

Modeling lactation curves of French dairy goats using random regression models

M.Arnal^{1,2}, H.Larroque¹, H.Leclerc³, V.Ducrocq⁴ & C.Robert-Granié¹

¹ GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 Castanet-Tolosan, France

² Idele, Chemin de Borde Rouge, 31326 Castanet-Tolosan cedex, France

³ Institut de l'Élevage, INRA UMR1313 GABI, 78350 Jouy-en-Josas, France

⁴ UMR1313 GABI, INRA, AgroParisTech, Univ. Paris-Saclay, 78350 Jouy-en-Josas, France
mathieu.arnal@idele.fr (Corresponding Author)

Summary

Random regression models are widely used to analyze longitudinal data in genetic evaluation. They permit to evaluate more precisely the environmental effects and the genetic value of an animal, by taking into account the test-day specific effects. Our objective was to implement a random regression model for the evaluation of French Saanen goats. The data consisted of milk test-days (TD) records from first lactation of Saanen goats. B-splines were used to model the environmental factors. The genetic and permanent environment effects were modeled through two Legendre polynomials of the same order. The goodness of fit and the genetic parameters being function of the order of the polynomials, orders 0, 2 and 4 were tested. Models with Legendre polynomials of order 4 were found to be the most parsimonious. A rank reduction of the variance covariance matrix was performed by eigenvalues decomposition in order to reduce computing time and complexity. With a reduction to rank 2, the first two eigenvectors well summed up the genetic variability of milk yield level and persistency.

Keywords: random regression model, reduced rank, test-day records, dairy goat, genetic parameters estimation

Introduction

Random regression test days models (RRM) are widely used in official genetic evaluations of test-day milk production (TD) in dairy cows. They increase the accuracy of breeding value predictions and fit more precisely the environmental effects. Besides describing the variability of genetic parameters along the lactation curve, the RRM allows the prediction of breeding values (EBV) for lactation persistency, comparing EBV estimated at any given time during lactation period. Lactation persistency, which is of interest to dairy producers, is defined here as the animal ability to maintain a high level of milk yield after reaching the lactation peak. In this study, French dairy goat TD milk production records were used to model the environmental effects on the lactation curve via B-spline functions and to estimate the genetic parameters for TD in a RRM using orthogonal Legendre polynomial functions with orders 0, 2 and 4. Because RRM are computationally expensive, Legendre polynomial functions have the computational advantage of reducing the correlation between estimated coefficients, thus facilitating convergence (Schaeffer and Jamrozik, 2008). They are also flexible, enabling the fitting of curves of different shapes. With eigendecomposition, it is possible to reduce the rank of the variance-covariance matrix to decrease computing time and estimation noise. The aim of this study was to test several RRM with different orders for Legendre polynomial functions with or without rank reductions to estimate genetic parameters for milk yield and to compute TD-EBV for French dairy goats.

Material and methods

Data: The data consisted of 400,557 milk test-day (TD) records from 57,747 first lactation Saanen goats from the French North-West region collected between 1995 and 2015. The pedigree consisted of 132,309 animals. Each lactation included at least 4 TD between the 7th and the 300th day in milk (DIM). The lactation had to last, at least, 180 to 350 days. Goats were milked twice a day and their records were summed to obtain their daily production. More than 4 animals by herd and test date were required. The sire of goats had to be an artificial insemination buck and to have at least 20 progeny in the dataset. The dams of the goats had to be known.

Random regression models (RRM): Several RRM were tested, all of the form:
$$y_{ijklm} = \mu + \pi_i + \delta_j + \gamma_k + \eta_l + a_m + p_m(t) + e_{ijklm}$$

where y_{ijklm} is the milk yield of goat m , within herd-test-date i (HTD_i), in subclass j for age at kidding (7 classes from 9 to 29 months), subclass k for month at kidding (11 classes), subclass l for the gestation stage at 300th DIM (16 classes), and a_m and $p_m(t)$ are the random additive genetic value and permanent environmental effects of animal m for the o th term of a Legendre polynomial of order q respectively, $P_o(t)$ is the value of the o th term of the Legendre polynomial at time t (DIM standardized from -1 to 1) and e_{ijklm} is the residual term. $p_m(t)$ was chosen to be a uniform cubic B-spline with 4 equidistant knots for all models tested because it was the best function to model the fixed effects (results not shown) in agreement with results of Druet et al. (2003). The model assumed a heterogeneous residual variance modeled with the fixed effect of 10 classes each 30 days in length.

Legendre polynomials of order 0, 2 and 4 were tested (leg0, leg2 and leg4) with the same order for both the random genetic and environmental effects. The EBV of goat m for a complete lactation from TD yields was obtained by summing the EBV for each DIM (called SUM_legx). Because the RRM facilitates the estimation of the shape of the lactation curve for each goat, EBV for persistency (denoted PERS_legx) can be computed as the additional genetic yield from the 60th DIM to 280th DIM relative to an average animal having the same yield on 60th DIM (Jamrozik et al., 1997).

Some rank reductions of the genetic and permanent environmental covariance matrices were tested using an eigen-decomposition of the covariance matrices. For a reduction to rank z , the first z eigenvectors obtained with the complete model were multiplied by the coefficients of Legendre polynomial as shown in Druet et al. (2003). The two matrices were reduced to the same order. We denote legxRz the Legendre polynomial function of order x with a covariance matrix reduced to rank z .

Multiple-trait model (MT): The evolution of the heritability over DIM, and genetic correlations between periods obtained with RRM were compared to those obtained with a MT model where 6 periods were created (1=[7,50], 2=[51,100], 3=[101,150], 4=[151,200], 5=[201,250], 6=[251,300]). TD on each period n ($n=1, \dots, 6$) were considered to be different traits with this MT model:

where y_{ijklm} is the average milk yield of the goat m for the period n ($n=1, \dots, 6$), in subclass i for age at kidding, subclass j for month at kidding, subclass k for the gestation stage at 300th DIM, and class l for herd-year (HY). The fixed effects are the same as the previous RRM model. a_m is the additive genetic breeding value of animal m for the n th period. The corresponding EBV of goat m from RRM was obtained by summing the EBVs for each DIM over the same period. Genetic parameters were estimated using the WOMBAT software (Meyer, 2007)

Model selection criteria: The goodness-of-fit of the models was assessed by comparing the

Bayesian Information Criterion (BIC) and the Mean Square Error (MSE) for each model.

Results and Discussion

Full rank models: BIC and MSE decreased rapidly with the increase in Legendre polynomial order from 0 to 2 (Table 1), the decay was smaller between orders 2 and 4 although 18 parameters were added. These results were consistent with Pool *et al.* (2000). The heritability estimated with the MT model increased up to maximum (0.31) in the middle of the lactation (Figure 1) and then decreased as reported by Pool *et al.*(2000). All models showed the same evolution of the heritability as the MT model. However, leg0 presented a smaller heritability after 100 days in milk. The mean levels of heritability were in accordance with those reported in goats by Zumbach *et al.* (2008) but higher than those obtained by Menéndez-Buxadera *et al.* (2010) (around 0.2). The shape of heritability curve was close to those reported in dairy cattle (Druet *et al.*, 2003) with a maximum heritability in the middle of the lactation, in contrast to the two previous studies.

The mean of the absolute difference between the correlation matrix of EBV from MT model and the correlation matrices of EBV from RRM by period varied from 0.14 with leg0 to 0.004 with leg4, showing that leg4 was very close to MT model. All the genetic correlations between two periods with either the MT or RRM models were positive (the weakest was 0.51) indicating that the selection of animals at any day of lactation will, on average, give positive responses to all other points of lactation curve. According to all BIC and MSE, leg4 appeared to be the best model to fit our data.

Reduced RRM (RRM*): In the leg4 model, the first 3 eigenvalues represented respectively for the genetic covariance matrix 87, 10 and 2%, (i.e. 99% of the total genetic variance) and for the permanent environmental covariance matrix, 76, 14 and 6%, (i.e. 96% of the total variance). The data fit was better when the first three principal components were kept (Table 1): BIC (expressed as a deviation of the best value) varied from -8,229 for leg4R2 to -1,924 for leg4R3. The leg4R2 model showed the highest residual variances. The main differences of residual variances between the two reduced models were located at the beginning and at the end of lactation (results not shown), indicating that leg4R3 was the best RRM*.

Nevertheless, the adjustments of the two RRM* were much better than the model assuming a constant genetic value throughout the lactation (leg0: BIC=- 44,174), which is similar to the model used nowadays for the genetic evaluation in France based on a total lactation milk yield.

Table 1 : Number of parameters, BIC and MSE for the different models tested

	Model	Number of parameters	Δ BIC ¹	MSE
Non-reduced RRM	leg0	12	-44,174	13.2
	leg2	22	-3,926	7.1
	leg4	40	0	5.8
Reduced RRM (RRM*)	leg4R2	16	-8,229	8.7
	leg4R3	22	-1,924	7

¹Values are expressed as deviations from the best value (leg4).

The evolution of the heritability over DIM with the RRM* was compared with MT model (Figure 1). The RRM* were close to MT, leg2 and leg4, and better than leg0. As expected, leg4R3 was closer to leg4 than leg4R2.

The first eigenvector was almost constant through DIM representing the average lactation

production whereas the second one was varying linearly explaining the persistency (Figure 2) as found by Van der Werf et al. (1998) and Druet et al. (2003). Correlations between the EBV of the sires obtained for the first dimension of RRM* and the EBV SUM_leg4 (computed from leg4 model) was equal to 1 for leg4R2 and leg4R3. Correlations between the EBV of the sires obtained for the second dimension of RRM* and the EBV PERS_leg4 was maximum for leg4R3 (0.98) and slightly lower for leg4R2 (0.95). Therefore, even leg4R2 allows a differentiated selection of the animals based on the two first components (total milk yield and persistency) as shown by Druet et al. (2003). For the future, with models combining milk production, fat percentage, protein percentage, and others parities, a model with only 2 components like leg4R2 seems most appropriate.

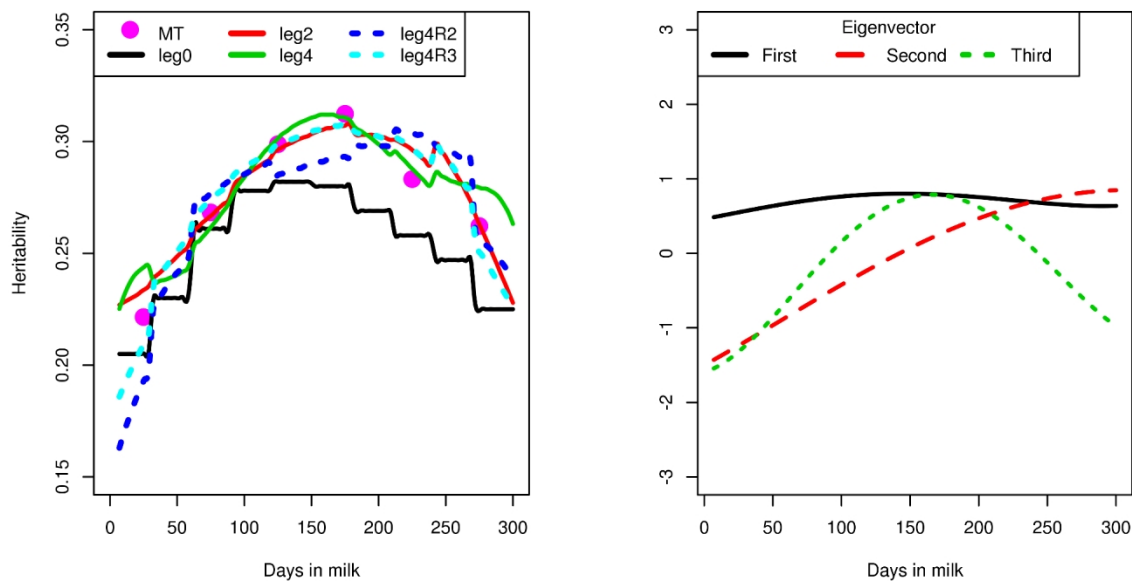


Figure 1: Evolution of the heritability according to the DIM for the non-reduced models, the reduced models and the MT model. Figure 2: Coordinates of the first three eigenvectors of the TD genetic covariance matrix computed with leg4 model.

Conclusion

In French Saanen dairy goat, the best random regression model to fit test-day milk production is obtained with a quartic Legendre polynomial to model genetic and permanent environmental effects. The rank reduction to the first two eigenvectors affects the quality of model adjustment but it summarizes well the level of the average lactation potential of an animal and its persistency.

The first author acknowledges financial support from the ANRT and Apis-Gene.

References

- Druet, T., Jaffrézic, F., Boichard, D., Ducrocq, V., 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy Sci.* 86, 2480–2490.
- Jamrozik, J., Schaeffer, L.R., Dekkers, J.C.M., 1997. Genetic evaluation of dairy cattle using test day yields and random regression model. *J. Dairy Sci.* 80, 1217–1226.
- Menéndez-Buxadera, A., Molina, A., Arrebola, F., Gil, M.J., Serradilla, J.M., 2010. Random regression analysis of milk yield and milk composition in the first and second

- lactations of Murciano-Granadina goats. *J. Dairy Sci.* 93, 2718–2726.
- Meyer, K., 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ.-Sci. B* 8, 815–821.
- Meyer, K., Kirkpatrick, M., 2005. Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. *Genet. Sel. Evol.* 37, 1.
- Pool, M.H., Janss, L.L.G., Meuwissen, T.H.E., 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. *J. Dairy Sci.* 83, 2640–2649.
- Schaeffer, L.R., Jamrozik, J., 2008. Random regression models: a longitudinal perspective. *J. Anim. Breed. Genet.* 125, 145–146.
- Van der Werf, J.H.J., Goddard, M.E., Meyer, K., 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *J. Dairy Sci.* 81, 3300–3308.
- Zumbach, B., Tsuruta, S., Misztal, I., Peters, K.J., 2008. Use of a test day model for dairy goat milk yield across lactations in Germany. *J. Anim. Breed. Genet.* 125, 160–167.