



Estimation of genetic parameters for test-day milk yield in Girolando cows using a random regression model

[*Estimação de parâmetros genéticos para produção de leite no dia do teste em vacas da raça Girolando utilizando modelo de regressão aleatória*]

E.P.B. Santos¹, G.L. Feltes², R. Negri², J.A. Cobuci³, M.V.G.B. Silva⁴

¹Aluno de pós-graduação – Universidade Estadual de Santa Cruz – Ilhéus, BA

²Aluno de pós-graduação – Universidade Federal do Rio Grande do Sul – Porto Alegre, RS

³Universidade Federal do Rio Grande do Sul – Porto Alegre, RS

⁴Embrapa Gado de Leite – Juiz de Fora, MG

ABSTRACT

The objective of this study was to estimate the components of variance and genetic parameters of test-day milk yield in first lactation Girolando cows, using a random regression model. A total of 126,892 test-day milk yield (TDMY) records of 15,351 first-parity Holstein, Gyr, and Girolando breed cows were used, obtained from the Associação Brasileira dos Criadores de Girolando. To estimate the components of (co) variance, the additive genetic functions and permanent environmental covariance were estimated by random regression in three functions: Wilmink, Legendre Polynomials (third order) and Linear spline Polynomials (three knots). The Legendre polynomial function showed better fit quality. The genetic and permanent environment variances for TDMY ranged from 2.67 to 5.14 and from 9.31 to 12.04, respectively. Heritability estimates gradually increased from the beginning (0.13) to mid-lactation (0.19). The genetic correlations between the days of the control ranged from 0.37 to 1.00. The correlations of permanent environment followed the same trend as genetic correlations. The use of Legendre polynomials via random regression model can be considered as a good tool for estimating genetic parameters for test-day milk yield records.

Keywords: genetic correlation, heritability, Legendre polynomials, Linear spline polynomials, wilmink function

RESUMO

O objetivo deste estudo foi estimar os componentes de variância e os parâmetros genéticos da produção de leite no dia do teste (TDMY) em vacas Girolando de primeira lactação, usando modelo de regressão aleatória. Foram utilizados 126.892 registros de produção de leite no dia controle de 15.351 vacas primíparas das raças Holandesa, Gir e Girolando, obtidas na Associação Brasileira dos Criadores de Girolando. Para estimar os componentes de (co) variância, as funções genéticas aditivas e de covariância ambiental permanente foram estimadas por regressão aleatória em três funções: Wilmink, polinômios de Legendre (terceira ordem) e polinômios splines lineares (três nós). A função polinomial de Legendre apresentou melhor qualidade de ajuste. As variâncias genéticas e de ambiente permanente para produção de leite no dia do controle variaram de 2,67 a 5,14 e de 9,31 a 12,04, respectivamente. As estimativas de herdabilidade aumentaram gradativamente do início (0,13) para o meio da lactação (0,19). As correlações genéticas entre os dias do controle variaram de 0,37 a 1,00. As correlações de ambiente permanente seguiram a mesma tendência das correlações genéticas. A utilização dos polinômios de Legendre via modelos de regressão aleatória pode ser considerada como uma boa ferramenta para estimação de parâmetros genéticos da produção de leite no dia do teste.

Palavras-chave: correlação genética, função de Wilmink, herdabilidade, polinômios de Legendre, polinômios splines lineares

Recebido em 3 de junho de 2020

Aceito em 23 de setembro de 2020

E-mail: nellepbs@gmail.com

INTRODUCTION

Crossbreeding has been adopted for many years in Brazilian dairy farming as a strategy for genetic improvement to combine the high milk production capacity and adaptability to tropical conditions. This strategy led to creation of Girolando breed (5/8 Holstein and 3/8 Gyr). Notable for their excellent productivity, Girolando and crossbred animals also stand out for their high fertility rates and good vigor. These traits made the breed conquer space and become predominant in dairy farms (Stafuzza *et al.*, 2017), representing 80% of milk production in Brazil (Nanzer, 2011).

Studies in the genetic field have been intensified due to importance of Girolando breed for tropical regions. Mathematical models applied in genetic evaluations of sires and cows seek to describe the different lactation curves and allow studying genetically the components of lactation. For this, the methodology that evaluates production on test-day through random regression models (RRM) has been widely used to improve the quality of genetic assessments (Herrera *et al.*, 2008). Associated with different mathematical functions in the most accurate search to infer genetically about the breed, the RRM methodology allows estimating genetic parameters, predicting breeding values, and identifying the most efficient and profitable genotypes (Daltro *et al.*, 2020), thus allowing higher genetic gains.

According to Schaeffer (2004), one of great advantages of RRM is to adjust the random lactation curve for each animal, expressing it as a deviation from an average curve of the population or a group of individuals. The regressions are adjusted using functions and trajectory modeling for the population average (fixed regression) and animal (random regression). Using complete lactation curve to select animals genetically has been a complex task in recent decades, the option is to investigate the feasibility of selecting specific patterns of lactation curves or derived traits (Oliveira *et al.*, 2019).

In addition to testing mathematical functions, already tested in other breeds but little

investigated in Girolando breed, the demand is to improve the efficiency of genetic evaluations and simplify the animal selection process. Thus, this study aimed to estimate the variance components and genetic parameters of test-day milk yield in first lactation Girolando animals, by adjusting for Wilmink functions, Legendre polynomials, and Linear spline polynomials, using a random regression model.

MATERIALS AND METHODS

The test-day milk yield (TDMY) records were obtained from the Associação Brasileira dos Criadores de Girolando, Brazil. The animals were of the Holstein (H), Gyr (G), and Girolando breeds: 1/4H:3/4G (1/4H), 3/8H:5/8G (3/8H), 1/2H:1/2G (1/2H), 5/8H:3/8G (5/8H), 3/4H:1/4G (3/4H), 7/8H:1/8G (7/8H), and the pure synthetic (PS) from the crossing of parents (5/8H). Animals from these groups are officially called Girolando in Brazil. Some restrictions were made in animal observations to improve the quality of information and consistency of analyses. Only the first-parity order was considered. Extreme records of milk yield (<4 or >40 kg), age at first calving (<22 or >48 months), days in milk (DIM, <5 or >305 days), and number of individual controls (<4) were excluded from the data set.

Contemporary groups (herd-year-season, HYS) considered four age classes (21 to 25 months; 26 and 27 months; 28 and 29 months, and over 29 months) and two calving seasons (rainy and dry). The minimum size of each contemporary group was three animals. Records of daughters of sires with at least one daughter in at least three herds were accepted to evaluation. Following these criteria, a total of 126,892 TDMY records from 15,351 first lactations, collected between 1990 and 2014, were analyzed. This database was used to adjust all the evaluated functions. The structure of the data set after editing is summarized in Table 1.

Table 1. Description of database of the first lactation of Brazilian Holstein, Gyr and Girolando cattle

Information	Statistics
Animals in pedigree file	36,740
Dams in pedigree file	18,033
Sires in pedigree file	4,011
Animals with records	15,351
H	346
G	84
1/2H	3,762
1/4H	479
3/4H	5,466
3/8H	580
5/8H	2,871
7/8H	1,600
PS	163
Contemporary groups	2,928
Test-day records	12,892
Mean records/animal	8.26
Milk yield mean, kg/day	17.06 ± 6.74

The analyses were performed using a random regression animal model in which the effects of the contemporary group and genetic group how fixed effects and the covariates cow age at calving as linear and quadratic effects, in addition to the average lactation curve of the population. The random, additive genetic, and permanent environmental effects of the animal were modeled using the Wilink parametric function, Legendre polynomials or Linear spline polynomials, and the fixed trajectory.

1. Wilink (Wilink, 1987):

$$y_t = a + be^{-kt} + ct$$

where, y is the TDMY, t is the DIM, a is the initial milk yield, b is the increase in milk yield until the peak lactation, and c the decrease in milk yield after the peak. The k parameter is related to the time of peak lactation and usually assumes a fixed value, derived from a preliminary analysis carried out on the average yield.

2. Legendre polynomial (Kirkpatrick et al., 1994):

$$y_t = \sum_{i=0}^n \alpha_i \Phi_i(d_t^*)$$

where, d_t^* represents the standardized time unit, which varies from -1 to $+1$, that is,

$$d_t^* = -1 + 2 \left(\frac{d_t - d_{\min}}{d_{\max} - d_{\min}} \right), \text{ in which } d_{\min}$$

and d_{\max} are the lowest and highest days of lactation that can be found in the data set, respectively; and d_t is the DIM. The standardized DIM can be defined as:

$$\Phi_{(d_t^*)^i} = \frac{1}{2^i} \sqrt{\frac{2i+1}{2}} \sum_{m=0}^{i/2} (-1)^m \binom{i}{m} \binom{2i-2m}{i} (d_t^*)^{i-2m}$$

where, d_t^* is the i -th DIM, i is the order of Legendre function (third degree in this study), and m is the number of indices needed to determine the polynomial.

3. Linear spline (Misztal, 2006):

Let T be a vector of n knots, then the covariates of linear spline for DIM $t(\Phi_i(t))$ located between the T_i and T_{i+1} knot can be calculated as:

$$\Phi_i(t) = \frac{t - T_i}{T_{i+1} - T_i},$$

$$\Phi_{i+1}(t) = \frac{T_{i+1} - t}{T_{i+1} - T_i} = 1 - \Phi_i(t), \text{ and}$$

$$\Phi_{1..i-1, i+2..n} = 0$$

if $t = T_i$, $\Phi_i(t) = 1$, and $\Phi_{1..i-1, i+2..n} = 0$.

The vector Φ of the DIM t has at most two non-zero elements, and the sum of all elements of the vector is equal to 1. The formula above assumes that $T_i \leq t < T_{i+1}$. In this case, 3 knots {5, 45, 305} were considered.

The residual variance was considered homogeneous for all functions. All analyses of variance components were estimated using the restricted maximum likelihood method, using the software REMLF90 (Misztal et al., 2002). The quality of fit of mathematical functions was compared using the Akaike's information criterion ($AIC = -2\log L + 2p$, where p is the number of parameters in the model); Schwarz's Bayesian information criterion ($BIC = -2\log L + p \log(\lambda)$, in which $\log(\lambda)$ is the natural logarithm of the sample size (or dimension of y))

and p is the number of parameters in the model); and the log-likelihood function ($-2\log$).

The heritability estimates for trait on the different DIM and the genetic and permanent environmental correlations between different DIM was estimated and plotted graphically to verify which of them were in accordance with the previously reported estimates for the same trait but other breeds or populations.

RESULTS AND DISCUSSION

Lower values of likelihood function, AIC and BIC were calculated for the model that considered the parametric function of Legendre polynomials, indicating it is the best model to be considered in the genetic evaluation of TDMY

for animals of the Girolando breed (Table 2). AIC takes into considers the complexity and predictive capacity of the data fit model, and although the models have the same number of parameters, BIC also indicated the Legendre polynomial function as the most appropriate. Moreover, a substantial difference was observed in the goodness-of-fit values between the tested models. The result corroborates with that obtained by Canaza-Cayo *et al.* (2015), who recommended to use the Legendre function, third order for genetic additive effect e fifth order about permanent environmental to describe changes in the components of (co)variance of milk yield during lactation of Girolando cows, as it is the most plausible in the specification of additive genetic effects and permanent environmental effects.

Table 2. Values of tests of goodness of fit obtained by random regression models using the Wilmink function, Legendre polynomials, and Linear spline polynomials

Model	P	-2log	AIC	BIC
Wilmink	11	354449	708914	708992
Legendre	11	353873	707768	707875
Spline	11	353999	708016	708104

P: number of estimated parameters; -2log: log-likelihood function; AIC: Akaike information criterion; BIC: Schwarz Bayesian information criterion.

The heritability estimates for TDMY during lactation varied little in the model using the Legendre polynomials (Figure 1). In this case, the highest heritability values were identified in the intermediate lactation period (125 to 185 days). Heritability for milk yield accumulated at 305 days was estimated at 0.26. The same behavior was observed for the genetic and permanent environmental variances during lactation. Considering the models with the Linear spline and Wilmink functions the heritability estimates for TDMY decreased after the beginning of lactation. However, the Linear spline function showed a higher variation, where the heritability estimates remained very low during the peak lactation phase. The heritability estimates for milk yield accumulated at 305 days were 0.16 and 0.27 for Linear spline and

Wilmink functions, respectively. The genetic and permanent environmental variances showed the same graphic behavior, being smaller at the beginning of lactation, highlighting the highest influence of environmental factors at this phase.

An important factor is that extrapolation of estimates may also be associated with the low number of phenotypic observations collected in the initial period of lactation (Meyer, 1998). In fact, its association with the effects of permanent environmental and residual variance leads to a bias in the heritability estimates at the beginning of the lactation. The differences between estimates may be associated with the mathematical functions used in the analyses, as the same database was used.

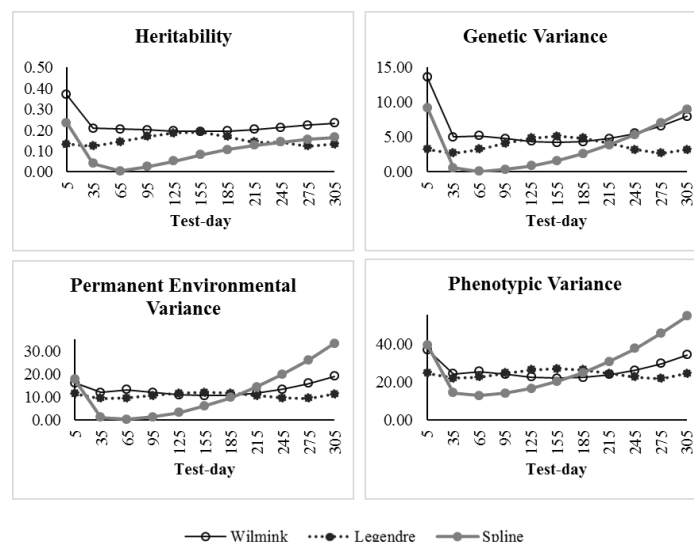


Figure 1. Heritability estimates and variances for test-day milk yield (TDMY) for Girolando cattle (Holsten x Gyr) by random regression models using the Wilmink function, Legendre polynomials, and Linear spline polynomials

The heritability estimates obtained in this study for the Girolando breed are within the range of magnitude of values observed in the literature for several dairy breeds. Freitas *et al.* (1996) used the production model of up to 305 days and obtained a heritability estimate of 0.26 for crossbred Holstein-Gyr cows. On the other hand, Facó *et al.* (2009) identified heritabilities from 0.24 to 0.31 for milk yield in the Girolando breed through an animal model that included only the cow genetic group. Canaza-Cayo *et al.* (2015) estimated heritability from 0.18 to 0.23 and 0.23 to 0.29, respectively, for this breed through the random regression.

Cobuci *et al.* (2005) used the production data on the test-day of Holstein animals and obtained heritability estimates between 0.15 and 0.31. On the other hand, Costa *et al.* (2005) used information from Gyr animals and estimated heritability from 0.20 to 0.70. Jamrozik and Schaeffer (1997) used the Wilmink function in random regression models to estimate the heritability of Holstein cattle in Canada and obtained a high estimate (0.38) at the beginning of lactation.

Genetic correlations between the test-days estimated by the different functions ranged from 0.37 to 1.00 by Legendre, -0.43 to 0.99 by Wilmink, and -0.38 to 1.00 by Linear spline polynomials (Figure 2). The highest correlation

estimates were found among adjacent controls, while the lowest, or intermediate and negative estimates were observed among the most distant controls. The genetic correlation estimated through the Legendre polynomials was similar, in terms of magnitude, to the estimates reported by Pereira *et al.* (2013) for Gyr breed animals and Gyr x Holstein crossbred animals, Herrera *et al.* (2008) for Gyr, and Padilha *et al.* (2019) for Holstein animals. Genetic correlations are estimates of great importance in well-structured breeding programs. Positive and high genetic correlations indicate that the selection to increase milk yield at a given lactation stage may have a positive influence on the other lactation stages, and thus provide greater gains in response to selection.

Using Wilmink or Ali and Schaeffer functions to model fixed and random regressions for milk production in Gyr cattle, Pereira *et al.* (2010) found negative genetic correlations between the TDMY at the beginning and end of lactation, which would be due to the bovine somatotropin application associated with the preferential treatment of animals in some herds. Negative correlations between the extreme points of the lactation curve were reported by Canaza-Cayo *et al.* (2015) for Girolando animals.

Estimation of genetic...

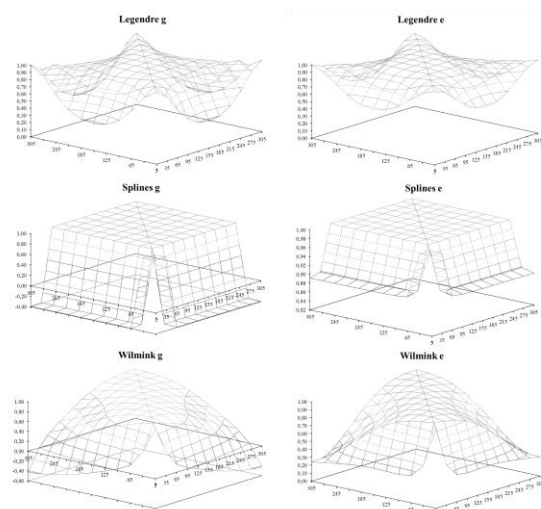


Figure 2. Genetic (g, on the right) and permanent environmental correlations (e, on the left) for test-day milk yield (TDMY) for Girolando cattle (Holsten x Gyr) by random regression models using the Wilmink function, Legendre polynomials, and Linear spline polynomials

Low or negative genetic correlations are not expected and indicate that this model has not been properly adapted to the data structure. This behavior is not expected in biological terms and has been attributed to the poor fit of the random regression model at the extreme points of the curve, probably due to the low number of observations during these periods (Meyer, 1999). In this sense, the Legendre polynomial function presents greater flexibility and practical applicability to estimate genetic component for test-day milk yield using random regression. The permanent environmental correlations between the test-days estimated by the different functions ranged from 0.58 to 0.99 by Legendre, 0.24 to 0.99 by Wilmink, and 0.89 to 1.00 by Linear spline polynomials. The highest estimates of correlations were found among adjacent controls, while the lowest estimates were observed among the most distant controls. Pereira *et al.* (2013) reported similar results for Gyr animals and the crossing between Gyr x Holstein.

Statistically and biologically, the inference to the permanent environment component is more complex when compared to the direct genetic component. According to Canaza-Cayo *et al.* (2015), a delicate factor when it comes to genetic estimation in Girolando cattle is the influence of the short period of lactation, typical of *Bos indicus* animals. That is why the impact between the controls is easily observed, whether they are adjacent or distant.

CONCLUSION

The results of this study show that third-order Legendre polynomials are the best alternative among functions with three parameters to be estimated. The heritability estimates and genetic and permanent environmental correlations estimated by this model are consistent with previous reports in the literature.

REFERENCES

- CANAZA-CAYO, A.W.; LOPES, P.S.; SILVA, M.V.G.B. *et al.* Genetic parameters for milk yield and lactation persistency using random regression models in Girolando cattle. *Asian Australas. J. Anim. Sci.*, v.28, p.1407-1418, 2015.
- COBUCCI, J.A.; EUCLYDES, R.F.; LOPES, P.S. *et al.* Estimation of genetic parameters for test-day milk yield in Holstein cows using a random regression model. *Genet. Mol. Biol.*, v.28, p.75-83, 2005.
- COSTA, C.N.; MELO, C.M.R.; MACHADO, C.H.C. *et al.* Parâmetros genéticos para a produção de leite de controles individuais de vacas da raça Gir estimados com modelos de repetibilidade e regressão aleatória. *Rev. Bras. Zootec.*, v.34, p.1519-1530, 2005.

- DALTRO, D.S.; SILVA, M.V.G.B.; GAMA, L.T. *et al.* Estimates of genetic and crossbreeding parameters for 305-day milk yield of Girolando cows. *Ital. J. Anim. Sci.*, v.19, p.86-94, 2020.
- FACÓ, O.; MARTINS FILHO, R.; LOBO, R.N.B. *et al.* Efeito da redução da variação da duração de lactação na avaliação genética de bovinos leiteiros mestiços. *Rev. Cienc. Agron.*, v.40, p.287-292, 2009.
- FREITAS, A.F.; TEIXEIRA, N.M.; VALENTE, J. Fatores genéticos e de ambiente sobre características produtivas e reprodutivas em rebanhos de animais mestiços. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 1996, Fortaleza. *Anais...* Fortaleza: SBZ, 1996. p.59-60, (Resumo).
- HERRERA, L.G.G.; EL FARO, L.; ALBUQUERQUE, L.G. *et al.* Estimativas de parâmetros genéticos para produção de leite e persistência da lactação em vacas Gir, aplicando modelos de regressão aleatória. *Rev. Bras. Zootec.*, v.37, p.1584-1594, 2008.
- JAMROZIK, J.; SCHAEFFER, L.R. Estimates of genetic parameters for a test day model with random regressions for yield traits. *J. Dairy Sci.*, v.80, p.762-770, 1997.
- KIRKPATRICK, M.; HILL, W.G.; THOMPSON, R. Estimating the covariance structure of traits during growth and aging, illustrated with lactations in dairy cattle. *Genet. Res.*, v.64, p.57-69, 1994.
- MEYER, K. Estimating covariance functions for longitudinal data using a random regression model. *Genet. Sel. Evol.*, v.30, p.221-240, 1998.
- MEYER, K. Estimates of genetic and phenotypic covariance functions for postweaning growth and mature weight of beef cows. *J. Anim. Breed. Genet.*, v.116, p.181-205, 1999.
- MISZTAL, I. Properties of random regression models using linear splines. *J. Anim. Breed. Genet.*, v.123, p.74-80, 2006.
- MISZTAL, I.; TSURUTA, S.; STRABEL, T. *et al.* BLUPF90 and related programs (BGF90). In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier. *Proceedings...* Montpellier, France; 2002. (CD-Rom Communication n° 28-07).
- NANZER, T.A.D.T. Produção de leite no Brasil e participação da genética Girolando com ênfase em reprodução. Associação Brasileira dos Criadores de Girolando, 2011. Disponível em: <http://www.girolando.com.br/index.php?paginas/Site/tecnico,39>. Acessado em: 12 out. 2019.
- OLIVEIRA, H.; BRITO, L.; SILVA, F. *et al.* Genomic prediction of lactation curves for milk, fat, protein, and somatic cell score in Holstein cattle. *J. Dairy Sci.*, v.102, p.452-463, 2019.
- PADILHA, A.H.; ALFONZO, E.P.M.; DALTRO, D.S. *et al.* Genetic trends and genetic correlations between 305-day milk yield, persistency and somatic cell score of Holstein cows in Brazil using random regression model. *Anim. Prod. Sci.*, v.59, p.207-215, 2019.
- PEREIRA, R.J.; AYRES, D.R.; EL FARO, L. *et al.* Genetic parameters for production traits of dairy Gyr (*Bos indicus*) × Holstein cattle estimated with a random regression model. *Livest. Sci.*, v.158, p.24-31, 2013.
- PEREIRA, R.J.; LOPES, P.S.; VERNEQUE, S. *et al.* Funções de covariância para produção de leite no dia do controle em bovinos Gir leiteiro Rodrigo. *Pesqui. Agropecu. Bras.*, v.45, p.1303-1311, 2010.
- SCHAEFFER, L.R. Application of random regression models in animal breeding. *Livest. Prod. Sci.*, v.86, p.35-45, 2004.
- STAFUZZA, N.B.; ZERLOTINI, A.; LOBO, F.P. *et al.* Single nucleotide variants and InDels identified from whole-genome re-sequencing of Guzerat, Gyr, Girolando and Holstein cattle breeds. *PLoS One*, v.12, p.1-15, 2017.
- WILMINK, J.B.M. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.*, v.16, p.335-348, 1987.