Scientific Papers. Series D. Animal Science. Vol. LXVIII, No. 1, 2025
ISSN 2285-5750; ISSN CD-ROM 2285-5769; ISSN Online 2393-2260; ISSN-L 2285-5750

RESEARCHES ON MODELING OF THE LACTATION CURVE IN ROMANIAN SPOTTED SIMMENTAL CATTLE BREED

Roxana-Bianca COȘA[1, 2], Horia Grosu[2]

[1]Institute of Hygiene and Veterinary Public Health, 37-39 Olteniței Road, District 4,
Bucharest, Romania
[2]University of Agronomic Sciences and Veterinary Medicine of Bucharest,
59 Marasti Blvd, District 1, Bucharest, Romania

Corresponding author email: dr.cosaroxanabianca@gmail.com

Abstract

Milk production in dairy cattle is influenced by multiple genetic and environmental factors, making precise lactation curve modeling essential for accurate milk yield estimation and genetic evaluation. In recent decades, numerous scientific studies have focused on improving these models to enhance the accuracy of milk production estimates for standard lactation. This study aims to identify the most suitable model for predicting breeding values using a limited number of test-day records at different lactation stages. Data from two lactations were analyzed. For the first lactation, 9235 test records from 1270 Romanian Spotted Simmental cows were examined, while the second lactation included 9984 test records from 1409 cows of the same breed. Production and pedigree data were modeled using three biometric functions: Legendre Orthogonal Polynomial, Natural Cubic Spline, and P-Spline Function. Among these, the P-Spline Function provided the most accurate results. These findings are particularly valuable in animal breeding, enabling more precise genetic evaluations and contributing to improved selection strategies for dairy cattle.

***Key words**: genetic progress, heritability, optimization of genetic selection, P-Spline, Test-day model.*

INTRODUCTION

The prediction of breeding values in cattle based on milk yield recorded during monthly official production control is significantly more accurate than systems relying on total milk yield over a standard lactation period of 305 days. This increased accuracy is due to the fact that monthly test-day records (every 28 days, method A4) allow for performance corrections based on environmental factors acting at the time of phenotype expression. In contrast, when using the total 305-day milk yield, it is assumed that the same environmental conditions persist throughout lactation, which may lead to biased genetic evaluations (White et al., 1999; Druet et al., 2003). Furthermore, test-day models provide a more dynamic representation of lactation performance, capturing short-term variations caused by factors such as feeding strategies, health status, and seasonal influences (Misztal, 2006). By incorporating these variations, test-day models enhance the accuracy of genetic evaluations, reducing the risk of overestimating or underestimating an animal's true genetic potential (Macciotta et al., 2005; Bohmanova et al., 2008). Consequently, they have become a standard approach in modern dairy cattle breeding programs, facilitating more precise selection decisions and accelerating genetic progress (Silvestre et al., 2010; Grosu & Rotar, 2015).

The mathematical model that incorporates monthly test-day records is referred to as the "test-day" model. Such models require mathematical functions capable of accurately describing the evolution of the lactation curve from calving to the end of lactation. These models are important in accurately capturing the dynamics of milk production, as they reflect the varying environmental conditions and physiological changes that occur throughout lactation. Over time, various biometric models have been developed, ranging from nonlinear parametric functions (Wilmink, 1987; Ali & Schaeffer, 1987) to orthogonal Legendre polynomials, spline-based functions, and fractional polynomials (Macciotta et al., 2005; 2008; 2010; Silvestre et al., 2005; Silvestre et al., 2010; Misztal, 2006; Bohmanova et al., 2008). These models enable more precise estimation of breeding values, as they allow for the incorporation of genetic and environmental factors in a more accurate manner than traditional models based on total lactation yield. Furthermore, the flexibility of these models to accommodate different lactation patterns has made them a valuable tool in genetic evaluations and selection decisions in dairy cattle breeding (Jensen, 2001; Lidauer et al., 2003; Macciotta et al., 2011; McCarthy & Veerkaamp, 2012; Swalve, 1995; Wahinya et al., 2020).

In a previous study, the authors found that among the ten monthly test-day records, the highest heritability estimates were observed between the 4th and 8th test-day records. Since genetic progress is directly proportional to heritability, these findings suggest that the number of test-day records required for accurate genetic evaluation could be reduced from ten to five, focusing on the middle of the lactation period (Silvestre et al., 2010; Macciotta et al., 2008). This result is consistent with other studies showing that heritability tends to be higher during the peak of lactation, which is typically reached around the 4th to 8th test-day. Furthermore, by reducing the number of test-day records, the efficiency of genetic evaluation is enhanced, leading to more cost-effective breeding programs without sacrificing accuracy (Druet et al., 2003; Bohmanova et al., 2008). Thus, optimizing the number of test-day records is essential for improving genetic selection in dairy cattle breeding programs.

Given this, the next challenge is identifying the optimal mathematical function that can accurately predict breeding values from a reduced number of test-day records. Selecting the most suitable function is important as it ensures precise genetic evaluations with minimal data. Several mathematical functions have been proposed in the literature, each with its strengths in fitting the lactation curve. In this study, three mathematical functions were investigated: Legendre Orthogonal Polynomials, P-Spline Function, and Natural Cubic Spline Function. These models have been widely used due to their ability to represent the complex, nonlinear nature of lactation curves and to accommodate irregular test-day records (Macciotta et al., 2005; Misztal, 2006; Bohmanova et al., 2008). Among these, the P-Spline function has been reported to provide the best balance between model complexity and prediction accuracy (White et al., 1999; Bohmanova et al., 2008). Therefore, understanding the advantages of each function is essential for improving genetic evaluations and optimizing selection strategies in dairy cattle breeding.

MATERIALS AND METHODS

Two lactations were analyzed in this study. The first lactation included a sample of 9,235 test-day records from 1,270 Romanian Spotted Simmental cows. For the second lactation, a sample of 9,984 test-day records was taken from 1,409 cows of the same breed. The trait analyzed was the quantity of milk, measured and recorded during the first 10 months of lactation for both lactations 1 and 2.

To achieve the study's objective, seven working variants were considered. One control variant was used, which involved 10 test-day records per cow, while six experimental variants were examined with different sets of test-day controls: from 1 to 5; from 2 to 6; from 3 to 7; from 4 to 8; from 5 to 9; and from 6 to 10.

To model the evolution of the lactation curve, three mathematical functions were employed: the Legendre Orthogonal Polynomial, the P-Spline function, and the natural cubic Spline function. Calculations were performed using an R script written by H. Grosu.

The following biometric model was used to describe the performance at the monthly control level:

$$y_{itklm} = HYS_l + AGE_m + DIM_t + \sum_{k=0}^{q}(\beta \cdot z_{itk}) + \sum_{k=0}^{q}(\alpha_{ik} \cdot z_{itk}) + \sum_{k=0}^{q}(\gamma_{ik} \cdot z_{itk}) + e_{itklm} \quad [1]$$

where:

y_{itklm} is the observation of cow i measured at time t within herd-year-season (HYS), subclass l and age (AGE) subclass m;

DIM_t is the fixed effect of DIM clases;

b is the fixed covariate effect on the ith cow;

α_{ik} and γ_{ik} are the *kth* random regression of additive genetic and permanent environmental effects for cow *i*, respectively;

z_{itk} is the *kth* order of LPs/P-Spline/NCS for cow i measured at DIM_t, DIM_t denote the order of LPs;

e_{itklm} is the random residual.

The covariates are customized, case by case, depending on the biometric model used and the function used (Legendre polynomial, P-Spline or Natural Cubic Spline functions).

In matrix notation, the model [1] can be rewritten as follows:

$$y = X_1b_1 + X_2b_2 + X_3b_3 + X_4\beta + Z_\alpha + Z_\gamma + e \qquad [2]$$

where:

y = vector of test-day (TD) milk yield records;

X_1= incidence matrix for the fixed effect of herd-year-season (HYS);

b_1 = vector of fixed effect for HYS;

X_2= incidence matrix of fixed effect for age at calving;

b_2 = vector of fixed effect for age at calving;

X_3= incidence matrix of fixed effect for days in milk (DIM);

b_3 = vector of fixed effect for days in milk;

X_4 = covariate matrix for days in milk;

β = vector of fixed regression coefficients;

Z_1= covariates matrix for all animals;

α = vector of random regression coefficients for the breeding value;

Z_2 = covariates matrix for cows with records;

γ = vector of random regression coefficients for the permanent environmental effect;

e = vector of residual effects.

The system of equations corresponding to model [2] is as follows:

$$\begin{bmatrix} X^T \cdot X & X^T \cdot Z_1 & X^T \cdot Z_2 \\ {Z_1}^T \cdot X & {Z_1}^T \cdot Z_1 + \sigma_e^2 \cdot (A^{-1} \otimes G^{-1}) & {Z_1}^T \cdot Z_2 \\ {Z_2}^T \cdot X & {Z_2}^T \cdot Z_1 & {Z_2}^T \cdot Z_2 + \sigma_e^2 \cdot (I \otimes P^{-1}) \end{bmatrix} \cdot \begin{bmatrix} \tilde{b}_1 \\ \hat{\alpha} \\ \gamma \end{bmatrix} = \begin{bmatrix} X^T \cdot y \\ Z_1 \cdot y \\ {Z_2}^T \cdot X \end{bmatrix}$$

Where: $X = (X_1, X_2, X_3, X_4)$.

The (co)variance structure was assumed for random effects of model:

$$V = \begin{bmatrix} A \otimes G & 0 & 0 \\ 0 & I \otimes P & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

where:

Var (a) = $A \otimes G$;

Var (a) = additive variance;

Where ⊗ is Kronecker product function;

Var (p) = $I \otimes P$;

Var (p) = environmental variance;

G and P are the matrices of genetic and permanent environmental variances and covariances between random regression coefficients.

I = represents the identity matrix with the size equal with the number of cows with records;

σ_e^2= residual variance for lactation assumed to be constant throughout the lactation.

The three models were compared according to the residual variances and also to the correlation between the breeding values (Grosu & Rotar, 2015).

RESULTS AND DISCUSSIONS

Average population performances

In Table 1 summarizes the key data on milk production performance and heritability estimates for the Romanian Spotted breed during their first two lactations. This data provides valuable insights into the genetic potential of the breed, highlighting both the average milk yield and the heritability values observed across the two lactation periods. Understanding these factors is important for optimizing breeding programs, as heritability estimates reflect the proportion of variation in milk production that can be attributed to genetic factors, which are essential for selecting animals with superior milk-producing capabilities.

Table 1. Average milk production performances and heritability for the first two lactations in the Romanian Spotted Cattle Breed

COP	**Lactation**					
	Lactation I			**Lactation II**		
	n	$\bar{X} \pm S_{\bar{X}}$	h^2	**n**	$\bar{X} \pm S_{\bar{X}}$	h^2
1	862	21.12±0.198	0.314	910	24.14±0.237	0.297
2	908	21.18±0.206	0.272	1,014	23.25±0.224	0.253
3	914	20.81±0.202	0.236	1,036	22.27±0.228	0.222
4	947	19.91±0.199	0.254	1,061	21.18±0.22	0.237

COP	Lactation					
	Lactation I			Lactation II		
	n	$\bar{X} \pm S_{\bar{X}}$	h^2	n	$\bar{X} \pm S_{\bar{X}}$	h^2
5	997	18.93±0.189	0.194	1,070	20±0.205	0.184
6	1,003	18.38±0.186	0.251	1,035	19.29±0.212	0.241
7	976	17.7±0.189	0.24	1,055	18.02±0.195	0.225
8	980	17.02±0.181	0.238	1,053	17.17±0.189	0.228
9	900	16.37±0.196	0.304	958	16.25±0.192	0.287
10	748	15.97±0.214	0.381	792	15.38±0.204	0.372
Overall	9,235	18.76±0.195	0.265	9,984	19.76±0.210	0.251

Correlation between improvement values

The error variance values corresponding to the three biometric models considered are presented for different segments of the lactation period (1-5, 2-6, 3-7, 4-8, 5-9 and 6-10) in the first lactation (Table 2). From the results, it is evident that the lowest residual variances were obtained in the first lactation period (controls 1-5) using the P-Spline function, followed by the Legendre polynomial, and finally, the Natural Cubic Spline function. Specifically, the error variances in the order presented are 3.450, 3.989, and 5.421. These findings indicate that the P-Spline function is the most effective in capturing the underlying patterns of milk production during early lactation, while the Natural Cubic Spline function tends to produce higher residuals. Such differences in error variance are important for improving the accuracy of genetic evaluations and selecting models that best represent the lactation curve dynamics. Therefore, choosing the appropriate model can have significant implications for the efficiency of breeding programs in dairy cattle.

Table 2. Error variance associated with biometric models, depending on the serial number of controls, for lactation I

	Mathematical function	Order number of performed controls						
		1-5	2-6	3-7	4-8	5-9	6-10	1-10
Residual variance	Legendre polynomial	3.989	4.129	4.244	4.327	4.527	4.664	3.853
	P-Spline	3.450	3.582	3.856	3.879	4.093	4.459	3.329
	Natural Cubic Spline Function	5.421	5.507	5.732	5.706	5.421	6.209	5.172

The values obtained are consistent with previous findings, which indicate an inversely proportional relationship between the level of heritability and the size of the residual variance. Specifically, as heritability increases, residual variance tends to decrease, suggesting that higher genetic variability is associated with more precise estimations of milk production. To further explore this relationship, the degree of concordance between the improvement values calculated from 10 control points (considered as the control variant) and those derived from experimental variants, which used five performances taken sequentially throughout the lactation period, was assessed by calculating the correlation (Table 3).

Table 3. Correlation between the improvement values calculated
between the control variant and the experimental variants for lactation I

	Mathematical function	Order number of performed controls						
		1-5	2-6	3-7	4-8	5-9	6-10	1-10
Correlation between improvement values	Legendre polynomial	0.890	0.884	0.881	0.878	0.872	0.860	1.00
	P-Spline	0.916	0.913	0.908	0.897	0.871	0.858	1.00
	Natural Cubic Spline Function	0.879	0.877	0.876	0.873	0.860	0.845	1.00

The results indicate that the lowest residual variances were obtained in the middle of the lactation period, specifically between the 4th and 8th test-day records. This suggests that mid-lactation provides the most reliable data for genetic evaluation, potentially offering a more accurate prediction of breeding values. These findings are important for refining breeding programs, as they emphasize the importance of selecting appropriate test-day intervals to optimize the accuracy of genetic evaluations.

The 1-5 control interval was identified as the optimal variant, as it maximizes genetic progress by maintaining a high level of heritability. This interval also ensures a sufficiently strong concordance between the improvement values calculated from the control variant and those derived from the experimental variants. Specifically, the highest correlation was obtained using the P-Spline function (0.916), followed by the Legendre Polynomial (0.890), and the Natural Cubic Spline function (0.879).

These findings from the first lactation serve as a basis for evaluating similar trends in the second lactation, where comparable analyses will be conducted to assess the consistency of these results across different lactation periods.

Table 4. Error variance associated with biometric models, depending on the serial number of controls, for lactation II

	Mathematical function	Order number of performed controls						
		1-5	2-6	3-7	4-8	5-9	6-10	1-10
Residual variance	Legendre polynomial	4.023	4.229	4.441	4.848	5.440	7.235	3.981
	P-Spline	3.490	3.513	3.741	3.993	4.730	4.924	3.218
	Natural Cubic Spline Function	6.094	6.239	6.629	7.049	7.427	7.580	5.729

As shown in Table 4, the analysis of residual variance values for the second lactation reveals that, across the various segments of the lactation period (1-5, 2-6, 3-7, 4-8, 5-9, 6-10, 1-10), the smallest error variances were observed at the first part of the lactation period, specifically at controls 1-5. The P-spline function yielded the lowest variance, followed by the Legendre polynomial, and lastly, the Natural Cubic Spline function. The error variance values for these models were 3.490, 4.023, and 6.094, respectively. These results are consistent with previous studies, which have also demonstrated that mid-lactation provides the most accurate predictions for breeding value estimation.

With regard to the degree of concordance between the values calculated from the control variant and the experimental variants (Table 5).

Table 5. Correlation between the improvement values calculated between the control variant and the experimental variants for lactation II

	Mathematical function	Order number of performed controls						
		1-5	2-6	3-7	4-8	5-9	6-10	1-10
Correlation between the improvement values	Legendre polynomial	0.864	0.846	0.842	0.839	0.827	0.773	1.00
	P-Spline	0.878	0.885	0.871	0.839	0.833	0.778	1,00
	Natural Cubic Spline Function	0.842	0.839	0.833	0.830	0.813	0.803	1.00

It is observed that the highest correlations were achieved for controls 1-5, with the P-Spline function yielding the highest value (0.878), followed by the Legendre polynomial (0.864), and lastly, the Natural Cubic Spline function (0.842). These findings suggest that the performances recorded during this period provide a more accurate estimate of the improvement value, reinforcing the results obtained in the first lactation. Such consistency across both lactation periods highlights the reliability of the selected control intervals for genetic evaluation. These results are consistent with similar studies by other authors (White et al., 1999; Macciotta, 2005; 2008; 2011; Silvestre, 2005; 2010; Misztal, 2006), who also found that mid-lactation controls offer the best predictive power for breeding value estimation.

CONCLUSIONS

The findings of this study emphasize that modeling the lactation curve using advanced biometric functions significantly enhances the accuracy of estimating the improvement value in Romanian Spotted Cattle. Among the three models analyzed (Legendre Polynomial, P-Spline, and Natural Cubic Spline), the P-Spline function demonstrated superior adjustment capacity, yielding the lowest error variances and the highest correlations between the improvement values estimated based on the 5-control variants and the control variant (10 controls).

The results further indicated that the heritability of milk production is highest in the early part of lactation (controls 1-5), which facilitates more precise genetic evaluation. Consequently, reducing the number of measurements from 10 to 5 does not compromise prediction accuracy but rather optimizes the genetic selection process by reducing the costs and resources associated with performance control.

Based on these findings, it is recommended that, in predicting the breeding value of Romanian Spotted Cattle, the P-Spline function be employed using performances measured between controls 1-5, as per the A4 method. This approach can substantially contribute to optimizing genetic progress while lowering the costs associated with conducting official performance control (COP).

REFERENCES

Ali, A.K.A., & Schaeffer, L.R. (1987). Accounting for covariances among test-day milk yields in dairy cows. *Canadian Journal of Animal Science*. *67*(3): 637-644.

Bohmanova, J., Miglior, F., Jamrozik, F., Misztal, I., & Sullivan P.G. (2008). Comparison of Random Regression Models with Legendre Polynomials and Linear Splines for Production Traits and Somatic Cell Score of Canadian Holstein Cows. *Journal of Dairy Science, 91,* 3627–3638.

Druet, T., Jaffrezic, F., Boichard, D., & Ducrocq, V. (2003). Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *Journal of Dairy Science*, *86*, 2480–2490.

Grosu, H., Rotar, M. C. (2015). *Estimarea valorii de ameliorare la taurine, pe baza modelelor zilei de control (test day models)*. Bucharest, RO: Ceres Publishing House.

Jensen, J. (2001). Genetic evaluation of dairy cattle using test-day models. *Journal of Dairy Science,* 84(12), 2803-2812.

Lidauer, M.H., Mantysaari, E.A., & Stranden, I. (2003). Comparison of test-day models for genetic evaluation of production traits in dairy cattle. *Livestock Production Science*, 79(1), 73-86.

Macciotta, N.P.P., Vicario, D., & Cappio-Borlino, A., (2005). Detection of different shapes of lacta tion curve for milk yield in dairy cattle by empirical mathematical models. *Journal of Dairy Science, 88*, 1178-1191.

Macciotta, N.P.P., Miglior, F., Cappio-Borlino, A., & Schaeffer, L.R., (2008c). Issues in modelling lactation curves with regression splines. *Journal of Dairy Science* , *91*(E-Suppl.1), 544 (abstr.).

Macciotta, N.P.P., Miglior, F., Dimauro, C., & Schaeffer, L.R., (2010). Comparison of parametric, orthogonal, and spline functions to model individual lactation curves for milk yield in Canadian Holsteins. *Italian Journal Animal Science*, *9,* 460-464.

Macciotta, P.P Nicolò, Dimauro, C., Salvatore, P.G., Steri, R., & Pulina, G. (2011). The mathematical description of lactation curves in dairy cattle. *Italian Journal of Animal Science*, *10*.

McCarthy, J., & Veerkaamp, R.F. (2012). Estimation of genetic parameters for test-day records of dairy traits in a seasonal calving system. *Journal of Dairy Science, 95*(9), 5365-5377.

Misztal, I., (2006). Properties of random regres sion models using linear splines. *Journal Animal Breeding and Genetics*, *123*, 74-80.

Silvestre, A.M., Petim-Batista, M.F., & Colaco, J. (2005). Genetic parameter estimates for milk, fat and protein using a spline test day model. *Journal of Dairy Science, 88,* 1225-1230.

Silvestre, A.M.D., Almeida, J.C.M., Santos, V.A.C., Fontes, P.J.P., & Alves, V.C. (2010). Modeling lactation curves of "Barrosă" beef cattle with Wood's model *Italian Journal of Animal Science, 9*, 243-247.

Swalve, H.H. (1995). The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *Journal of Dairy Science,78*(4), 928-938.

Wahinya, P.K., Jeyaruban, M.G., Swan, A.A., Gilmour, A.R., & Magothe T.M. (2020). Genetic parameters for test-day milk yield, lactation persistency, and fertility in low-, medium-, and high-production systems in Kenya. *Journal of Dairy Science, 103*(11), 10399-10413.

White, I. M. S., Thompson, R., & Brotherstone, S. (1999). Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science, 82.*

Wilmink, J.B.M. (1987). Adjustment of Test-Day Milk, Fat and Protein Yield for Age, Season and Stage of Lactation. *Livestock Production Science, 16*(2), 335-348.