Canadian Journal of Animal Science
Revue canadienne de science animale

Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines

Journal:	<i>Canadian Journal of Animal Science</i>
Manuscript ID	CJAS-2017-0019.R1
Manuscript Type:	Article
Date Submitted by the Author:	15-Jun-2017
Complete List of Authors:	Brito, Luiz; University of Guelph, Department of Animal Biosciences Gomes da Silva, Felipe Rojas de Oliveira, Hinayah Souza, Nadson Caetano, Giovani Costa, Edson Romeiro de Oliveira Menezes, Gilberto Puerro de Melo, Ana Lúcia Teixeira Rodrigues, Marcelo de Almeida Torres, Robledo
Keywords:	genetic evaluation, Alpine, milk yield, Segmented polynomials, test-day

SCHOLARONE™
Manuscripts

Running head: Modelling lactation curves in goats

**Modelling lactation curves of dairy goats by fitting random regression models
using Legendre polynomials or B-splines**

**Brito, L.F.^{1,*}, Silva, F.G.², Oliveira, H.R.^{1,3}, Souza, N.O.³, Caetano, G.C.³, Costa,
E.V.³, Menezes, G.R.O.⁴, Melo, A.L.P.⁵, Rodrigues, M.T.³, Torres, R.A.³**

¹Centre for Genetic Improvement of Livestock, Department of Animal Biosciences,
University of Guelph, Guelph, Ontario, Canada, N1G 2W1.

²Departamento de Zootecnia e Extensão Rural, Universidade Federal de Mato Grosso,
Cuiabá, Mato Grosso, Brazil, 78060-900.

³Departamento de Zootecnia, Universidade Federal de Viçosa, Viçosa, Minas Gerais,
Brazil, 36570-000.

⁴Embrapa Gado de Corte, Campo Grande, Mato Grosso do Sul, Brazil, 79106-550.

⁵Instituto de Zootecnia, Universidade Federal Rural do Rio de Janeiro, Seropédica, Rio
de Janeiro, Brazil, 23890-000.

***Corresponding author:** Luiz Fernando Brito

E-mail address: lbrito@uoguelph.ca

Postal address: Centre for Genetic Improvement of Livestock, Department of Animal
Biosciences, University of Guelph, Guelph, Ontario, Canada, N1G 2W1.

ABSTRACT

A total of 17 356 test-day milk yield (TDMY) records from 642 first lactations of Alpine goats were used to model variations in lactation curve using random regression models (RRM). Orthogonal Legendre polynomials and B-splines were evaluated in order to obtain adequate and parsimonious models for the estimation of genetic parameters. The analysis were performed using a single-trait RRM, including the additive genetic, permanent environmental and residual effects. We estimated the mean trend of milk yield, and the additive genetic and permanent environmental covariance functions through random regression using different orders of Orthogonal Legendre polynomial (three to six) and B-spline functions (linear, quadratic and cubic, with three to six knots). This study further evaluated different number of classes of residual variances. The covariance components and the genetic parameters were estimated using the Restricted Maximum Likelihood method. Heritability estimates presented similar trends for both functions. RRM with a higher number of parameters better described the genetic variation of TDMY throughout the lactation. The most suitable RRM for genetic evaluation of TDMY of Alpine goats is a quadratic B-spline function with six knots, for the mean trend, curves of additive genetic and permanent environmental effects and five classes of residual variance.

Key words: genetic evaluation, Alpine, milk yield, test-day, segmented polynomials.

INTRODUCTION

The world goat population is around 1.1 billion animals (FAOSTAT, <http://www.fao.org/faostat/>) and specially the dairy goat industry is under great expansion (Liang and Devendra 2014). In order to continue expanding and remain

competitive with other dairy industries, goat producers need to improve the production efficiency of their herds. Genetic improvement is the most attractive approach to promote permanent gains in livestock species. Milk production is among traits with highest economic relevance for the industry and is a great example of longitudinal traits, characterized by repeated measures in the same individual over time. For the analysis of longitudinal data, random regression models (RRM) have been widely used in various livestock species to analyse a variety of traits (Flores and Werf 2015; Kheirabadi and Rashidi 2016; Rovadoscki et al. 2016; Oliveira et al. 2016; Sasaki et al. 2015; Englishby et al. 2016; Pereira et al. 2013). In RRM, a fixed curve for the population is calculated and individual curves are fitted as deviation from the population curve. RRM require the use of a function to model the fixed and random curves. Orthogonal Legendre polynomials (LEG) are among the most popular ones (Bignardi et al. 2011; Meyer 2005; Sarmiento et al. 2008; Kheirabadi et al. 2014). However, these models may require higher polynomial degrees and uneven adjustment throughout the lactation curve when compared to other polynomials (Meyer 2005). In addition, LEG may provide implausible parameter estimates for the beginning and the end of the trajectory by overemphasizing observations at the extremes (Meyer 2005). Considering that, B-spline (segmented polynomials) functions have been suggested as an alternative to high degree LEG (Silvestre et al. 2006; DeGroot et al. 2007; Bohmanova et al. 2008; Bignardi et al. 2011; Sánchez et al. 2008; Huisman et al. 2002). However, studies investigating the use of B-splines are still scarce for dairy goats and therefore, there is a need to investigate these models for modelling lactation curves of dairy goats.

According to Bohmanova et al. (2008), different functions can be used to fit random and fixed regression curves. The first applications used parametric functions and functions of lactation shape such as Ali and Schaeffer (Ali and Schaeffer 1987) and

Wilmink (Wilmink 1987). However, these functions have been replaced with LEG and more recently with B-spline functions. According to Meyer (2005), the use of splines is interesting when a low polynomial degree does not fit well the data. The choice of an appropriate statistical model will allow the application of RRM to genetic evaluations of dairy goats aiming to improve milk yield over the whole lactation. Therefore, the objective of this study was to investigate the variation in test-day milk production in dairy goats fitting RRM using either Orthogonal Legendre polynomials or B-spline functions.

MATERIALS AND METHODS

Animal care

Animal welfare and use committee approval was not needed for this study as datasets were obtained from pre-existing databases.

Data

A total of 17 356 Test-Day Milk Yield (TDMY) records from 642 first lactations of Alpine goats, from a single herd owned by the Universidade Federal de Viçosa (Viçosa, Minas Gerais, Brazil) were used for this study. Animals were kept in collective stalls under a free stall system and were fed diets based on corn silage and hay, as roughage, and concentrate mixture, available according to the nutritional needs of the animals. Milk control was performed weekly and milking was carried out twice daily. TDMY is the sum of the milk collected in these two periods.

The data editing and consistency analysis were performed using SAS software (SAS Institute 2014). Only the following records were kept for further analysis: first lactation of does that had their first kidding between 10 and 30 months of age, lactation

length longer than 100 days; type of kidding (one to three kids), days in milk greater than five and control intervals shorter than 28 days. Lactations were truncated at 270 days. Pedigree was recorded and the numerator relationship matrix coefficients contained 4 767 animals.

105

Statistical analysis

After data editing, the influence of the following fixed effects on TDMY was investigated: year and season of kidding (years = 2000 to 2011; and season 1, from March to September, and season 2 from October to February), genetic grouping (breed composition), type of kidding (one to three kids) and age of goat at kidding as a covariate, considering linear and quadratic effects. Only animals with the genetic composition of Alpine breed equal or greater than 50% were kept for further analysis. They were grouped in classes as: 1 = 50% of Alpine breed; 2 = greater than 50% and equal to or less than 75%; 3 = greater than 75% and equal to or less than 93%, and 4 = greater than 93%.

To check the influence of environmental effects, analyses of variance for unbalanced data was performed through the Generalized Least Square (GLS) method using the procedure GLM of SAS software (SAS Institute 2014). The model used to explain variation of the traits due to known environmental effects was:

$$y_{ijkl} = \mu + T_i + G_j + YS_k + b_1(P_{ijkl} - \bar{P}) + b_2(P_{ijkl} - \bar{P})^2 + \varepsilon_{ijkl}$$

where, y_{ijkl} is the observation of the random animal l , with type of kidding i , genetic grouping j and year-season k ; μ is the overall mean; T_i is the fixed effect of type of kidding i ; G_j is the fixed effect of genetic grouping j ; YS_k is the fixed effect of year-season k ; b_1 and b_2 are the linear and quadratic coefficients of regression of the analyzed trait in terms of age of dam at kidding (ADK, included as covariate), respectively; P_{ijkl}

125 is the age of the dam l at kidding; $\bar{P}$ is the overall mean of ADK; and, ϵ_{ijkl} is the
 126 residual term associated to each observation y_{ijkl} .

127

128 ***Orthogonal Legendre polynomials***

129 The single-trait animal model for random regression with LEG is described in a
 130 generalized manner as:

$$y_{ijk} = FE_k + \sum_{m=0}^{k_b-1} b_m \phi_m(t_{ij}) + \sum_{m=0}^{k_\alpha-1} \alpha_{im} \phi_m(t_{ij}) + \sum_{m=0}^{k_\gamma-1} \gamma_{im} \phi_m(t_{ij}) + \epsilon_{ij}$$

131 where, y_{ijk} is the milk production on the j control of the i^{th} goat within fixed effects k ;
 132 FE_k are the fixed effects, which consisted of genetic grouping (breed composition),
 133 year-season of kidding (contemporary group), and type of kidding. ADK was included
 134 as a co-variable with linear and quadratic effects; b_m is the m^{th} regression coefficient of
 135 the TDMY on the LEG for modeling the average curve of the population; α_{im} and γ_{im}
 136 are the m^{th} regression coefficient of the additive genetic and permanent environment
 137 effects, respectively, for the i^{th} goat; k_b , k_α , and k_γ are the degree of LEG; t_{ij} is the
 138 control variable, lactation week of the i^{th} goat standardized for the -1 to 1 interval, as
 139 described by Kirkpatrick et al. (1990); $\phi_m(t_{ij})$ is the LEG function for parameter m
 140 evaluated for age t_{ij} ; and, ϵ_{ij} is the specific effect for each observation that was not
 141 explained by the correction factors or the regression of the model. The tested models
 142 had combinations of distinct adjustment orders for polynomials of fixed, random
 143 genetic, and permanent environmental curves, besides a number of classes for residual
 144 variance. The first models evaluated ranged from third to sixth order for curves: fixed,
 145 additive genetic and permanent environment, considering homogeneity of variances.
 146 Only functions of the same order of adjustment were considered to allow possibilities of
 147 equal variations of both components, as recommended by López-Romero et al. (2004).

After choosing the most suitable model considering homogeneity of residual variance, a total of five different models considering heterogeneity of residual variance were investigated in this second step: two (6th to 160th and 161th to 270th days in milk), three (from 6th to 60th, from 61th to 150th and 151th to 270th days in milk), four (6th to 50th, 51th to 100th, 101th to 200th and 201th to 270th days in milk), five (from 6th to 30th, 31th to 90th, 91th to 140th, 141th to 200th and 201th to 270th days in milk) and six (6th to 30th, 31th to 90th, 91th and 120th, 121th to 160th, 161th to 210th and 211th to 270th days in milk) classes. The classes of residual variances were formed based on a graphical inspection of TDMY records.

The notation of the models fitted using LEG follows the pattern: LEG kHx , with k referring to the order of the polynomial curve to model the population mean, the additive genetic effects and permanent environmental and x is the number of classes of residual variances. For instance, a model with regression of mean curve, additive genetic and permanent environment of order 3 and that had four classes of heterogeneous residual variance was coded as LEG3H4.

B-splines (segmented polynomials)

Analysis were performed using single-trait random regression models. For all models the random effects, additive genetic and permanent environmental, were included in the analysis. The fixed effects included in the analysis were: year-season of kidding (contemporary group), genetic grouping (breed composition), type of kidding and the linear and quadratic effects of the covariate age of dam at kidding. The matrix form of the single-trait animal model using random regression was given by: $y = Xb + Za + Wc + e$, where, y is the vector of observations, b is the vector of fixed effects and coefficients of fixed regression, a is the vector of coefficients of additive genetic

173 regression, $\mathbf{c}$ is the vector of coefficients of random permanent environment regression,
 174 $\mathbf{e}$ is the vector of random residual effect; and $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{W}$ are the incidence matrices of
 175 fixed, additive genetic and permanent environmental effects, respectively.

176 The assumptions regarding to the components of the model are:

$$177 \quad E \begin{bmatrix} y \\ a \\ c \\ e \end{bmatrix} = \begin{bmatrix} \mathbf{X}b \\ 0 \\ 0 \\ 0 \end{bmatrix}, V(\mathbf{a}) = \mathbf{K}_A \otimes \mathbf{A}; V(\mathbf{c}) = \mathbf{K}_C \otimes \mathbf{I}_{Nd} \text{ and, } V(\mathbf{e}) = \mathbf{R}, \text{ where: } \mathbf{K}_A \text{ and } \mathbf{K}_C \text{ are}$$

178 the matrices of (co)variance between the regression coefficients random additive genetic
 179 and permanent environmental effects, respectively. $\mathbf{A}$ is the numerator relationship
 180 matrix, $\mathbf{I}_{Nd}$ is an identity matrix, and, $\mathbf{R}$ represents a matrix containing residual
 181 variances.

182 Additive genetic, permanent environmental and average lactation curves were
 183 modeled by B-spline functions (linear, quadratic and cubic), with equidistant knots. As
 184 suggested by Meyer (2005), the B-spline of degree $p = 0$ have values equal to unity for
 185 all points (t) at a given interval, and zero otherwise. For the k^{th} segment given for knots
 186 T_k and T_{k+1} with $T_k \leq T_{k+1}$:

$$B_{k,0}(t) = \begin{cases} 1, \text{ if } T_k \leq t < T_{k+1}; \\ 0, \text{ otherwise.} \end{cases}$$

187 Functions with degree $B_{k,p}$, for $p > 0$, can be represented by:

$$B_{k,p}(t) = \frac{t - T_k}{T_{k+p} - T_k} \cdot B_{k,p-1}(t) + \frac{T_{k+p+1} - t}{T_{k+p+1} - T_{k+1}} \cdot B_{k+1,p-1}(t)$$

188 The individual segments have been considered as linear (L), quadratic (Q) or
 189 cubic (C) order polynomials, with base function of degree $p = 1, 2$ and 3 , joined at
 190 knots, which allows the function to become continuous. Polynomials of same degree
 191 were considered in the model for the fixed curve and for additive genetic and permanent
 192 environment effects.

First, we evaluated the B-spline functions for all curves considering the homogeneity of residual variances. In possession of the most suitable model under homogeneity of residual variances, two forms of residual variance structures were evaluated: residual variance constant throughout lactation or residual variance constant within lactation, but varying between classes of days in milk. If heterogeneous, we considered the same classes as specified for LEG. The m knots were chosen to divide weeks of lactation in $m - 1$ equidistant intervals and external knots were located on the 6th and 270th days in milk for all models. The same number of knots was adjusted for all effects. The number of random regression coefficients for modeling the trajectory of linear, quadratic and cubic bases functions were given by m , $m + 1$ and $m + 2$, respectively. The extremity knots do not interfere in the number of parameters of the model.

The B-spline RRM_s were coded as: "BSX k H y ", where X = L, Q or C, are the degrees of the linear, quadratic or cubic polynomial segments, respectively; k is the number of knots of the fixed curve, additive genetic and permanent environmental effects, and y is the number of classes of residual variances. Therefore, as illustration, a model with the following representation BSL3H2 is a linear B-spline function, considering two classes of residual variances, three knots for all effects, and two segments ($k - 1$).

Variance components, model comparison and genetic parameter estimates

Variance components were estimated via Restricted Maximum Likelihood (REML) approach, using WOMBAT software (Meyer 2007). The results obtained when fitting RRM using B-spline functions were also compared with the best model obtained using LEG. Criteria for choosing models in this study were: Logarithm of the Restricted

Maximum Likelihood function (LogL), Akaike's Information Criterion (AIC) (Akaike 1998) and Bayesian Information Criterion (BIC) (Schwarz 1978). The information criteria AIC and BIC can be described as $AIC = -2\log L + 2p$ and $BIC = -2\log L + p\log(N - r)$, where p 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. Lower values for AIC and BIC and greater values for LogL indicate better model fit. For comparison between full and reduced models we also used the Likelihood Ratio Test (LRT, Wolfinger 1993), at 1% of probability.

The components of (co)variance and genetic parameters estimated over the lactation curve obtained by the models that provided better adjustments were also investigated. In addition, monitoring was done in order to verify the occurrence of convergence problems or estimates that were not consistent with the biology of the trait.

RESULTS AND DISCUSSION

Analysis of fixed effects

All fixed effects (genetic grouping, year-season of kidding, age of dam at kidding with linear and quadratic effect, and type of kidding) were significant at 1% level of significance and therefore were included in the final model for the study of TDMY. The average number of observations per test-day was 456 ± 91 . Figure 1 shows the TDMY as a function of the lactation week. Milk production varied accordingly to a standard lactation curve for specialized dairy goat breeds, starting with 2.16 kg in the first milk control, increasing to 2.66 kg in the tenth week after kidding and gradually reducing in subsequent milk recordings until 1.67 in the 39th week of lactation.

Fit and model comparison

243 *Orthogonal Legendre polynomials.* Table 1 shows the random regression
 244 models fitted by LEG considering homogeneity of residual variances and model
 245 comparison. It was observed that values of LogL, AIC and BIC improved with higher
 246 polynomial orders (Table 1). In accordance with LRT, there were significant differences
 247 ($P < 0.01$) between the full and reduced models. In general, higher values for the LogL
 248 criteria will always be expected for the more parameterized models because they are
 249 more flexible. However, it is important to use other comparison criteria that provide a
 250 choice of more parsimonious models, such as the AIC and BIC criteria. The LEG6H1
 251 model was the one that best fitted the data according to all criteria. Nevertheless, it is
 252 important to evaluate the need to consider heterogeneity of residual variances
 253 throughout lactation, as there may be improvements in the partition of the total variation
 254 and consequently, generate better estimates of genetic parameters (Olori et al. 1999).

255 For greater number of classes there are more parameters to be estimated in the
 256 maximization of the likelihood function. More parameterized models may have bigger
 257 problems of convergence and estimation. It can hamper the use of RRM for genetic
 258 evaluations using large datasets. Therefore, it is important to use more parsimonious
 259 models without losing quality of fit of the random effects included in the model. The
 260 results of the models considering heterogeneity of residual variances are shown in Table
 261 1. The model that considered homogeneity of residual variance attained the worst fit,
 262 which is in agreement with other studies (Menezes et al. 2011; Silva et al. 2013; Breda
 263 et al. 2006; Assis et al. 2006). The residual variance is influenced by many factors
 264 which may vary throughout lactation. For instance, the effect of scale of production that
 265 is greater in the early stage of lactation compared to the lactation end. This can be
 266 verified by the result of the division into only two classes of residual variances, which
 267 has provided a large reduction in the values of AIC and BIC, indicating better fit. In

addition, the increase in number of classes considerably improved the modeling of random effects included in the analysis. These results are in agreement with those presented by Breda et al. (2006) and Silva et al. (2013) for Alpine goats, where a similar trend of fit improvement with increased number of residual variance classes was reported. The LEG6H6 model best fit based on AIC and LogL. However, it presented a slightly smaller BIC value compared to LEG6H5. This can be explained by the fact that BIC criterion applies a greater penalty to more parameterized models (Vrieze, 2012), and LEG6H6 has one more parameter than the previous model. Furthermore, no significant differences based on the LRT were observed between these two models. As a greater number of parameters generates greater computational demand, more issues with convergence of algorithms can be expected for higher parameterized models. Therefore, the model that considered five classes of residual variances (LEG6H5) seemed to be the most suitable model.

Sarmiento et al. (2006) also working with data from Alpine goats, indicated the model of fifth order for the additive genetic effect and seventh order for the permanent environmental effect as the best option. However, they suggested that the use of a lower parameterized model of fourth order for the additive genetic effect and sixth order for the permanent environmental effects would give similar results. Silva et al. (2013) investigating Orthogonal Legendre polynomials reported that the most suitable model for genetic evaluation of TDMY in dairy goats was the one that considered the fixed curve of order 4, the curve of additive genetic effects of order 2, the curve of permanent environmental effects of order 7 and at least five classes of residual variances.

After choosing the best model using LEG, the variance components and heritability for all points of lactation and genetic correlations between milk yield on days six, 39, 71, 104, 137, 170, 203, 236 and 270 were investigated. The highest values

293 of additive genetic and residual variance were observed in the stage of production peak,
294 which occurred around 40 days after kidding (Figures 2a and 2b). A possible
295 explanation could be the effect of production scale, since at that stage it was observed
296 the largest milk yields of the lactation. These results differ from those found by Silva et
297 al. (2013), that observed increasing genetic variance along the lactation curve.

298 The phenotypic and permanent environmental variances were lower in early
299 lactation and increased at the end of lactation compared to the other lactation stages.
300 This may be due to the smaller number of records at the beginning and end of lactation,
301 which does not represent major problems for the prediction of breeding values as RRM
302 were used. However, missing data in the beginning or end of lactation will require
303 extrapolations to predict values associated to permanent environmental effects of
304 animals in those lactation stages. As shown in Figures 2a and 2b, the estimates of
305 variances in the final period seemed worse than the estimates of the onset of lactation,
306 which was expected as there was a greater reduction in the number of records at the end
307 of the lactation. Menezes et al. (2011) also found higher estimates of permanent
308 environmental effects at the end of lactation. Heritability estimates ranged from 0.15 to
309 0.40 throughout the lactation (Figure 3). It was observed that the heritability estimates
310 were lower in early and late lactation. This was possibly due to the greater influence of
311 environmental effects on these stages and/or proportionally lower genetic additive
312 variance on this stages. For instance, in early lactation, first parity goats are usually not
313 well adapted to the milking process and they have a deep postpartum stress due to
314 negative energy balance. On the other hand, in late lactation many of them are in
315 advanced stage of pregnancy, which can justify the decrease at this stage. Furthermore,
316 as already mentioned, in these two extremes there was a greater amount of missing
317 records, which can also cause errors in the estimation of variance components. Similar

318 results were reported by Sarmento et al. (2008). However, Silva et al. (2013) found
319 higher estimates in the final third of lactation and Menezes et al. (2011) found higher
320 values at the beginning of the lactation.

321 Table 2 presents the estimates of genetic correlations obtained by LEG6H5
322 model between milk production on days six, 39, 71, 104, 137, 170, 203, 236 and 270.
323 Genetic correlations in this study, obtained by LEG6H5 model, ranged from 0.21 to
324 0.95. Silva et al. (2013) found estimates of genetic correlations ranging from 0.98 to
325 1.00. Genetic correlations were smaller as the intervals between TDMYs increased. This
326 result was consistent with Oliveira et al. (2016), who also found higher estimates of
327 genetic correlations for shorter TDMY intervals in Alpine and Saanen goats. One
328 possible explanation is that gene groups and the expression of genes related to milk
329 production varies throughout lactation, and this expression is more similar to adjacent
330 TDMY. However, other studies in this field are still needed (Strucken et al. 2015).
331 Despite the smaller genetic correlations between more distant TDMYs, all estimates
332 were positive, indicating that the selection of animals for milk production on any given
333 day of lactation will bring positive responses to all other points on the curve. However,
334 for the points where the genetic correlations were small (even if favorable), it is
335 important to include additional curve points in the selection index in order to make
336 progress in the whole pattern of the lactation curve.

337 **B-splines.** The models that were tested considering homogeneity of residual
338 variances are shown in Table 3. The WOMBAT software (Meyer 2007) does not allow
339 evaluation of models BSC3 (three knots and cubic order), as it is expected to have at
340 least one parameter influencing the dependent variable in all three segments. Therefore,
341 at least four knots are needed to be evaluated for cubic order B-splines. Regardless of
342 the B-spline functions used (linear, quadratic or cubic), when we increased the number

of knots, AIC and the BIC showed higher absolute values indicating better fit (Table 3). Considering only the number of knots, the values of AIC and BIC were always higher in models for second-degree polynomial (quadratic), indicating that these polynomials fitted better to the data. Among B-spline functions considering homogeneity of residual variances, the model that best fit the data was BSQ6H1. However, the criteria LRT does not allow comparisons between the BSQ6H1 and the other B-spline models with one class of residual variance (BSL6H1 and BSC6H1). This happened because there is no sufficient degrees of freedom for this comparison, since the number of parameters in the three models was the same (i.e., 43). Thus, different numbers of classes for the residual variances of B-splines of degrees 1, 2 and 3 (BSL6, BSQ6 and BSC6) were also investigated. Among the models with the same number of parameters, but with different degrees of the B-spline function (linear, quadratic and cubic), those using a quadratic B-spline fitted better in accordance with the adopted comparison criteria (Table 4). Moreover, for all polynomial orders evaluated, the most parsimonious B-spline model that better fitted the data had five residual variance classes, similarly to the best model adjusted for Orthogonal Legendre polynomials (i.e., LEG6H5). These findings indicate that five is actually the most suitable number of residual variance classes to be used in analyzes of TDMY for Alpine goats. The most appropriate division of classes of residual variance also influence the number of classes to be used.

Among all the models evaluated using B-splines, BSQ6H5 was the one that better fitted the data according to the comparison criteria used. The fact that the model BSQ6H5 have fitted the data better than the BSL6H5 and BSC6H5 models may be due to how each of them considers the points that influence the parameters of each knot, as shown by Meyer (2005). Models using linear B-splines ignore the estimation of the influence of a curve point in more distant points, while the cubic B-spline functions take

368 into consideration the influence of more distant points. The linear B-spline functions
369 require a smaller number of points (records) to estimate the coefficients and therefore,
370 have less multicollinearity issues. In the case of cubic B-spline functions, more points
371 are used to estimate the coefficients (Meyer, 2005). However, besides the higher
372 multicollinearity, late lactation records may influence the early lactation records and
373 vice-versa. This is not desirable, as genes that influence peak production, lactation
374 length and lactation persistence may not be the same (Strucken et al., 2015). In this
375 study, quadratic B-spline functions seem to have better balanced the advantages and
376 disadvantages of each order.

377 The results of Tables 1 and 4 show that the models that considered homogeneous
378 residual variances presented the worst adjustments to the data according to the
379 comparison criteria used. This highlights the need to consider a heterogeneous structure
380 of residual variances, since the variances tend to differ along the lactation. Similar
381 results were obtained by Bignardi et al. (2011), when evaluating models for adjustments
382 of lactation curves of Holstein cows and the authors concluded, according to the AIC
383 and BIC criteria, that models containing homogeneous residual variances provided
384 worst fits, regardless of the function used to model the random effects.

385 After choosing the best model using B-splines, variance components were
386 investigated (i.e. heritability for all points throughout the lactation and genetic
387 correlations between milk yield on days six, 39, 71, 104, 137, 170, 203, 236 and 270.
388 The largest variance estimates were observed for additive genetic and residual for the
389 production peak stage, which occurs around 40 days after birth (Figure 2b). It could be
390 due to a scale effect since the highest TDMYs were observed in that lactation stage. The
391 residual variance was greatest in the lactation peak and gradually reduced until the end
392 of the lactation. The phenotypic and permanent environment variances seemed to be

underestimated in the early stage of lactation and overestimated at the end. This may be due to the smaller number of records available at the beginning and especially at the end of lactation, as a consequence of the proximity of the goat dry-period and the fact that several goats are already in advanced pregnancy stage.

The random effect of permanent environment ranged along the lactation curve. This demonstrates that adjusting the coefficients for heterogeneity of residual variances allowed separating more efficiently the non-genetic term from the genetic variation, reflecting in more accurate heritability estimates. The heritability estimates were higher in late lactation, which is consistent with the dairy cattle literature (e.g. Cobuci et al. 2005; Costa et al. 2008). However, especially for dairy goats this pattern was not reported in the literature yet, since in late lactation a decrease in heritability have been observed in other studies (Assis et al., 2006; Oliveira et al., 2016; Sarmento et al., 2008). Heritability estimates obtained by BSQ6H5 model were smaller in the extremities of the lactation curve (Figure 3). This can be explained by the lower additive genetic variances in those stages. Table 2 presents the genetic correlations for TDMY between six, 39, 71, 104, 137, 170, 203, 236 and 270, obtained using BSQ6H5 model. Estimates of genetic correlations were positive and larger between adjacent TDMY and reduced to the extent that increased the interval between milking controls. For points where the genetic correlations have shown small magnitude, it is important to include additional curve points in the selection index in order to make progress in the whole lactation curve. These results were very similar to those observed for LEG6H5 model.

Comparison between models using orthogonal Legendre polynomials and B-splines

The best models using LEG or B-splines were compared in order to make a recommendation of the best model to be used in genetic evaluations of dairy goats. In

418 general, more parameterized models tend to be more penalized, as parsimonious models
419 are desirable (Vrieze, 2012). In this study, the two best models have the same number of
420 parameters. BSQ6H5 presented greater LogL value compared to LEG6H5 model,
421 indicating better fit. This was possibly due to the similar number of parameters in both
422 models. However, major multicollinearity issues are expected when using high degree
423 LEG. It may happen due to the fact that dependency problems of parameter estimates
424 from the beginning to the end of the regression are reduced when using B-splines.

425 The phenotypic, additive genetic, permanent environmental and residual
426 obtained by the two models (LEG6H5 and BSQ6H5) were very similar and showed the
427 same trend throughout lactation (Figures 2a and 2b). The number of residual variance
428 classes was the same for both models (five). Furthermore, both models yielded higher
429 estimates of the variance of permanent environmental effects in the end of the lactation
430 curve, possibly due to the smaller number of records. Meyer (2005) stated that RRM
431 using B-splines tend to be less susceptible to problems of erratic estimates at the
432 extremes of the curve, a problem often observed when using LEG.

433 The heritability estimates obtained by LEG6H5 model were higher throughout
434 almost the entire lactation when compared to BSQ6H5 model (Figure 3). However, it
435 showed the same trend of variation throughout lactation, indicating that periods with
436 higher heritability coincided for both models. Heritability estimates obtained by the two
437 models suggest that there was enough genetic variability to make genetic progress for
438 TDMY in dairy goats. According to LogL criteria, the model that best fitted the data
439 was BSQ6H5. In addition, this model contained individual segments of polynomials of
440 smaller degrees, which allows better control of the overall influence of individual
441 observations and tend to provide better genetic parameter estimates. In agreement with
442 our results, León et al. (2012) and Thepparat et al. (2015), working with Murciano-

443 Granadina and Thailand goats, respectively, reported better fits when using spline
444 functions. The findings obtained in this study indicates that there is enough genetic
445 variability to make genetic progress for test-day milk yield in dairy goats and it will be
446 of great relevance to the implementation of genetic evaluations in dairy goats.

447 **CONCLUSIONS**

448 For the genetic evaluation of dairy goats aiming to improve test-day milk yield,
449 we recommend to fit a random regression model using B-splines of degree 2 (quadratic)
450 for the curves: fixed of the population, additive genetic and permanent environment,
451 with five segments (six knots) and five classes of residual variances.

452 Heritability estimates obtained for the lactation curve were moderate to high,
453 indicating that there is enough genetic variability to make genetic progress for test-day
454 milk yield in dairy goats.

455
456
457
458
459
460
461
462
463
464
465
466
467

TABLES

Table 1. Models, number of parameters (p), LogL, AIC, BIC and LRT for Orthogonal Legendre polynomials.

Model ^a	p	LogL	AIC	BIC	LRT
LEG3H1	13	2,374.138	-4,722.276	-4,621.400	**
LEG4H1	21	2,714.039	-5,386.076	-5,223.122	**
LEG5H1	31	2,941.077	-5,820.154	-5,579.608	**
LEG6H1	43	3,050.944	-6,015.888	-5,682.228	Full model
LEG6H2	44	3,142.035	-6,196.070	-5,854.652	**
LEG6H3	45	3,155.987	-6,221.974	-5,872.796	**
LEG6H4	46	3,153.432	-6,214.864	-5,857.928	**
LEG6H5^b	47	3,179.123	-6,264.246	-5,899.548	ns
LEG6H6	48	3,182.000	-6,268.000	-5,895.544	Full model

** = $P < 0.01$.^aLEG x H y , where x is the polynomial order for the fixe curve and for the additive genetic and permanent environmental effects and y is the number of residual variance classes.^bValues in bold indicate the best model based on LogL, AIC, BIC and LRT criterions.

Table 2. Estimates of genetic correlations for test-day milk yield (TDMY) obtained for the model LEG6H5^a (above diagonal) and for BSQ6H5^b (below diagonal).

TDMY	6	39	71	104	137	170	203	236	270
6		0.97	0.95	0.88	0.84	0.76	0.71	0.43	0.25
39	0.98		0.97	0.90	0.88	0.80	0.74	0.45	0.29
71	0.96	0.97		0.95	0.94	0.84	0.75	0.51	0.21
104	0.90	0.92	0.96		0.95	0.87	0.84	0.49	0.33
137	0.86	0.89	0.94	0.96		0.92	0.88	0.54	0.40
170	0.77	0.79	0.84	0.91	0.93		0.91	0.63	0.51
203	0.68	0.72	0.74	0.83	0.87	0.89		0.84	0.78
236	0.43	0.45	0.47	0.56	0.61	0.71	0.76		0.83
270	0.19	0.21	0.18	0.33	0.35	0.70	0.75	0.83	

^aLegendre polynomial of order 6 and 5 classes of residual variance.
^bQuadratic B-spline model of polynomial order 6 and 5 classes of residual variance.

Table 3. Models, number of parameters (p), LogL, AIC, BIC and LRT for B-splines considering homogeneity of residual variances.

Model ^a	p	LogL	AIC	BIC	LRT
BSL3H1	13	2,201.291	-4,376.582	-4,275.706	**
BSL4H1	21	2,631.513	-5,221.026	-5,058.074	**
BSL5H1	31	2,876.156	-5,690.312	-5,449.766	**
BSL6H1^b	43	3,066.300	-6,046.600	-5,712.942	Full model
BSQ3H1	13	2,339.923	-4,653.846	-4,552.978	**
BSQ4H1	21	2,711.716	-5,381.432	-5,218.480	**
BSQ5H1	31	2,952.442	-5,842.884	-5,610.338	**
BSQ6H1^b	43	3,092.906	-6,099.812	-5,766.154	Full model
BSC4H1	21	2,717.819	-5,393.638	-5,230.686	**
BSC5H1	31	2,958.138	-5,852.276	-5,611.728	**
BSC6H1^b	43	3,079.741	-6,073.482	-5,739.824	Full model

** = $P < 0.01$.

^aBSXkHy, where X = L, Q and C represents the degree of linear, quadratic and cubic polynomials, respectively, k is the polynomial order for the fixed curve and for the additive genetic and permanent environmental effects and y is the number of residual variance classes.

^bValues in bold indicate the best model based on LogL, AIC, BIC and LRT criterions.

515
516
517
518

Table 4. Models, number of parameters (p), LogL, AIC, BIC and LRT B-splines considering heterogeneity of residual variances.

Model ^a	p	LogL	AIC	BIC	LRT
BSL6H1	43	3,066.300	-6,046.600	-5,712.942	**
BSL6H2	44	3,155.888	-6,223.776	-5,882.358	**
BSL6H3	45	3,192.276	-6,294.552	-5,945.374	**
BSL6H4	46	3,184.473	-6,276.946	-5,920.010	**
BSL6H5	47	3,209.369	-6,324.738	-5,960.042	ns ^c
BSL6H6	48	3,208.161	-6,320.322	-5,947.866	Full model
BSQ6H1	43	3,092.906	-6,099.812	-5,766.154	**
BSQ6H2	44	3,185.676	-6,283.352	-5,941.934	**
BSQ6H3	45	3,202.330	-6,314.660	-5,965.482	**
BSQ6H4	46	3,197.208	-6,302.416	-5,945.478	**
BSQ6H5^b	47	3,223.592	-6,353.184	-5,988.488	ns^c
BSQ6H6	48	3,224.676	-6,353.352	-5,980.896	Full model
BSC6H1	43	3,079.741	-6,073.482	-5,739.824	**
BSC6H2	44	3,171.470	-6,254.940	-5,913.522	**
BSC6H3	45	3,185.888	-6,281.776	-5,932.598	**
BSC6H4	46	3,182.311	-6,272.622	-5,915.684	**
BSC6H5	47	3,208.625	-6,323.250	-5,958.554	ns ^c
BSC6H6	48	3,210.447	-6,324.894	-5,952.438	Full model

** = p < 0.01.
^aBSXkHy, where X = L, Q e C represents the degree of linear, quadratic and cubic polynomials, respectively, k is the polynomial order for the fixed curve and for the additive genetic and permanent environmental effects and y is the number of residual variance classes.
^bValues in bold indicate the best model based on LogL, AIC, BIC and LRT criterions.
^cns = no significant difference.

519
520
521
522
523

524

525

526

527

FIGURE CAPTIONS

528 **Figure 1.** Test-day milk yield (TDMY) per week for Alpine goats.

529 **Figure 2a.** Estimates of variances obtained for the LEG6H5 model.

530 **Figure 2b.** Variance estimates for the BSQ6H5 model.

531 **Figure 3.** Heritability estimates for test-day milk yield (TDMY) over the lactation for
532 the models **LEG6H5** (Legendre polynomial of order 6 and 5 classes of residual
533 variance) and **BSQ6H5** (Quadratic B-spline model of polynomial order 6 and 5 classes
534 of residual variance).

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573

REFERENCES

Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. *Selected Papers of Hirotugu Akaike*. Springer.

Ali, T., and Schaeffer, L. 1987. Accounting for covariances among test day milk yields in dairy cows. *Can. J. Animal Sci.*, **67**: 637-644.

Araújo, C.V., Torres, R.A., and Costa, C.N. 2006. Uso de funções ortogonais para descrever a produção de leite no dia de controle por meio de modelos de regressão aleatória. *Rev. Bras. de Zootecnia*, **35**: 967-974. [in Portuguese, English abstract].

Assis, G.M.L.D., Albuquerque, L.G.D., Sarmento, J.L.R., Carneiro Júnior, J.M., Lopes, P.S., and Rodrigues, M.T. 2006. Estimacão de parâmetros genéticos em caprinos leiteiros por meio de análise de regressão aleatória utilizando-se a Amostragem de Gibbs. *Rev. Bras. de Zootecnia*, **35**: 706-714. [in Portuguese, English abstract].

Bignardi, A., El Faro, L., Torres Júnior, R., Cardoso, V., Machado, P., and Albuquerque, L.G.D. 2011. Random regression models using different functions to model test-day milk yield of Brazilian Holstein cows. *Gen. Mol. Res.*, 3565-3575.

Bohmanova, J., Miglior, F., Jamrozik, J., Misztal, I., and Sullivan, P. 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.

- 574 Breda, F.C., Albuquerque, L.G., Yamaki, M., Reis Filho, J.C., Sarmiento, J.L.R., Lopes,
575 P.S., and Rodrigues, M.T. 2006. Estimação de parâmetros genéticos para
576 produção de leite de cabras da raça Alpina. *Rev. Bras. de Zootecnia*, **35**: 396-
577 404. [in Portuguese, English abstract].
- 578 Cobuci, J.A., Euclides, R.F., Lopes, P.S., Costa, C.N., Torres, R.D.A., and Pereira, C.S.
579 2005. Estimation of genetic parameters for test-day milk yield in Holstein cows
580 using a random regression model. *Gen. Mol. Biol.*, **28**: 75-83.
- 581 Costa, C.N., Melo, C.M.R.D., Packer, I.U., Freitas, A.F.D., Teixeira, N.M., and Cobuci,
582 J.A. 2008. Genetic parameters for test day milk yield of first lactation Holstein
583 cows estimated by random regression using Legendre polynomials. *Rev. Bras.*
584 *de Zootecnia*, **37**: 602-608.
- 585 DeGroot, B., Keown, J.F., Van Vleck, L.D., and Kachman, S.D. 2007. Estimates of
586 genetic parameters for Holstein cows for test-day yield traits with a random
587 regression cubic spline model. *Genet. Mol. Res.*, **6**: 434-444.
- 588 Englishby, T., Banos, G., Moore, K., Coffey, M., Evans, R., and Berry, D. 2016.
589 Genetic analysis of carcass traits in beef cattle using random regression models.
590 *J. Animal Sci.*, **94**: 1354-1364.
- 591 Flores, E. and WERF, J. 2015. Random regression test day models to estimate genetic
592 parameters for milk yield and milk components in Philippine dairy buffaloes. *J.*
593 *Animal Breed. and Gen.*, **132**: 289-300.
- 594 Huisman, A., Veerkamp, R., and Van Arendonk, J. 2002. Genetic parameters for
595 various random regression models to describe the weight data of pigs. *J. Animal*
596 *Sci.*, **80**: 575-582.
- 597 Kheirabadi, K. and Rashidi, A. 2016. Genetic description of growth traits in Markhoz
598 goat using random regression models. *Small Rum. Res.*, **144**: 305-312.

- 599 Kheirabadi, K., Rashidi, A., Alijani, S., and Imumorin, I. 2014. Modeling lactation
600 curves and estimation of genetic parameters in Holstein cows using multiple-
601 trait random regression models. *Animal Sci. J.*, **85**: 925-934.
- 602 Kirkpatrick, M., Lofsvold, D. & Bulmer, M. 1990. Analysis of the inheritance, selection
603 and evolution of growth trajectories. *Genetics*, **124**: 979-993.
- 604 León, J.M., Macciotta, N.P.P., Gama, L.T., Barba, C., and Delgado, J.V. 2012.
605 Characterization of the lactation curve in Murciano-Granadina dairy goats.
606 *Small Rumin. Res.*, **107**: 76-84.
- 607 Liang, J. and Devendra, C. 2014. Expanding the contribution of dairy goats in efficient
608 and sustainable production systems. *Animal Prod. Sci.*, **54**: 1198-1203.
- 609 López-Romero, P., Rekaya, R., and Carabano, M. 2004. Bayesian comparison of test-
610 day models under different assumptions of heterogeneity for the residual
611 variance: the change point technique versus arbitrary intervals. *J. Animal Breed.*
612 *Gen.*, **121**: 14-25.
- 613 Menezes, G.R.D.O., Torres, R.D.A., Sarmiento, J.L.R., Rodrigues, M.T., Brito, L.F.,
614 Lopes, P.S., and Silva, F.G.D. 2011. Random regression models in the milk
615 yield evaluation in Saanen goats. *Rev. Bras. de Zootecnia*, **40**: 1526-1532.
- 616 Meyer, K. 2005. Random regression analyses using B-splines to model growth of
617 Australian Angus cattle. *Gen. Sel. Evol.*, **37**: 1.
- 618 Meyer, K. 2007. WOMBAT - A tool for mixed model analyses in quantitative genetics
619 by restricted maximum likelihood (REML). *Journal of Zhejiang University*
620 *Science B*, **8**: 815-821.
- 621 Oliveira, H.R.D., Silva, F., Siqueira, O., Souza, N., Junqueira, V., Resende, M.,
622 Borquis, R., and Rodrigues, M.T. 2016. Combining different functions to

- 623 describe milk, fat, and protein yield in goats using Bayesian multiple-trait
624 random regression models. J. Animal Sci., **94**: 1865-1874.
- 625 Olori, V., Hill, W., Mcguirk, B., and Brotherstone, S. 1999. Estimating variance
626 components for test day milk records by restricted maximum likelihood with a
627 random regression animal model. Livest. Prod. Sci., **61**: 53-63.
- 628 Pereira, R., Bignardi, A., El Faro, L., Verneque, R., Vercesi Filho, A., and Albuquerque,
629 L. 2013. Random regression models using Legendre polynomials or linear
630 splines for test-day milk yield of dairy Gyr (*Bos indicus*) cattle. J. Dairy Sci.,
631 **96**: 565-574.
- 632 Rovadoscki, G.A., Petrini, J., Ramirez-Diaz, J., Pertile, S.F., Pertille, F., Salvian, M.,
633 Iung, L.H., Rodriguez, M.A.P., Zampar, A., and Gaya, L.G. 2016. Genetic
634 parameters for growth characteristics of free-range chickens under univariate
635 random regression models. Poultry Sci. 167.
- 636 Sánchez, J., Misztal, I., Aguilar, I., and Bertrand, J. 2008. Genetic evaluation of growth
637 in a multibreed beef cattle population using random regression-linear spline
638 models. J. Animal Sci., **86**: 267-277.
- 639 Sarmento, J.L.R., Albuquerque, L.G.D., Torres, R.D.A., Rodrigues, M.T., Lopes, P.S.,
640 and Reis Filho, J.C. 2008. Comparison of random regression models for the
641 estimation of genetic parameters in dairy goats. Rev. Bras. de Zootecnia, **37**:
642 1788-1796.
- 643 Sarmento, J.L.R., Reis Filho, J.C., Albuquerque, L.G.D., Lopes, P.S., and Rodrigues,
644 M.T. 2006. Genetic evaluation of Alpine goats using test day milk yield. Rev.
645 Bras. de Zootecnia, **35**: 443-451.
- 646 SAS Institute 2014. *Base SAS 9.4 Procedures Guide*, SAS Institute.

- 647 Sasaki, O., Aihara, M., Nishiura, A., Takeda, H., and Satoh, M. 2015. Genetic analysis
648 of the cumulative pseudo-survival rate during lactation of Holstein cattle in
649 Japan by using random regression models. *J. Dairy Sci.*, **98**: 5781-5795.
- 650 Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics*, **6**:
651 461-464.
- 652 Silva, F., Torres, R., Brito, L., Euclides, R., Melo, A., Souza, N., Ribeiro Júnior, J., and
653 Rodrigues, M.T. 2013. Random regression models using Legendre orthogonal
654 polynomials to evaluate the milk production of Alpine goats. *Gen. Mol. Res.*,
655 **12**: 6502-6511.
- 656 Silvestre, A., Petim-Batista, F., and Colaco, J. 2006. The accuracy of seven
657 mathematical functions in modeling dairy cattle lactation curves based on test-
658 day records from varying sample schemes. *J. Dairy Sci.*, **89**: 1813-1821.
- 659 Strucken, E. M., Laurenson, Y. C., and Brockmann, G. A. 2015. Go with the flow—
660 biology and genetics of the lactation cycle. *Frontiers Gen.*, **6**:118.
- 661 Thepparat, M., Boonkum, W., Duangjinda, M., Tumwasorn, S., Nakavisut, S., and
662 Thongchumroon, T. 2015. Genetic evaluation using random regression models
663 with different covariance functions for test-day milk yield in an admixture
664 population of Thailand goats. *Animal Sci. J.*, **86**: 655-660.
- 665 Vrieze, S. I. 2012. Model selection and psychological theory: a discussion of the
666 differences between the Akaike information criterion (AIC) and the Bayesian
667 information criterion (BIC). *Psyc. Meth.*, **17**(2), 228.
- 668 Wilmlink, J. 1987. Adjustment of test-day milk, fat and protein yield for age, season and
669 stage of lactation. *Liv. Prod. Sci.*, **16**: 335-348.
- 670 Wolfinger, R. 1993. Covariance structure selection in general mixed models.
671 *Communications in statistics-Simulation and computation*, **22**: 1079-1106.

672

For Review Only

Figure 1. Test-day milk yield (TDMY) per week for Alpine goats.

Figure 2a. Estimates of variances obtained for the LEG6H5 model.

Figure 2b. Variance estimates for the BSQ6H5 model.

Figure 3. Heritability estimates for test-day milk yield (TDMY) over the lactation for the models **LEG6H5** (Legendre polynomial of order 6 and 5 classes of residual variance) and **BSQ6H5** (Quadratic B-spline model of polynomial order 6 and 5 classes of residual variance).