended lactations for the US Holsteins is useful because a majority (>55%) of the cows in the present population produce lactations longer than 305 d. In this study, 9 empirical and mechanistic models were compared for their suitability for modeling 305-d and 999-d lactations of US Holsteins. A pooled data set of 4,266,597 test-day yields from 427,657 (305-d complete) lactation records from the AIPL-USDA database was used for model fitting. The empirical models included Wood (WD), Wilmink (WIL), Rook (RK), monophasic (MONO), diphasic (DIPH), and lactation persistency (LPM) functions; Dijkstra (DJ), Pollott (POL), and new-multiphasic (MULT) models comprised the mechanistic counterparts. Each model was separately tested on 305-d (>280 days in milk) and 999-d (>800 days in milk) lactations for cows in first parity and those in third and greater parities. All models were found to produce a significant fit for all 4 scenarios (2 parity groups and 2 lactation lengths). However, the resulting parameter estimates for the 4 scenarios were different. All models except MONO, DIPH, and LPM yielded residuals with absolute values smaller than 2 kg for the entire period of the 305-d lactations. For the extended lactations, the prediction errors were larger. However, the RK, DJ, POL, and MULT models were able to predict daily yield within a ±3 kg range for the entire 999-d period. The POL and MULT models (having 6 and 12 parameters, respectively) produced the lowest mean square error and Bayesian information criteria values, although the differences from the other models were small. Conversely, POL and MULT were often associated with poor convergence and highly correlated, unreliable, or biologically atypical parameter estimates. Considering the computational problems of large mechanistic models and the relative predictive ability of the other models, smaller models such as RK, DJ, and WD were recommended as sufficient for modeling extended lactations unless mechanistic details on the extended curves are needed. The recommended models were also satisfactory in describing fat and protein yields of 305-d and 999-d lactations of all parities.

Key words: dairy cow, Holsteins, lactation curve, modeling

INTRODUCTION

Lactation curves are a valuable tool in designing suitable breeding and management strategies for cattle (Papajcsik and Bodero, 1988; Beever et al., 1991; Pietersma et al., 2001) as well as other species (Gipson and Grossman, 1990; Ruiz et al., 2000). Population- and time-specific lactation models help genetic selection (Dekkers et al., 1996; VanRaden et al., 2006), predict yield from incomplete data, analyze yield responses to management and environmental changes, diagnose problems, and identify opportunities for increased net merit effectively (Scott et al., 1996; Pietersma et al., 2001; Val-Arreola et al., 2004).

Presently in a number of countries, many cows have lactations extended beyond 305 d (Vargas et al., 2000). Lactation length has increased by about 30 d over the last decade in some populations (Gonzalez-Recio et al., 2004). Recent studies show that over 55% of US Holstein cows record lactations longer than 305 d (Tsuruta et al., 2005; VanRaden et al., 2006). The undesirable trend that exists with loss of fertility and reproductive failures in dairy cattle (Butler, 1998; Silvia, 2003) is a well-known contributor to extended lactations. However, extended lactations could be a part of management strategy (Tarazon-Herrera et al., 2000; Gonzalez-Recio et al., 2006). Almost all validations and uses of the lactation models reported in literature have been for 305-d or shorter lactations, with rare exceptions (Vargas et al., 2000; Grossman and Koops, 2003) in which lactations extending up to 18 mo have been examined. Papajcsik and Bodero (1988) cited a list of 20 different empirical formulas developed since 1923. The incomplete gamma function (**WD**) proposed by Wood (1967) was the earliest popular model conceived for the whole lactation, although inverse polynomial (**IP**; Nelder, 1966) is still a favorite choice for modeling (Batra, 1986; Scott et al., 1996). Subsequent attempts to

Received November 27, 2006.
Accepted March 28, 2007.
[1]Corresponding author: mahinda@vt.edu

Table 1. Number of records, average 15th day and peak milk yields (kg), and peak day in milk for 305-d (>280 DIM) and 999-d (>800 DIM) lactations for first and third and greater parities

Parity	Lactation length (d)	Lactation records, n	Test-day records, n	d 15 yield	Peak yield	Peak DIM
First	305	235,241	2,128,015	26.4	33.1	94
	999	1,692	38,495	25.4	35.0	98
Third and greater	305	237,416	2,138,582	38.0	44.3	51
	999	1,012	22,487	35.2	45.6	70

improve WD with respect to its functional form, mathematical properties, and forecasting ability were reviewed by Beever et al. (1991). Wilmink's exponential (**WIL**) function (Wilmink, 1987) has an advantage over WD and IP because the initial yield is not forced to zero. Rook et al. (1993) showed that their Mitscherich × exponential and Michaelis-Menten × exponential functions, which described the lactation curve as a product of growth and death processes of mammary cells, fit better than the WD model. Val-Arreola et al. (2004) found that the Michaelis-Menten × exponential function (**RK**) is the most superior model of Rook et al. (1993). Grossman and Koops (1988) showed that both WD and IP overpredict yield during early lactation and underpredict the peak, and produce autocorrelated residuals. Alternatively, they proposed a multiphasic curve using sums of logistic functions to overcome those problems. Their monophasic (**MONO**) and diphasic (**DIPH**) variants (with 3 and 6 parameters, respectively) were supposed to be the optimal versions (Gipson and Grossman, 1990). However, many have criticized the multiphasic model for its lack of a biological basis (Beever et al., 1991; Rook et al., 1993). Grossman et al. (1999) also proposed a model to measure persistency

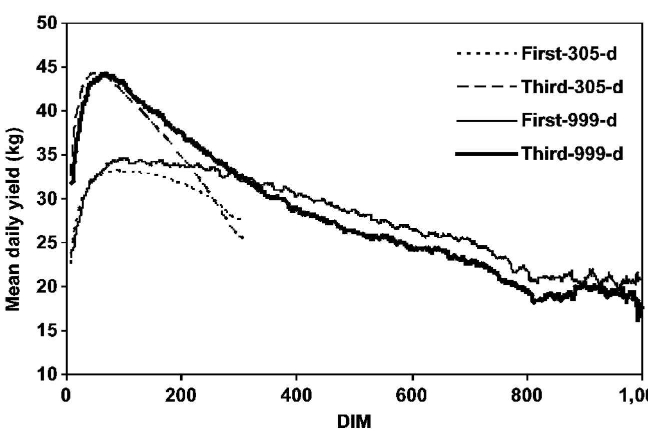

Figure 1. Observed daily mean yields from 305-d and 999-d lactations of cows in the 2 parity groups. First-305-d and first-999-d = first-parity cows with >280 DIM and >800 DIM, respectively; third-305-d and third-999-d = cows in third and greater parities with >280 DIM and >800 DIM, respectively.

of lactation. The proposed lactation persistency model (**LPM**) assumes a constant yield in mid lactation, and the length of this period is the measure of lactation persistency.

Alternatively, several mechanistic models have been developed to simulate the metabolic response of the whole animal with each term of the model having a biological interpretation (Beever et al., 1991). The initial mechanistic model of Neal and Thornley (1983) was based on mammary cell differentiation and programmed cell death (apoptosis), but had limited practical use due to nonavailability of information on inputs required. Subsequently, Dijkstra et al. (1997) developed a 4-parameter model that describes the mammary growth pattern (cell proliferation and apoptosis) of mammals during pregnancy and lactation. The Dijkstra model (**DJ**) was found to fit well for lactations of dairy cows (Dijkstra et al., 1997; Val-Arreola et al., 2004). Pollott (2000) proposed a model that mimics 3 processes including mammary cell differentiation, apoptosis, and milk secretion rate per cell. The basic Pollott model (**POL**) contained 6 parameters, with an additional parameter for each new factor (secondary growth, pregnancy, etc.). An alternative mechanistic version of the multiphasic model was proposed by Grossman and Koops (2003). This new multiphasic model (**MULT**) contained 13 parameters that could be interpreted with respect to the shape of the curve. Such large models are expected to have better predictive ability but could be computationally demanding.

Alternative models have been compared with respect to their ability to fit individual lactation curves of various shapes (Macciotta et al., 2005; Silvestre et al., 2006). Alternatively, the models have been used to describe average lactation curves of various groups such as parities and lactation lengths (Tozer and Huffaker, 1999; Vargas et al., 2000; VanRaden et al., 2006). Group-average lactation curves are useful in design of optimization models, simulations, reproductive and management strategies (Freeze and Richards, 1992; Vargas et al., 2000; Pietersma et al., 2001), and breeding (VanRaden et al., 2006).

Until recently, lack of sufficiently large datasets on extended lactations has been a hindrance to modeling