Random regression models using different functions to model milk flow in dairy cows

M.M.M. Laureano[1], A.B. Bignardi[2], L. El Faro[2], V.L. Cardoso[2], H. Tonhati[1] and L.G. Albuquerque[1]

[1]Departamento de Zootecnia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, SP, Brasil
[2]Agência Paulista de Tecnologia dos Agronegócios, Pólo Regional Centro Leste, Ribeirão Preto, SP, Brasil

Corresponding author: M.M.M. Laureano
E-mail: monyka.laureano@gmail.com

ABSTRACT. We analyzed 75,555 test-day milk flow records from 2175 primiparous Holstein cows that calved between 1997 and 2005. Milk flow was obtained by dividing the mean milk yield (kg) of the 3 daily milking by the total milking time (min) and was expressed as kg/min. Milk flow was grouped into 43 weekly classes. The analyses were performed using a single-trait Random Regression Models that included direct additive genetic, permanent environmental, and residual random effects. In addition, the contemporary group and linear and quadratic effects of cow age at calving were included as fixed effects. Fourth-order orthogonal Legendre polynomial of days in milk was used to model the mean trend in milk flow. The additive genetic and permanent environmental covariance functions were estimated using random regression Legendre polynomials and B-spline functions of days in milk. The model using a third-order Legendre polynomial for additive genetic effects and a sixth-order polynomial for permanent environmental effects, which contained 7 residual classes, proved to be the most adequate to describe variations in milk flow, and was also the

most parsimonious. The heritability in milk flow estimated by the most parsimonious model was of moderate to high magnitude**.**

Key words: B-spline function; Functional trait; Legendre polynomials; Milkability; Milk production

INTRODUCTION

Dairy cattle breeding programs are mainly based on traits related to the production of milk and its components. However, research has indicated that functional traits should be included in these programs to minimize any possible negative effects of selection for milk yield on reproductive and health traits (Cobuci et al., 2006). Hence, milk flow represents an important trait because it is associated with milk production efficiency or ability, as well as with the susceptibility of animals to mastitis (Zwald et al., 2005).

Milk flow, or milking time, has received limited research focus because this trait is difficult to measure. Most studies traditionally employ repeatability and single-trait or multiple trait models to evaluate this trait. Repeatability models have been used to investigate milk flow in primiparous Holstein cows using either a scale of visual scores ranging from 1 to 3 (Erf et al., 1992) or objective measures (milking time, kg/min) (Rensing and Ruten, 2005). Although milk flow is the data obtained subjectively during lactation usually, this parameter tends to be analyzed by repeatability or ordinary multi-trait models (König et al., 2006; Dodenhoff and Emmerling, 2009; Gray et al., 2011, 2012; Samoré et al., 2011; Laureano et al., 2012).

Random regression models (RRM) that apply Legendre polynomials have been used to describe variations in the longitudinal data of dairy cattle, including milk production (Costa et al., 2008; El Faro et al., 2008; Bignardi et al., 2009a,b, 2011). In these studies, high-order polynomials were found to be the most adequate way of modeling changes in the mean and variance over time (El Faro et al., 2008; Bignardi et al., 2009a). This approach increases the number of parameters being estimated, with consequently high computational requirements to estimate variance components. However, according to Meyer (2005), high-order polynomials might lead to errors in the estimates of genetic parameters, mainly because of oscillations at the extremes of the curve. One alternative that is currently being studied to reduce the order of these polynomials, and to minimize estimation problems, is the application of segmented polynomials, or spline functions. These functions are formed by segments of low-degree polynomials that are joined together at fixed points, termed knots. Spline functions have some advantages depending on the way they are modeled. Such advantages include their being easy to estimate, with simple biological interpretations. In addition, because the curve is formed by segments, this permits greater flexibility in modeling possible seasonal variations in the observations (Meyer, 2005). In animal breeding, segmented polynomials have been used in random regression analyses to fit the lactation curve of dairy cattle (White et al., 1999; Bignardi et al., 2011) and the growth curve of beef cattle (Meyer, 2005). Among the different types of spline curves, the so-called B-splines have been extensively studied because of their excellent numerical properties (Eilers and Marx, 2010), as well as their efficiency in modeling random effects in mixed models and in estimating covariance functions (Rice and Wu, 2001).

The objective of the present study was to estimate covariance functions and genetic parameters for milk flow in Holstein cows, applying Legendre polynomials and B-spline functions in RRM, to evaluate the possible inclusion of this trait in selection programs.

MATERIAL AND METHODS

Data were provided by the Agrindus Corporation, one of the most important farms in the State of São Paulo, southeastern Brazil. A total of 75,555 once weekly test-day milk flow records from 2175 primiparous Holstein cows that calved between 1997 and 2005 were analyzed. Daily records were obtained from cows that were milked 3 times a day with an electronic milking machine. The management herd was completely intensive, and the cows had a high genetic potential for milk yield.

Milk flow was obtained by dividing the mean milk yield (kg) of the 3 milkings by the total milking time (min) and was expressed as kg/min. Milk flow was grouped into 43 weekly classes. Lactation with at least 3 individual records was included. The contemporary group was defined as year and month of test, with the restriction that each group should contain a minimum of 7 records. The contemporary groups were defined as herd-test-week, for a total of 364 groups, and contained at least 8 animals per class. The relationship matrix included 9572 animals.

The analyses were performed using a single-trait RRM that included direct additive genetic, permanent environmental, and residual random effects. In addition, the contemporary group and linear and quadratic effects of cow age at calving were included as fixed effects. Fourth-order orthogonal Legendre polynomial of days in milk was used to model the mean trend. The additive genetic and permanent environmental covariance functions were estimated using random regression Legendre polynomials and B-spline functions of days in milk flow.

The B-spline functions were used to fit trends using segmented functions, as described by Meyer (2005). Each individual segment was considered to be linear (L), with a baseline function of p = 1 degree. Polynomials of different orders were included in the model for direct additive and permanent environmental random effects. Up to 7 equally spaced knots were considered for direct additive genetic and animal permanent environmental effects. According to Meyer (2005), in RRM using B-spline functions as a baseline function (a type of penalized spline), the choice of knot position is less crucial. M knots were selected to divide age at recording into m-1 equally spaced intervals, and the external knots are located in classes 1 and 43 for all models. The number of random regression coefficients to model the trajectory of the linear basis functions is given by m.

For models using Legendre polynomials, step functions were initially fitted using 43 classes of residual variance. Next, the pattern of variation was analyzed to define step functions with 7, 10, and 20 classes of residual variance. In this case, the weeks of lactation were divided as follows: 1, 2, 3, 4-9, 10-19, 20-39, and 40-43 weeks for the model with 7 classes; 1, 2, 3, 4-9, 10-15, 16-19, 20-30, 31-39, 40, and 41-43 weeks for the model with 10 classes, and 1, 2, 3, 4, 5-8, 9, 10-15, 16-19, 20-22, 23-25, 26-27, 28-30, 31, 32-33, 34-35, 36-37, 38-39, 40, 41, and 42-43 weeks for the model with 20 classes. For the B-spline model, residual variances were modeled by a step function with 7 classes of variance.

The matrix representation of the model is:

$$y = Xb + Za + Wpe + e \qquad \text{(Equation 1)}$$

where y is the vector of observations measured in Nd animals; b is the vector of fixed effects; a is the vector of additive genetic random coefficients; pe is the vector of animal permanent

environmental random coefficients; *e* is the vector of N different residuals; and *X, Z,* and *W* are the corresponding incidence matrices for fixed and additive genetic and permanent environmental random effects, respectively.

The (co)variance components and genetic parameters were estimated by the restricted maximum likelihood method (REML) using the Wombat package (Meyer, 2007).

The RRM were compared using Akaike's (AIC) and Schwarz's Bayesian (BIC) information criteria (Wolfinger, 1993), as well as by the inspection of the estimates of variance and of correlations between milk flow on different test-days. The AIC and BIC allow non-hierarchical models to be compared, and to penalize models that contain a larger number of parameters, with the BIC attributing a more rigorous penalty.

The model selection criteria are described as follows:

$$AIC = -2\log L + 2np \qquad \text{(Equation 2)}$$

$$BIC = -2\log L + p\log(N - r) \qquad \text{(Equation 3)}$$

where *np* is the number of parameters in the model; *N* is the total number of observations; *r* is the rank of the incidence matrix of fixed effects in the model, and *log L* is the logarithm of the restricted maximum likelihood function. Lower AIC and BIC values indicate a better fit. In addition, the variance components and genetic parameters estimated for each RRM were also used as criteria to select the most adequate model to describe the covariance structure of the data.

Spearman correlation coefficients between test-day milk flow breeding values estimated in different analyses were calculated. The breeding values for 305-day accumulated milk flow records were predicted for each sire based on the set of random regression coefficients.

The RRM are referred to as Mk$_a$k$_{pe}$_r, with M corresponding to the Legendre polynomials (LEG) or B-spline functions (BSL = linear, BSQ = quadratic, and BSC = cubic), k$_a$ and k$_{pe}$ correspond to the order for additive genetic and permanent environmental covariance functions, respectively, and r is the residual structure of variance.

RESULTS AND DISCUSSION

Mean milk flow on the test-day was 2.36 kg/min, with a standard deviation of 0.64 kg/min and coefficient of variation of 27.4%. Figure 1 shows that milk flow increased until week 25 (2.50 kg/min), after which there was a slight decline until the end of lactation (2.31 kg/min). The largest variation in milk flow was observed at the end of lactation (29.5%). The lowest mean milk flow rates were observed at the beginning of lactation, and were attributed to lower milk production. In addition, first lactation was studied, which was when cows were at the beginning of their productive life, i.e., when they were milked for the first time.
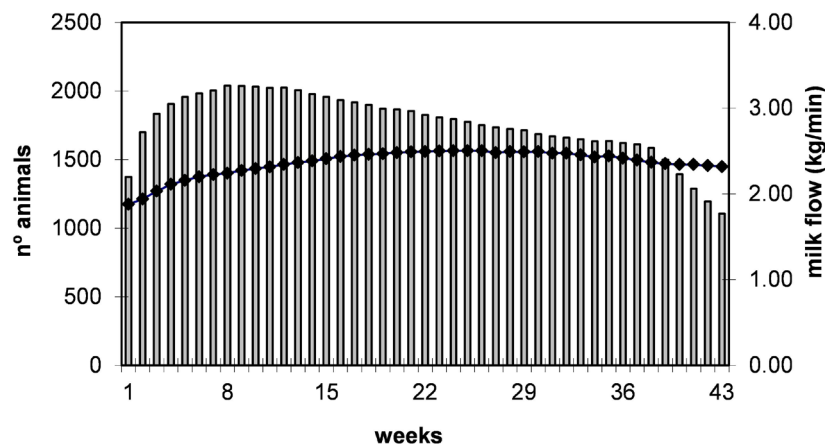
Figure 1. Number of animals (bars) and milk flow (lines).

An order of 3 for k_a and k_{ap} was used in the first 5 RRM (LEG33_1, LEG33_7, LEG33_10, LEG33_20, and LEG33_43), and only the number of residual variance classes was changed to determine which structure would be more appropriate to model this effect (Table 1). The model considering residual variance homogeneity (LEG33_1) provided the poorest fit as indicated by the model selection criteria. This result indicates that the behavior of residual variance differs across lactation, and indicates the need to consider the heterogeneity of variance. The likelihood ratio test of nested models only demonstrated significant differences between the model containing 7 residual variance classes (LEG33_7) and the model considering variance homogeneity. According to the AIC criterion, the model LEG33_10 provided the best fit among the models considering different residual variance classes. However, the BIC criterion, which more rigorously penalizes models containing a large number of parameters, selected the model LEG33_7 as the best. Based on the BIC, the model considering 7 residual variance classes was selected as the most adequate, since it contained a smaller number of parameters, which is desired to prevent convergence problems in the estimation of covariance components (Rekaya et al., 1994). Olori et al. (1999) demonstrated that third-order models for additive genetic and permanent environmental effects resulted in few changes in mean genetic parameter estimates when the residual variance structure was modeled using 1, 4, 10, and 37 classes. This observation indicates that the use of more fragmented residual classes is often not necessary, since, in addition to models containing fewer residual classes being parsimonious, parameter estimates are not affected.

An increase in the order of the covariance functions for additive genetic effects from 3 to 4 resulted in little change in log L values (Table 1). Improvement of log L, AIC, and BIC values was observed when the order of polynomials for additive genetic effects was fixed, and the order of polynomials for permanent environmental effects was modified. According to log L and AIC values, the model LEG310_7 provided the best fit, whereas the BIC criterion selected the LEG39_7 as the best model. The likelihood ratio test showed no significant difference in the likelihood function between models LEG39_7 and LEG310_7, indicating that ninth-order polynomials were sufficient to model permanent environmental effects. The variance component estimates obtained with model LEG36_7 were similar to those obtained with models LEG39_7 and LEG310_7. The latter 2 models also provided similar estimates. However, LEG36_7 contained a smaller number of parameters (34) compared to models

LEG39_7 (58) and LEG310_7 (68), which permitted the model to achieve faster convergence in the estimation of variance components. This model was, therefore, more parsimonious.

The RRM employing linear, quadratic, and cubic B-spline functions was fitted by always considering the same polynomial order for additive genetic and permanent environmental effects. Among the models studied, log L values were found to increase with increasing number of parameters, irrespective of the B-spline function tested (Table 2). In addition, the most complex function (cubic B-spline) provided the best log L value, when comparing models with the same number of parameters using linear, quadratic, and cubic functions. Comparison of models with the same number of parameters according to the AIC and BIC criteria showed that models employing cubic B-splines were consistently superior.

According to the criteria shown in Table 2, BSC88 (5 segments) was the model that best fitted the data, followed by model BSQ88 (6 segments), with both models containing 79 parameters. However, the model BSQ88, which contained a larger number of segments and knots, presented convergence problems in the estimation of variance components, and its use was less advantageous when compared to BSC88. According to White et al. (1999), the computational requirements of models that employ spline functions depend on the number and position of fixed knots.

The log L value of model BSC88 (3050) was similar to that obtained when third- and ninth-order orthogonal Legendre polynomials with 7 classes of heterogeneous variance were used to model additive genetic and permanent environmental effects, respectively (log L = 3046). The AIC values obtained for the 2 models were also similar. In contrast, the BIC criterion, which rigorously penalizes more parameterized models, demonstrated the superiority of the model employing Legendre polynomials (58 parameters, LEG39_7).

Comparison of the parameter estimates obtained with models BSC88 and BSQ77, which contained the same number of segments but used different B-splines (cubic and quadratic, respectively), revealed practically no differences in the variance component estimates (not shown), although AIC and BIC values indicated the superiority of model BSC88.

Table 1. Order of the covariance function for additive genetic (k_a) and permanent environmental (k_{pe}) effects, residual variance structure (e), number of variance components (np), log likelihood function (log L), Akaike's information criterion (AIC), Schwarz Bayesian information criterion (BIC), and likelihood ratio test (LRT) for the different models tested.

Model	k_a	k_{pe}	e	np	log L	AIC	BIC	LRT[a]
LEG33_1	3	3	1	13	283	-540	-420	(2-1) 550**
LEG33_7	3	3	7	19	833	-1627	-1452	(3-2) 15[NS]
LEG33_10	3	3	10	22	848	-1652	-1449	(4-3) 13[NS]
LEG33_20	3	3	20	32	861	-1657	-1362	(5-4) 11[NS]
LEG33_43	3	3	43	55	872	-1633	-1126	-
LEG34_7	3	4	7	23	1716	-3385	-3173	(7-6) 557**
LEG35_7	3	5	7	28	2273	-4489	-4231	(8-7) 388**
LEG36_7	3	6	7	34	2661	-5255	-4941	(9-8) 180**
LEG37_7	3	7	7	41	2841	-5601	-5222	(10-9) 117**
LEG38_7	3	8	7	49	2958	-5818	-5366	(11-10) 88**
LEG39_7	3	9	7	58	3046	-5975	-5440	(12-11) 55[NS]
LEG310_7	3	10	7	68	3101	-6066	-5438	-
LEG44_7	4	4	7	27	1726	-3398	-3149	(14-13) 555**
LEG45_7	4	5	7	32	2281	-4497	-4202	(14-15) 390**
LEG46_7	4	6	7	38	2671	-5265	-4914	(16-15) 180**
LEG47_7	4	7	7	45	2851	-5612	-5196	(17-16) 116**
LEG48_7	4	8	7	53	2967	-5829	-5339	(17-18) 88**
LEG49_7	4	9	7	62	3055	-5985	-5413	-

[a]Likelihood ratio test between hierarchical models. **P < 0.01; NS = not significant.

Table 2. Number of parameters (*np*) in the model, number of segments (*s*), position of knots (*pk*), log likelihood function (log L), Akaike's information criterion (AIC), and Schwarz Bayesian information criterion (BIC) for the different models tested.

Model	*np*	*s*	*pk*	log L	AIC	BIC
BSL33	19	2	1, 22, 43	821	-1604	-1429
BSL44	27	3	1, 15, 29, 43	1686	-3317	-3068
BSL55	37	4	1, 12, 22, 33, 43	2109	-4144	-3803
BSL66	49	5	1, 10, 18, 27, 35, 43	2405	-4713	-4261
BSL77	63	6	1, 8, 15, 22, 29, 36, 43	2655	-5184	-4603
BSQ44	27	2	1, 22, 43	1805	-3556	-3307
BSQ55	37	3	1, 15, 29, 43	2278	-4483	-4141
BSQ66	49	4	1, 12, 22, 33, 43	2623	-5148	-4696
BSQ77	63	5	1, 10, 18, 27, 35, 43	2876	-5625	-5044
BSQ88	79	6	1, 8, 15, 22, 29, 36, 43	3016	-5874	-5145
BSC55	37	2	1, 22, 43	2327	-4579	-4238
BSC66	49	3	1, 15, 29, 43	2695	-5292	-4840
BSC77	63	4	1, 12, 22, 33, 43	2914	-5720	-5121
BSC88	79	5	1, 10, 18, 27, 35, 43	3050	-5942	-5213

BSX k, where *X* corresponds to linear (L), quadratic (Q) or cubic (C), i.e., the degree of the polynomial segment, and *k* is the specific number of random regression coefficients for direct additive genetic and permanent environmental effects, respectively. For all models residual variances were modeled by a step function with 7 classes of variance.

Figure 2 shows the phenotypic (σ^2_p), genetic (σ^2_a), and permanent environmental ($\sigma^2_{pe} + \sigma^2_e$) variance estimated with selected RRMs (LEG36_7, LEG39_7, LEG310_7, BSQ77, and BSC88). The phenotypic variance estimated with the models showed the same trend, increasing until week 7, and then remaining constant until week 40, after which it again increased until the end of lactation, when the highest values were observed. Similar results have been reported by Laureano et al. (2012), who analyzed the same dataset by multi-trait analysis. The additive genetic variance obtained with models LEG36_7, LEG39_7, and LEG310_7 was similar throughout the milk flow curve and was of lower magnitude at some points when compared to multi-trait analysis (Laureano et al., 2012). In contrast, the additive genetic variance estimated by models BSQ77 and BSC88 was higher compared to that obtained when orthogonal Legendre polynomials were used, especially during the first half of lactation. The variance estimated using orthogonal Legendre polynomials was constant, and showed no major oscillations across lactation, resulting in a smoother curve compared to that obtained by the models using B-splines. The additive genetic variance estimated by models BSQ77 and BSC88 was also higher compared to those obtained by Laureano et al. (2012), who used a multi-trait model, particularly during the first half of lactation.

Similar residual variance was obtained by the models fitting B-spline functions and orthogonal Legendre polynomials. Thus, compensation of permanent environmental variance in relation to additive genetic variance occurred when using models that fitted Legendre polynomials. The highest residual variance was observed at the end of lactation for all models. This finding might be explained by the smaller number of animals with records close to the dry period.

Since models LEG36_7, LEG39_7, and LEG310_7 estimated similar levels of variance, the model including the smallest number of parameters (LEG36_7) could be used to analyze milk flow without any loss of precision.

Similar heritability was obtained by the models using Legendre polynomials, ranging from 0.23 to 0.33 for LEG36_7 and from 0.22 to 0.37 for LEG39_7 and LEG310_7 (Figure
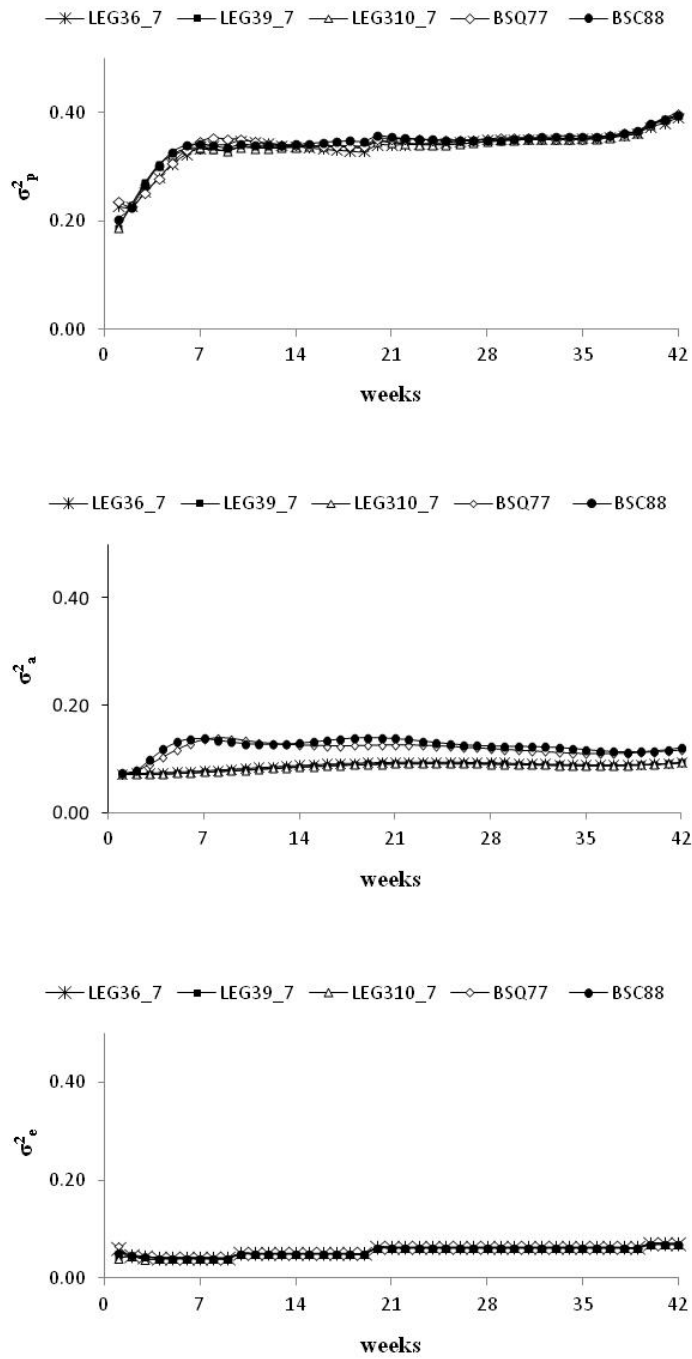
Figure 2. Phenotypic (σ^2_p), genetic (σ^2_a), and residual variances estimated with the LEG36_7, LEG39_7, LEG310_7, BSQ88, and BSC77 random regression models.

3). For these 3 models, the highest heritability was observed during the first 3 weeks of lactation, and between weeks 19 and 27. The heritability estimates obtained by the models using B-splines was higher compared to those estimated using Legendre polynomials. Similar heritability was estimated by models BSQ77 and BSC88 (Figure 3), ranging from 0.28 to 0.39 for the model fitting a quadratic function and from 0.30 to 0.40 for the model fitting a cubic function. The heritability obtained by the 2 models showed no important oscillations until week 21, after which it decreased until the end of lactation. The tendencies of the heritability estimates were similar to those reported by Zwald et al. (2005), Dodenhoff and Emmerling (2009), Gray et al. (2011, 2012), and Laureano et al. (2012).

Smaller oscillations in the heritability estimates after week 7 of lactation were observed for the models using Legendre polynomials. This result indicates that segmented polynomials might not be adequate for the analysis of milk flow, since these functions are adequate for data that present wide variation over time, which was not the case in the present study. When studying the growth data of beef cattle, Meyer (2005) found no important differences between estimates obtained with RRM using B-splines and those fitting orthogonal Legendre polynomials. In addition, the author reported that B-spline functions were more adequate for fitting the data, since they are more robust to estimation problems at the extremes of the growth curve.

The heritability estimates for milk flow reported in the literature vary widely. This variation probably occurs because this trait may be measured in different ways. The heritability estimates for objectively measured flow rate generally seem to be higher compared to those for subjectively scored milking speed (Boettcher et al., 1998, Rupp and Boichard, 1999; Ilahi and Kadarmideem, 2004; Sewalem et al., 2011). The most common practice reported in the literature is to obtain only 1 measure of scores during lactation, and to analyze this measure together with traits related to health and conformation using single-trait or multiple trait models (Luttinem and Juga, 1997; Rupp and Boichard, 1999; Vicario et al., 2006). However, the scores used in these studies differ in terms of the scale adopted, attributing scores of 1 to 3 or 1 to 8 for milk flow, a fact impairing the comparison of heritability found for this trait. Despite the divergence in the measurement and analysis of milk flow, the heritability obtained in the present study was similar to that reported in the published literature.
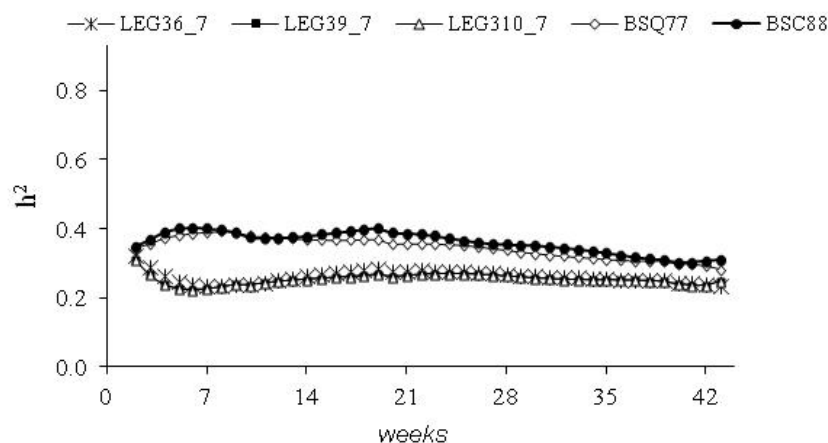

Figure 3. Heritability (h^2) estimates for weekly test-day milk yield obtained with the LEG36_7, LEG39_7, LEG310_7, BSQ88, and BSC77 random regression models.

Since the order of the covariance function for genetic effects was low, important oscillations in the genetic correlation estimates were not observed for the Legendre models (Figure 4). The genetic correlations were positive, and ranged from 0.54 to 0.99 for LEG36_7 and from 0.51 to 0.99 for LEG39_7 and LEG310_7. The phenotypic, genetic, and permanent environmental correlations estimated with models BSQ88 and BSC77 are presented in Figure 5. The genetic correlation between milk flow on different test-days was positive, and ranged from 0.51 to 1.0 and from 0.26 to 1.0 for these 2 models, respectively. Despite similar values, the correlations oscillated less when a quadratic, rather than a cubic, function was used. The phenotypic correlations estimated by model LEG36_7 were similar to those obtained by the models using B-splines. However, the genetic and permanent environmental correlations estimated using orthogonal polynomials oscillated less compared to those obtained by the models fitting B-splines. This finding differs to those reported by other investigators who observed more uniform growth and milk yield curves when the data were fitted using B-spline functions (Meyer, 2005; Bignardi et al., 2011).

For all of the tested models, the highest genetic correlations were observed between adjacent test-days, with the magnitude of the correlations decreasing with increasing distance between test-days. These results indicate that selection for increased milk flow at a certain point during lactation has a positive effect on any other point of the curve. Karacaören et al. (2006) reported the same trend in genetic correlations between daily milk flow records, with values being close to unity for correlations between adjacent test-days and values of about 0.23 for correlations between more distant test-days (e.g., 270 to 305 days of lactation). Similar patterns were reported in studies on milk yield (Cobuci et al., 2005; Araújo et al., 2006; Bignardi et al., 2009a).
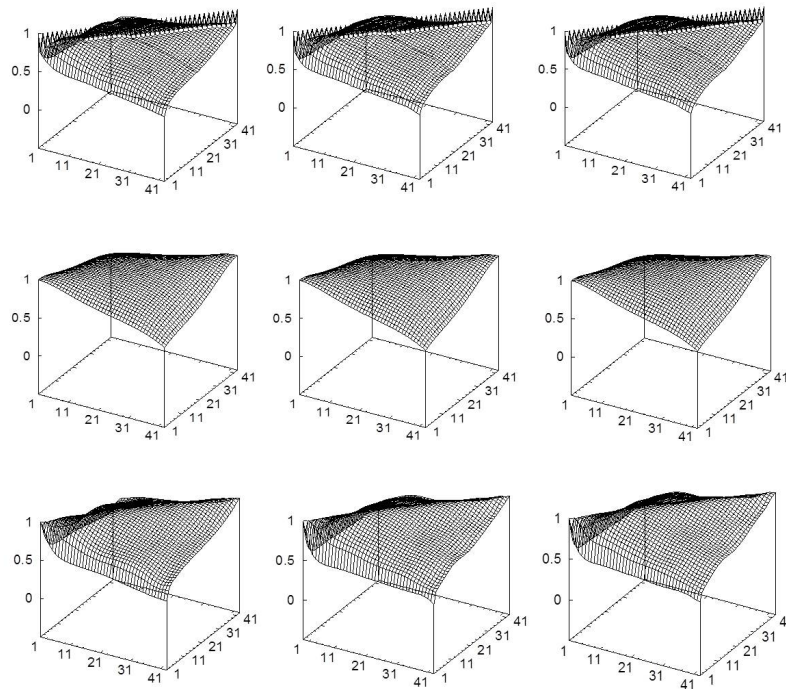

Figure 4. Phenotypic (1st row), genetic (2nd row), and permanent environmental (3rd row) correlation estimates between weekly milk flow obtained with the LEG36_7, LEG39_7, and LEG310_7 models.
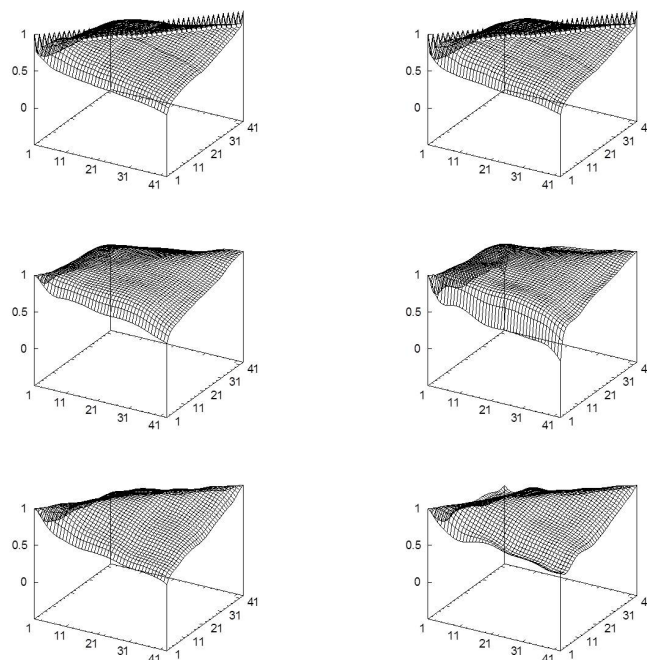
Figure 5. Phenotypic (1st row), genetic (2nd row), and permanent environmental (3rd row) correlation estimates between weekly milk flow obtained with the BSQ77 and BSC88 models.

The variance components and genetic parameters estimated with models LEG36_7, LEG39_7, and LEG310_7 were similar. However, the log L, AIC, and BIC values differed between these models, favoring those that contained a larger number of parameters. Since little change was observed in the genetic parameters obtained by these models, LEG36_7 could be used to evaluate milk flow without any loss of precision, even with a reduction in the number of parameters being estimated. Difficulties in the selection of the most adequate models to describe variations in growth and milk yield traits were also reported by other researchers (Jaffrézic et al., 2002; Bignardi et al., 2009a). These authors suggested selecting models that more accurately predict breeding values, or that promote greater responses to selection (Jaffrézic et al., 2002).

The rank correlations (Spearman) between sire breeding values estimated by models LEG36_7 and LEG39_7, LEG36_7 and LEG310_7, and LEG39_7 and LEG310_7 were 0.98, 0.97, and 0.99, respectively. These correlations indicate that the application of any of these models would practically lead to the selection of the same animals. This finding may be explained by the similar variance components and genetic parameters obtained by these models. When selecting 10% (20) of the best sires based on the breeding values estimated by model LEG36_7, 95% of the animals would be the same if selection were based on breeding values obtained by models LEG39_7 and LEG310_7. However, when selecting 10% of the best sires based on the breeding values estimated by model LEG39_7, 100% of the animals would be the same if selection were based on breeding values obtained by model LEG310_7. This observation indicates that selection of animals with superior breeding values according to the models described here would not lead to important changes in animal ranking. Therefore, considering the existence of coincidence in the selection process (i.e., animals selected by one criterion would also be selected

by another criterion), it would be more advantageous to select the most parsimonious model.

In addition to the variance components and genetic parameters, the rank correlation (Spearman) between sire breeding values obtained by models BSQ88 and BSC77 was highly similar. This correlation was 0.99, indicating that the application of either model would lead to the selection of the same animals. Based on the breeding values predicted by models BSQ77 and BSC88, all animals would be the same when selecting the best 10% (16) of sires. Thus, considering the existence of coincidence in the selection process, the choice of the most parsimonious models (i.e., the model that uses a quadratic function) would be advantageous.

High-degree Legendre polynomials were not necessary to fit additive genetic effects in the present study. However, a sixth-order polynomial was required to adequately model the permanent environmental effect. High-degree polynomials increase computational requirements, with consequent difficulties in achieving the convergence of the iterative process for the estimation of variance components. Similarly, the model using a B-spline function that provided the best data fit also contained a large number of parameters requiring estimation. This issue often limits the application of this method to genetic evaluations that use large data sets. In addition, the present results showed that B-spline functions were not as adequate as orthogonal Legendre polynomials to model the milk flow curve and dispersion. These findings contradict those reported by White et al. (1999) and Meyer (2005) for milk yield and growth, respectively. These authors found that B-spline functions, rather than Legendre polynomials, were more adequate as base functions for RRM to describe the covariance structure of the data. The authors came to this conclusion because B-spline functions are less likely to produce erratic estimates at the extremes of the curve compared to Legendre polynomials.

Although RRM fitting B-spline functions are more parameterized compared to models using Legendre polynomials, they achieve faster convergence in the estimation of variance components because they consist of segments of low-degree (quadratic or cubic) polynomials. This phenomenon favors analyses that contain a large volume of data. In addition, the various segments that make up spline curves confer greater flexibility to these models. As a consequence, RRMs using B-splines tend to be affected by the distribution of the data and, even more, by the choice of knots and degree of the spline function (Misztal, 2006). Since these models were not found to be adequate to explain the variation in milk flow in the present study, other models should be tested; for example, models including different numbers of knots that are not equidistant and models with different degrees of these B-spline functions. According to White et al. (1999), the correct fixation of the number and position of knots results in models that provide a good fit, with computational analysis becoming viable. Misztal (2006) described 2 alternatives for the choice of the number and position of knots in random regression analyses using linear splines. The first possibility is to start the analysis with a large number of equidistant knots, and then to eliminate highly correlated knots. Another possibility would be to start the analysis with a few knots, to estimate variance, and finally add other knots in such a way that the correlation between adjacent knots would be within the range of 0.6 to 0.8.

It might be advantageous to use an RRM for milk flow genetic evaluations, rather than a single trait model. This is because the former produced moderate to high heritability estimates, leading to more accurate breeding value predictions. In a simulation study, Meyer (2004) showed that breeding values were 5.9% more accurate using a RRM compared to a multi-trait model. Moreover, it is not necessary to define class ranges when using RRM, which potentially results in records being eliminated. Hence, the use of such models could prevent considerable loss of data.

Gäde et al. (2007) suggested that milkability should be included in breeding deci-

sions, because this parameter is related to labor efficiency and udder health. Greater milk flow is associated with a decrease in labor time, and thus a decrease in expense. In addition, faster milking implies a reduction in electricity costs and machine maintenance. Prins et al. (2002) estimated economic value for milking times ranges from $2.14 to $34.02/minute/cow/year in Germany, depending on the size of the milking parlor.

Typically, a group of cows enters the milking parlor at the same time, with all cows in the group remaining in the parlor until the last cow has finished milking. Thus, the slowest cow in the group sets the pace. Milk flow has an intermediate optimum; hence, there is value in reducing variation around that optimum, but it is not easily accomplished. To obtain a more uniform milking duration among the cows, and to simultaneously avoid deterioration in udder health, cows with low or high milk flow should be discarded from breeding group (Gäde et al., 2007).

CONCLUSIONS

Heterogeneity in residual variance should be considered when modeling milk flow. The model containing 7 residual classes is the most adequate for describing the trajectory of this effect across lactation. The model using a third-order Legendre polynomial for additive genetic effects and a sixth-order polynomial for permanent environmental effects is the most adequate for describing variation in milk flow and is also the most parsimonious. The heritability of milk flow estimated by the most parsimonious model was of moderate to high magnitude, indicating that selection for this trait would result in rapid genetic gain. The genetic correlation between milk flow on different test days estimated by the same model was positive, and indicated that selection for increased milk flow on a certain test-day would imply an increase in milk flow on other test-days.

ACKNOWLEDGMENTS

REFERENCES

Araújo CV, Torres RA, Costa CN, Torres Filho RA, et al. (2006). Random regressions models to describe the genetic variation of milk yield in Holstein breed. *Rev. Bras. Zootec.* 35: 975-981.

Bignardi AB, El Faro L, Albuquerque LG, Cardoso VL, et al. (2009a). Random regression models to estimate test-day milk yield genetic parameters Holstein cows in southeastern Brazil. *Livest. Sci.* 123: 1-7.

Bignardi AB, El Faro L, Cardoso VL, Machado PF, et al. (2009b). Parametric correlation functions to model the structure of permanent environmental (co)variances in milk yield random regression models. *J. Dairy Sci.* 92: 4634-4640.

Bignardi AB, El FL, Torres Junior RA, Cardoso VL, et al. (2011). Random regression models using different functions to model test-day milk yield of Brazilian Holstein cows. *Genet. Mol. Res.* 10: 3565-3575.

Boettcher PJ, Dekkers JC and Kolstad BW (1998). Development of an udder health index for sire selection based on somatic cell score, udder conformation, and milking speed. *J. Dairy Sci.* 81: 1157-1168.

Cobuci JA, Euclydes RF, Lopes PS, Costa CN, et al. (2005). Estimation of genetic parameters for test-day milk in Holstein cows using a random regression model. *Genet. Mol. Biol.* 28: 75-83.

Cobuci JA, Costa CN, Teixeira NM and Freitas AF (2006). Use of Legendre polynomials and Wilmink function in genetic evaluations for persistency of lactation in Holstein cows. *Arq. Bras. Med. Vet. Zootec.* 58: 614-623.

Costa CN, Melo CNR, Pacher IU, Freitas AF, et al. (2008). Genetic parameters for test-day milk yield of first lactation Holstein cows estimated by random regression using Legendre polynomials. *Rev. Bras. Zootec.* 37: 602-608.

Dodenhoff J and Emmerling R (2009). Genetic parameters for milkability from the first three lactations in Fleckvieh cows. *Animal* 3: 329-335.

Eilers PHC and Marx BD (2010). Splines, Knots and Penalties. Available at [http://www.stat.lsu.edu/faculty/marx/Skip.pdf]. Accessed March 30, 2013.

El Faro L, Cardoso VL and Albuquerque LG (2008). Variance component estimates applying random regression models for test-day milk yield in Caracu heifers (*Bos taurus* Artiodactyla, Bovidae). *Genet. Mol. Biol.* 31: 665-673.

Erf DF, Hansen LB and Lawstuen DA (1992). Inheritance and relationships of workability traits and yield for Holsteins. *J. Dairy Sci.* 75: 1999-2007.

Gäde S, Stamer E, Bennewitz J, Junge W, et al. (2007). Genetic parameters for serial, automatically recorded milkability and its relationship to udder health in dairy cattle. *Animal* 1: 787-796.

Gray KA, Vacirca F, Bagnato A, Samoré AB, et al. (2011). Genetic evaluations for measures of the milk-flow curve in the Italian brown Swiss population. *J. Dairy Sci.* 94: 960-970.

Gray KA, Cassady JP, Huang Y and Maltecca C (2012). Effectiveness of genomic prediction on milk flow traits in dairy cattle. *Genet. Sel. Evol.* 44: 24.

Ilahi H and Kadarmideen HN (2004). Bayesian segregation analysis of milk flow in Swiss dairy cattle using Gibbs sampling. *Genet. Sel. Evol.* 36: 563-576.

Jaffrézic F, White IM, Thompson R and Visscher PM (2002). Contrasting models for lactation curve analysis. *J. Dairy Sci.* 85: 968-975.

Karacaören B, Jaffrézic F and Kadarmideen HN (2006). Genetic parameters for functional traits in dairy cattle from daily random regression models. *J. Dairy Sci.* 89: 791-798.

König S, Köhn F, Kuwan K, Simianer H, et al. (2006). Use of repeated measures analysis for evaluation of genetic background of dairy cattle behavior in automatic milking systems. *J. Dairy Sci.* 89: 3636-3644.

Laureano MM, Bignardi AB, El FL, Cardoso VL, et al. (2012). Genetic parameters for first lactation test-day milk flow in Holstein cows. *Animal* 6: 31-35.

Luttinen A and Juga J (1997). Genetic Relationships Between Milk Yield, Somatic Cell Count, Mastitis, Milkability and Leakage in Finnish Dairy Cattle Population. Proceedings of the INTERBULL Meeting, Uppsala, 78-83.

Meyer K (2004). Scope for a random regression model in genetic evaluation of beef cattle for growth. *Livest. Prod. Sci.* 86: 69-83.

Meyer K (2005). Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genet. Sel. Evol.* 37: 473-500.

Meyer K (2007). WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B* 8: 815-821.

Misztal I (2006). Properties of random regression models using linear splines. *J. Anim. Breed. Genet.* 123: 74-80.

Olori VE, Hill WG, McGuirk BJ and Brotherstone S (1999). Estimating variance components for test-day milk records by restricted maximum likelihood with a random regression animal model. *Livest. Prod. Sci.* 61: 53-63.

Prins D, Groen AF and Saatkamp H (2002). Economic Value of Milkability in Dairy Cattle. Master's thesis, Wageningen University, Wageningen.

Rekaya R, Carabano MJN and Toro MA (1994). Random Regression in Animal Models for Test-Day Production in Dairy Cattle. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Guelph, 443-446.

Rensing S and Ruten W (2005). Genetic Evaluation for Milking Speed in German Holstein Population Using Different Traits in a Multiple Trait Repeatability Model. Proceedings of the INTERBULL Meeting, Uppsala, 167-170.

Rice JA and Wu CO (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* 57: 253-259.

Rupp R and Boichard D (1999). Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. *J. Dairy Sci.* 82: 2198-2204.

Samoré AB, Roman-Ponce SI, Vacirca F, Frigo E, et al. (2011). Bimodality and the genetics of milk flow traits in the Italian Holstein-Friesian breed. *J. Dairy Sci.* 94: 4081-4089.

Sewalem A, Miglior F and Kistemaker GJ (2011). Short communication: Genetic parameters of milking temperament and milking speed in Canadian Holsteins. *J. Dairy Sci.* 94: 512-516.

Vicario D, Carnier P and Degano L (2006). Test-Day Model for National Genetic Evaluation of Somatic Cell Count in Italian Simmental Population. Proceedings of the INTERBULL Meeting, Uppsala, 171-175.

White IM, Thompson R and Brotherstone S (1999). Genetic and environmental smoothing of lactation curves with cubic splines. *J .Dairy Sci.* 82: 632-638.

Wolfinger R (1993). Covariance structure selection in general mixed models. *Commun. Stat.* 22: 1079-1106.

Zwald NR, Weigel KA, Chang YM, Welper RD, et al. (2005). Genetic evaluation of dairy sires for milking duration using electronically recorded milking times of their daughters. *J. Dairy Sci.* 88: 1192-1198.