

## Estimation of genetic parameters for milk yield using a random regression test-day model in first parity dairy cows under pasture-based systems of Los Ríos region in Chile

Héctor Uribe<sup>a\*</sup>, Felipe Lembeye<sup>b</sup>

**ABSTRACT.** In dairy cattle, genetic selection for milk yield was generally based on 305 days lactation records that were calculated from available monthly test-day milk yield records. A test-day milk yield record, multiplied by the number of days between the current and following test-day, was the monthly milk yield and summed to all other monthly milk yields represented a 305 days lactation yield. Cows that for any reason did not complete their lactation got a 305 days yield via correction factors assuming a common lactation curve. Random regression models allow individual deviation from a common curve. The objective of this study was to estimate genetic parameters for daily milk yield using a random regression model (RRM) in Chilean dairy cows. A data set containing 97,683 monthly test-day records of 10,528 cows from 15 commercial dairy herds of Los Ríos Region in southern Chile was used. Days in milk (DIM) were modelled using the fourth-order Legendre polynomials and the model also included, as fixed effects, contemporary group and cow age at test-day as a covariate. The average daily milk yield was  $17.83 \pm 5.25$  kg. Average estimated heritability and repeatability from five to 305 DIM was  $0.26 \pm 0.02$  and  $0.61 \pm 0.04$ , respectively. The heritability estimate varied from 0.23 to 0.31. Both parameters did not vary dramatically except after 270 DIM when repeatability increased while heritability decreased. Although the estimated genetic parameters did not seriously depart from the most recent results available in the Chilean literature, they are mathematically more precise for estimating the true parameters than those calculated using adjustment factors, suggesting that the model used could be the starting point to develop a genetic evaluation system for dairy cattle in Chile.

*Key words:* milk, test day, random regression, genetic parameters.

### INTRODUCTION

Test-day models for the genetic evaluation of dairy yield traits were first used in Canada by Ptak and Schaeffer (1993) and in the case of somatic cells, they were used by Reents *et al* (1994). The task was to model test-day milk yields as a function of a known day in milk (DIM) fixed lactation curve. This procedure allowed a more accurate estimation of environmental effects by accounting for their influence in a particular DIM. Hence, the estimated breeding value of an animal in lactation includes information from all available test-day records and the estimation accuracy is therefore improved (Ptak and Schaeffer 1993). Another advantage of test-day models is that uncompleted lactation test-day records can also be used in the analyses without needing adjustment factors (Jensen 2001). However, test-day models using fixed parameters of DIM (lactation curve) assume that all cows follow the same shape of the chosen lactation curve, these do not allow for random individual cow deviations from the fixed lactation curve used in a particular model (Jamrozik and Schaeffer 1997). Test-day models using random regressors allow for individual cow departure from a common fixed lactation curve, they

were first implemented by Jamrozik and Schaeffer (1997) on a commercial basis for Canadian Holstein. The basic principle of random regression models applied to dairy cattle consists of fitting common to all animals' average lactation curves and specific curves describing individual random deviations from the average curve (Bohmanova *et al* 2008). Uribe (2001) provided a basic introduction of RRM applied to milk yield test-day records. Schaeffer (2004) thoroughly documented the theoretical aspects, application, and structure of RRM in animal breeding including test-day records.

Orthogonal polynomials (Kirkpatrick *et al* 1990) are widely used in RRM because they are easy to fit and the correlation among parameters has been proven low, albeit it has no biological meaning (Schaeffer 2004, Pool and Meuwissen 1999). The use of orthogonal polynomials to model lactation curves in genetic evaluations of dairy cattle has been implemented in New Zealand (Harris *et al* 2007), Germany, Canada, and the United Kingdom (Strabel *et al* 2005). On the other hand, Chile has not implemented a centralized dairy cattle genetic improvement program like other countries.

Some attempts have been made to estimate the genetic parameters of the Chilean Holstein population but none of them has used RRM (Elzo *et al* 2004, Montaldo *et al* 2017, Uribe *et al* 2017). The objective of this study was to estimate genetic parameters such as heritability ( $h^2$ ) and repeatability (*rep*) for daily milk yield, using orthogonal polynomials in a random regression test-day model in cows of Los Ríos Region, Chile.

Received: 09.04.2020.

Accepted: 24.06.2020.

<sup>a</sup>Departamento de Producción Animal, Facultad de Ciencias Agronómicas, Universidad de Chile, Santiago, Chile.

<sup>b</sup>Soprole S.A., Departamento Agropecuario, Gerencia de Materias Primas Lácteas, Santiago, Chile.

\*Corresponding author: H Uribe; hector.a.uribe@gmail.com

## MATERIAL AND METHODS

A data set containing 97,683 monthly test-day records of 10,528 first parity cows was used in this study and, from pedigree files, 2,350 ancestors without records were included. Data were gathered from 1996 to 2019 in 15 commercial dairy herds of Los Ríos Region, southern Chile. Cow's breed composition was predominantly Holstein Friesian although other dairy breeds and crosses are also part of the southern Chile dairy population. Unfortunately, the exact breed identification was not available in the data set.

Calving age was available in the data set, therefore, only heifers calving for the first time (from 20.5 to 40 months of age), were included in this study. Milk yield test-day records below five and above 35 kg of milk were deleted as well as records below six and above 305 DIM.

To account for DIM monthly milk yield test-day records were modelled using the fourth-order orthogonal polynomials, as described by Kirkpatrick *et al* (1990). Let  $y_t$  be the dependent variable kg of milk measured on day  $t$  of lactation, the polynomial equation can be written as:

$$y_t = b_0P_0 + b_1P_1 + \dots + b_nP_n \quad Eq. 1$$

Where  $b_i$  are the estimated regression coefficients and  $P_i$  is standardized to the unit of time orthogonal polynomial (Schaeffer 2004). Following Schaeffer (2004) the standardized trajectory ( $x_t$ ) chosen in this study, from six to 305 DIM, can be expressed from  $-1$  (six days) to  $+1$  (305 days) as follows:

$$x_t = -1 + 2\left(\frac{t - t_{min}}{t_{max} - t_{min}}\right) \quad Eq. 2$$

Where:  $t$  is the DIM of a given test-day record,  $t_{min}$  and  $t_{max}$  are six and 335 DIM, respectively. Thus, the standardized fourth-order plus the intercept orthogonal polynomials for a daily milk yield observation can be obtained as follows (Spiegel 1971):  $L_0(t) = 1$ ,  $L_1(t) = x_t$ ,  $L_2(t) = \frac{1}{2}(3x_t^2 - 1)$ ,  $L_3(t) = \frac{1}{2}(5x_t^3 - 3x_t)$ ,  $L_4(t) = \frac{1}{8}(35x_t^4 - 30x_t^2 + 15)$ .

The random regression animal model was:

$$y_{ijkl} = C_i + b_{ijkl}(Age) + \sum_{h=0}^4 \beta_{jh} L_{ijkl} + \sum_{h=0}^4 \alpha_{kh} L_{ijkl} + \sum_{h=0}^4 \lambda_{kh} L_{ijkl} + e_{ijkl} \quad Eq. 3$$

Where:  $y_{ijkl}$  = record  $l$  of cow  $k$  of herd  $j$  in contemporary group  $i$ .  $C_i$  fixed effect of the contemporary group  $i$ .  $b_{ijkl}$  fixed regression coefficient of daily milk yield on the age of the cow ( $Age$ ).  $\beta_{jh}$  = fixed regression coefficients within-herd  $j$ .  $\alpha_{kh}$  = genetic random regression coefficients for

animal  $k$ .  $\lambda_{kh}$  permanent environmental random regression coefficients for animal  $k$ .  $e_{ijkl}$  = random residual effect associated with  $y_{ijkl}$ .

In a particular contemporary group ( $C_i$ ) were included all cows tested in the same herd (15), year (24), and test-day season (3). The test-day season had three levels, with cows tested from March to June, July to October (spring), and November to February being levels one, two, and three, respectively.

In matrix notation the model can be written as:

$$y = Xb + Za + Wp + e \quad Eq. 4$$

Where vector  $b$  includes  $C_i$ ,  $b_{ijkl}$  and  $\beta_{jh}$ , vector  $a$  includes five random regression coefficients per each animal included in the analysis,  $p$  is a vector of five random permanent environmental coefficients per each cow with records, and  $e$  is a vector of heterogeneous residual effects. Matrices  $X$ ,  $Z$  and  $W$  are the incidence and covariate matrices. The (co)variance structure of the model was:

$$\begin{pmatrix} a \\ p \\ e \end{pmatrix} \sim N(0, V) \quad Eq. 5$$

With

$$V = \begin{pmatrix} G \otimes A & 0 & 0 \\ 0 & P \otimes I & 0 \\ 0 & 0 & R \end{pmatrix} \quad Eq. 6$$

The genetic covariance matrix of the five random regression coefficients is  $G$ , with elements denoted by  $g_{ij}$  for  $i$  and  $j$  going from 0 to 4.  $A$  is the additive genetic relationship matrix and  $\otimes$  is the Kronecker product function.  $P$  is a 5x5 permanent environmental covariance matrix with elements denoted by  $p_{ij}$ .  $I$  is an identity matrix of order equal to the number of cows with records.  $R$  is a diagonal matrix with elements that depend on DIM, let  $R = \text{diag}(\sigma_e^2)$ , following Jaffrezic *et al* (2000) and Tsuruta *et al* (2004) the  $i^{\text{th}}$  diagonal element that assumes the heterogeneous structure across DIM of  $R$  was estimated as:

$$\ln \sigma_e^2 = L'c = (L_0 \quad L_1 \quad L_2 \quad L_3 \quad L_4) \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} \quad Eq. 7$$

Where  $L$  is a vector of orthogonal polynomials for a given DIM and  $c_i$  is the natural log of the estimated residual coefficients.

Estimates of genetic and permanent environmental variances at DIM  $i$  were obtained as:

$$\sigma_{a_i}^2 = L_i' \hat{G} L_i \quad Eq. 8$$

$$\sigma_{p_i}^2 = L_i' \hat{P} L_i \quad Eq. 9$$

Where  $\hat{G}$  and  $\hat{P}$  are the estimated covariance matrixes for the random regression coefficients and permanent environmental elements, respectively, and  $L_i$  is a vector of orthogonal polynomials affecting DIM  $i$ . Likewise, the genetic covariance between any two DIM ( $\sigma_{a_{i,j}}$ ) were estimated as:

$$\sigma_{a_{i,j}} = L_i' \hat{G} L_j \quad \text{Eq. 10}$$

Where  $L_j$  is a vector of orthogonal polynomials affecting the  $j^{\text{th}}$  DIM.

Genetic correlations, for daily milk yield between the  $i^{\text{th}}$  and the  $j^{\text{th}}$  DIM ( $r_{a_{i,j}}$ ) were estimated as:

$$r_{a_{i,j}} = \frac{\sigma_{a_{i,j}}}{\sqrt{\sigma_{a_i}^2} * \sqrt{\sigma_{a_j}^2}} \quad \text{Eq. 11}$$

$h^2$  and  $rep$  at DIM  $i$  were calculated as:

$$h_i^2 = \frac{\sigma_{a_i}^2}{\sigma_{a_i}^2 + \sigma_{p_i}^2 + \sigma_{e_i}^2} \quad \text{Eq. 12}$$

$$rep_i = \frac{\sigma_{a_i}^2 + \sigma_{p_i}^2}{\sigma_{a_i}^2 + \sigma_{p_i}^2 + \sigma_{e_i}^2} \quad \text{Eq. 13}$$

To solve the linear model and estimate variance components the data were processed using the AIREML software<sup>1</sup>. For the final estimates, a value lower than  $10^{-10}$  of the squared differences between two consecutive solution estimates was defined as the convergence criterion.

## RESULTS AND DISCUSSION

The average DIM for all test-day records was  $152.50 \pm 84.19$  days while the same parameter, considering only the last test-day of each cow, was  $279.19 \pm 22.36$  days. The shortest lactation included in the study lasted 73 DIM, test-day records per cow ranged from one to 11 and the average was  $9.28 \pm 1.08$ .

Table 1 presents average, minimum and maximum estimated values across 300 days in milk for  $h^2$ ,  $rep$  and genetic, permanent environmental, and residual variances. Compared with Cobucci *et al* (2005) the pattern across DIM for genetic variances was similar except that in the latter the genetic variance steadily increased from 180 DIM toward the end of the lactation. However, Gebreyohannes *et al* (2016) and Bohmanova *et al* (2008) indicated that genetic variance consistently increased from the beginning to the end of lactation, the increment changed from 0.14 toward 1.0 ( $\ln(\text{kg/day})^2$ ) for 30 to 300 DIM, respectively.

The estimate of  $h^2$  across DIM reached its peak at 270 DIM (0.31) and fell to 0.24 at 305 DIM (table 2).

**Table 1.** Average estimate, standard deviation (SD) and minimum (Min) and maximum (Max) values throughout 300 days in milk for heritability ( $h^2$ ), repeatability ( $rep$ ), genetic ( $\sigma_a^2$ ), permanent environmental ( $\sigma_p^2$ ) and residual ( $\sigma_e^2$ ) variances.

Parameter	Average	SD	Min	Max
$h^2$	0.26	0.02	0.23	0.31
$rep$	0.61	0.04	0.58	0.78
$\sigma_a^{2*}$	3.89	0.33	3.34	4.95
$\sigma_p^{2*}$	5.14	0.80	4.42	8.73
$\sigma_e^{2*}$	5.85	0.79	3.58	7.20

\* =  $\text{kg}^2$ .

**Table 2.** Additive genetic ( $\sigma_a^2$ ), permanent environmental ( $\sigma_p^2$ ) and residual variances ( $\sigma_e^2$ ), heritability ( $h^2$ ), repeatability ( $t$ ), at selected days in milk (DIM), genetic correlation ( $r_{a_{i,j}}$ ) between the sixth and the other selected DIM, and the corresponding average daily milk yield (AMY).

DIM	$\sigma_a^{2*}$	$\sigma_p^{2*}$	$\sigma_e^{2*}$	$h^2$	$t$	$r_{a_{i,j}}^{**}$	AMY (kg)
6	4.95	8.14	7.20	0.24	0.65	1	16.44
30	3.83	5.79	7.07	0.23	0.58	0.85	19.79
60	4.10	5.63	6.64	0.25	0.59	0.62	20.30
90	4.09	5.09	6.17	0.27	0.60	0.53	19.41
120	3.80	4.71	5.86	0.26	0.59	0.50	18.72
150	3.48	4.61	5.72	0.25	0.59	0.50	18.01
180	3.34	4.48	5.72	0.25	0.58	0.49	16.96
210	3.57	4.43	5.74	0.26	0.58	0.45	16.69
240	4.13	4.67	5.58	0.29	0.61	0.39	16.19
270	4.39	4.93	5.00	0.31	0.65	0.37	15.61
305	3.80	8.73	3.58	0.24	0.78	0.46	15.40

\* =  $\text{kg}^2$ , \*\* = genetic correlation between the sixth and the corresponding ( $j$ ), in the first column, day in milk.

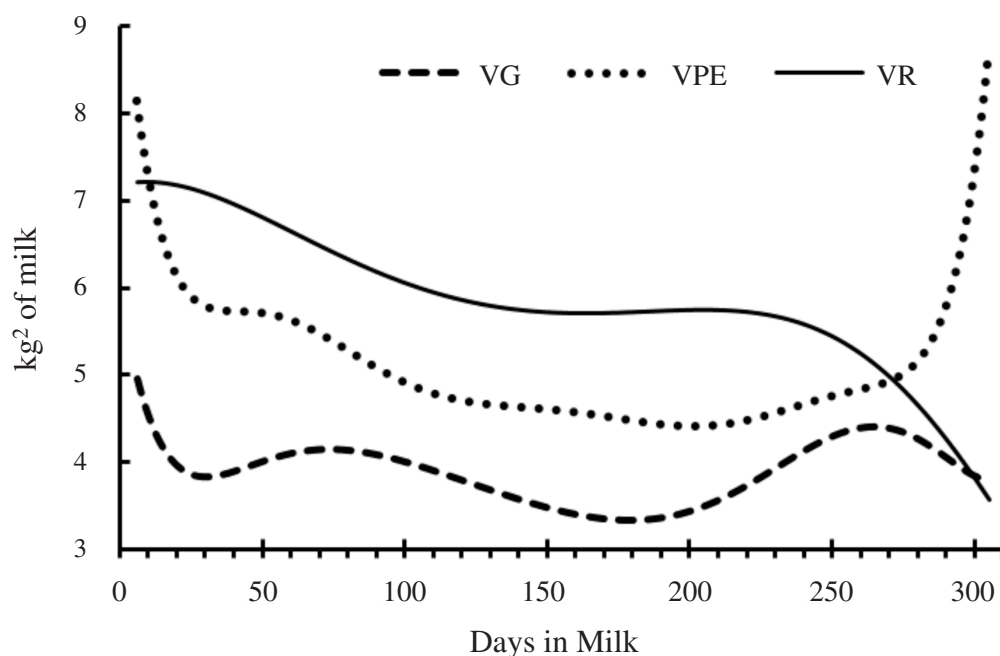
<sup>1</sup> Misztal I, Tsuruta S, Lourenco DAL, Masuda Y, Aguilar I, *et al*. 2018. Manual for BLUPF90 family programs. University of Georgia. <http://nce.ads.uga.edu/wiki/doku.php?id=documentation>

The estimates for heritability of Gebreyohannes *et al* (2016) varied from 0.17 and 0.42 for the first and the last test-day, respectively, and were similar to the  $h^2$  estimates reported in this study only for the first 50 DIM, and the authors partially attributed their high  $h^2$  estimate due to the degree of variability of the multi-breed population used in their research. Similarly, the  $h^2$  estimates of Cobuci *et al* (2005) varied from 0.15 to 0.31 and agreed with the results presented in this study. After ignoring the estimates of the first week of lactation, the  $h^2$  estimates of Bignardi *et al* (2011) varied from 0.20 to 0.35, their highest estimates were obtained between 196 and 273 DIM and their results are similar to those presented in this study, particularly the pattern followed by the  $h^2$  across DIM.

The average  $h^2$  estimate in this study was 0.26 (table 1) which is lower than the estimates reported by Elzo *et al* (2004) and Montaldo *et al* (2017) (0.29-0.34), but greater than that published by Uribe *et al* (2017) (0.16) for milk yield. The estimates of  $h^2$  reported in the present study are important because variance components calculated using a RRM are mathematically more precise for estimating the true parameters than those calculated using adjustment factors (Schaeffer 2004), which is the case of those previously published in Chile. The expected genetic change ( $\Delta G$ ) using estimates of a RRM, based in the formula proposed by Rendel and Robertson (1950) ( $\Delta G = i \times r_{A,\hat{A}} \times \sigma_g$ ), should be higher than the genetic progress based on Elzo *et al* (2004), Montaldo *et al* (2017), and Uribe *et al* (2017) because of their lower breeding value estimation accuracy ( $r_{A,\hat{A}}$ ).

Repeatability estimates across DIM fluctuate between 0.58 and 0.78 (table 1), the pattern across DIM is very similar to that followed by the permanent environmental variance (figure 1). Higher *rep* estimates are seen at the beginning and the end of the lactation, exactly as shown for permanent environmental variance estimates (figure 1). Gebreyohannes *et al* (2016) reported higher *rep* estimates that ranged from 0.84 at the beginning of the lactation to 0.94 at the last DIM, their trajectory pattern was equal to that of their  $h^2$  estimates and this was consistently increasing as lactation progressed. Other estimates of *rep* for daily milk yield were not found in the literature reviewed in this study.

Genetic correlations presented in table 2 between the sixth and some selected DIM were all positive and high between adjacent test-days and, as expected, they tended to decline as the distance between them increased. However, the genetic correlation among the sixth and above 265 DIM began to increase (data are partially shown in the seventh column of table 2). The lowest genetic correlation was 0.37, estimated between the sixth and DIMs from 260 to 266. Gebreyohannes *et al* (2016) reported a constant and declining genetic correlation trend across all over the lactation but at a lesser extent, for instance, their lowest estimated genetic correlation was 0.78 between the 30<sup>th</sup> and the 300<sup>th</sup> DIM. Bignardi *et al* (2011) also described that genetic correlations were high for adjacent test-days and decreased with increasing distance between them, however, these authors obtained negative genetic correlations between the initial and final test-days with the four models



**Figure 1.** Changes of the genetic (VG), permanent environmental (VPE) and residual (VR) variance estimates from 6 to 305 days in milk.

used in their research, irrespective of the function used to fit the lactation curve. Bignardi *et al* (2011) attributed those negative correlations to the difficulty of modelling initial test-day milk yields because in this period the cow suffers from post-calving stress and a negative energy balance. Jamrozik and Schaeffer (1997) and Cobuci *et al* (2005) also reported negative genetic correlations among test-day milk yields, however, in both studies the random regression coefficients were not included to model permanent environmental effects.

The objective of this study was to estimate genetic parameters for daily milk yield using a random regression test-day model and data from the Los Ríos region in Chile. Although the methodology was accessible in Chile since 2001 (Uribe 2001), its implementation using Chilean dairy records has not been reported. During the last two decades, available computing technology<sup>1</sup> and model fitting strategies have improved dramatically and some studies in which different functions and models were compared are available in the literature reviewed here (Cobucci *et al* 2005, Bignardi *et al* 2011, Gebreyohannes *et al* 2016). To model the lactation curve, orthogonal polynomials were chosen in this study because their properties and advantages, compared to other mathematical functions, have extensively been documented (Schaeffer 2004, Bohmanova *et al* 2008, Bignardi *et al* 2011). In this type of model, each animal gets five regression coefficient estimates corresponding to its random additive genetic, expression of breeding values for a given animal in its *i*<sup>th</sup> DIM as presented by Uribe (2001).

Test-day models using random regression have been implemented for routine genetic evaluation of dairy cattle in several countries, however, the Chilean dairy cattle industry has not implemented a public genetic evaluation system yet. The use of orthogonal polynomials in a random regression model to estimate heritability and repeatability in dairy cattle of Los Ríos Region, Chile, showed good results. Although the estimated genetic parameters did not radically depart from those reported in the Chilean literature, they are mathematically more precise, for estimating the true heritability and repeatability, than those calculated using adjustment factors suggesting that the model used in this study could be the starting point to develop a robust national genetic evaluation system for dairy cattle.

#### ACKNOWLEDGEMENTS

The authors would like to thank Dr. Ignacy Misztal, University of Georgia, USA, for allowing the use of the AIREML computer program. They also thank to Dr. Shogo Tsuruta, University of Georgia, USA, for providing valuable insights on fitting in the model heterogeneous residual variance.

#### REFERENCES

- Bignardi AB, El Faro L, Torres Júnior RAA, Cardoso VL, Machado PF, *et al*. 2011. Random regression models using different functions to model test-day milk yield of Brazilian Holstein cows. *Genet Mol Res* 10, 3565-3575.
- Bohmanova J, Miglior F, Jamrozik J, Misztal I, Sullivan PG. 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.
- Cobucci JA, Euclides RF, Lopes PS, Costa CN, Torres RdA, *et al*. 2005. Estimation of genetic parameters for test-day milk yield in Holstein cows using a random regression model. *Genet Mol Biol* 28, 75-83.
- Elzo MA, Jara A, Barría N. 2004 Genetic parameters and trends in the Chilean multibreed dairy cattle population. *J Dairy Sci* 87, 1506-1518.
- Gebreyohannes G, Koonawootrittriron S, Elzo M, Suwanasopee T. 2016. Estimation of genetic parameters using a random regression monthly test-day model in an Ethiopian dairy cattle population. *Agriculture and Natural Resources* 50, 64-70.
- Harris BL, Winkelman AM, Johnson DL, Montgomerie WA. 2007. Test-day model for national genetic evaluation of milk production traits. *Proceedings of the New Zealand Society of Animal Production* 67, 382-387.
- Jaffrezic FL, White MS, Thompson R, Hill WG. 2000. A link function approach to model heterogeneity of residual variances over time in lactation curve analyses. *J Dairy Sci* 83, 1089-1093.
- Jamrozik J, Schaeffer LR. 1997. Estimates of genetic parameters for a test-day model with random regressions for yield traits of first lactation Holsteins. *J Dairy Sci* 80, 762-770.
- Jensen J. 2001. Genetic evaluation of dairy cattle using test-day models. *J Dairy Sci* 84, 2803-2812.
- Kirkpatrick M, Madsen P, Bulmer M. 1990. Analysis of inheritance, selection, and evolution of growth trajectories. *Genetics* 124, 979-993.
- Montaldo H, Trejo C, Lizana C. 2017. Genetic parameters for milk yield and reproduction traits in the Chilean Dairy Overo Colorado cattle breed. *Cien Inv Agr* 44, 24-34.
- Pool MH, Meuwissen THE. 1999. Prediction of daily milk yields from limited number of test days using test day models. *J Dairy Sci* 82, 1555-1564.
- Ptak E, Schaeffer LR. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest Prod Sci* 34, 23-34.
- Reents R, Dekkers JCM, Schaeffer LR. 1994. Analysis of test day for somatic cells from dairy cows. *J Anim Sci* 72, 267 (Abst.).
- Rendel JM, Robertson A. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J Genet* 50, 1-8.
- Schaeffer LR. 2004. Application of random regression models in animal breeding. *Livest Prod Sci* 86, 35-45.
- Spiegel MR. 1971. *Advanced mathematics for engineers and scientists*. McGraw-Hill, New York, USA.
- Strabel T, Szyda J, Ptak E, Jamrozik J. 2005. Comparison of random regression test-day models for Polish Black and White cattle. *J Dairy Sci* 88, 3688-3699.
- Tsuruta S, Misztal I, Lawlor TJ, Klei L. 2004. Modeling final scores in US Holsteins as a function of year of classification using a random regression model. *Livest Prod Sci* 91, 199-207.
- Uribe H. 2001. Test day model: A new statistical tool for genetic evaluation of dairy cattle. *Agr Tec* 61, 500-511.
- Uribe H, González H, Gatica C. 2017. Genetic parameter estimation to milk yield and fat and protein yield deviated from 3% of concentration in milk, in dairy herds of southern Chile. *Austral J Vet Sci* 49, 71-76.

