Pre-Adjusting Data for the Shape of the Lactation Curves in Test-Day Model

H. Leclerc, D. Duclos, A. Barbat, T. Druet and V. Ducrocq
UR337 INRA, Station de Génétique Quantitative et Appliquée, 78 352 Jouy-en-Josas, France

Abstract

A study on environmental effects included in a test-day model genetic evaluation revealed that the shape of lactation curves did not change over years. This feature allows a 2-step approach where test-day records are pre-adjusted in a first step for time independent fixed effects (shape of lactation curves). Random effects and year dependent fixed effects are estimated in a second step. Genetic and permanent environment effects estimated in such a 2-step procedure were found to be virtually identical to those from a 1-step procedure. Two main advantages can be put forward from this approach: it requires less computing time and less computer memory.

Introduction

During the last decade, a lot of interest has been expressed about test-day models (TDM) and many countries have already implemented such models in routine genetic evaluations. Genetic evaluations based on TDM present many advantages compared with 305-day lactation models as mentioned in several reviews (in particular Swalve, 2000 and Jensen, 2001). In addition to well known benefits from a genetic viewpoint, TDM are useful for prediction and management purposes, but only a few published studies specifically refer to these aspects (Mayeres *et al.*, 2004).

Countries which already moved from a lactation model to a test-day one have developed their own TDM for routine genetic evaluations for their own population. These models differ in the way the lactation curve is modelled as a function of days in milk (DIM) – with fixed classes, parametric or semi-parametric (splines) curves – in the way the genetic and permanent environment components are described (fixed or random regression using Legendre polynomials or others), and in the way heterogeneous residual variances are accounted for. In France, regression splines (White *et al.*, 1999) were used to model lactation curves (Druet *et al.*, 2003).

In a recent study, Leclerc *et al*. (2007) showed that the shape of the lactation curves depends on parity, age at calving, month of calving, length of dry period, gestation, but does not depend on the year of calving. Therefore, the effect of calving year can be modelled just as a constant over the whole lactation. This feature makes it possible to consider a 2-step approach where test-day data are first pre-adjusted for time independent fixed effects (describing the shape of the lactation curve) before a regular TDM is fitted.

The aim of this paper was to implement the 2-step procedure mentioned above to investigate the limits and opportunities of such an approach and to compare it with a 1-step procedure in a large dataset test-day model genetic evaluation, in particular in terms of computing time and requirements.

Material & Methods

Data

Data were test-day (TD) yields of milk, fat and protein, fat and protein percentages for the first three lactations of French Montbéliarde (with at least 7/8 of Montbéliarde blood) collected between September 1988 and December 2005. Days in milk (DIM) ranged from 7 to 335 days. Cows were required to have a first-lactation record and known sire and dam to be included in the data file. Only cows with at least three TD were considered. To define a herd by test-date effect (HTD), at least five records were required.

To reduced computational requirement, most of the analyses were made using a regional subset of the full dataset, considering 2.5 million TD, hereafter referred to the "Jura" dataset.

Model

$$y_{ijklnopqtt'vw} = HTD_i + ymonth_j + ydry_k + yage_l + spl_{DIM}(month_n, t) + spl_{DIM}(dry_o, t) + spl_{DIM}(age_p, t) + spl_{DCC}(parity_q, t') + \sum_{r=1}^{4}(\upsilon_{tr} a_{vr}) + \sum_{r=1}^{4}(\zeta_{tr} p_{vr}) + \sum_{r=1}^{6}(\tau_{tr} h_{wr}) + e_{ijklnopqtt'vw}$$

where $y_{ijklnopqtt'vw}$ is the record of animal *v* in lactation *q* on DIM *t* and days carried calf (DCC) *t'*. HTD_i is the i^{th} herd-test-date effect. The interaction terms ymonth, ydry and yage correspond to region-year-parity class for calving month, length of dry period and calving age respectively. These terms are constant over the lactation. $spl_{var}(eff)$ denotes a regression spline function of the variable *var* specific to the effect *eff*. Four spline functions were included to model lactation curves according to the effects region-parity class of calving month (month), region-parity class of dry period length (dry) and region-parity class of calving age (age) as a function of DIM *t*, spl_{DCC} is a regression spline modeling the effect of gestation of region-parity m as a function of DCC *t'* and e is the residual term.

The random additive genetic, permanent environment and herd-year of calving-parity random regression effects are a_{vr}, p_{vr} and h_{wr} respectively for animal *v* and herd-year of calving-parity *w* with associated coefficients υ_{tr}, ζ_{tr} and τ_{tr}, at DIM *t*.

Environmental fixed effects

Curves for month of calving, age at calving and length of dry period were modelled using 6-knot regression splines (White *et al*., 1999), with knots positioned at DIM 7, 20, 50, 135, 245 and 335. DCC, defined as DIM minus the number of days from calving to successful insemination, was used to model the effect of gestation. When DCC was smaller than 100 days, the effect of gestation on the analysed traits was ignored. The DCC effect ranged from 100 to 265 days. Four-knot regression splines with knots at DCC 100, 150, 200 and 265 (Druet *et al*., 2003) were used to model the effect of gestation. These four effects were assumed to be independent and therefore the overall fixed part of the lactation curve of an animal can be described as the sum over the relevant effects of elementary curves.

The definition of fixed effect classes was based on the current national evaluation. Three regions were defined in the national data file, based on seasonal patterns and feeding systems. Twelve seasons of calving were defined based on calendar months. Most of calving occurred during the autumn (Sept-Dec = 83.5%). For the calving age effect, 24 age classes were formed: ten for first lactation (22-23, 24-25, …, ≥ 40), seven for second (≤ 36, 37-39, …, ≥ 52), and seven for third lactation (≤ 48, 49-51, …, ≥ 64). For first lactation cows, the median class was 34-35 months corresponding to an extensive management of heifers. For length of dry period, five classes were created for second and third lactation cows: ≤ 25 days, 26 - 45 days, 46 - 60 days, 61 - 75 days, and ≥ 76 days. A specific class was created for heifer effects (first lactation cows).

Genetic, permanent environment and herd by year effects

The phenotypic (co)variance matrix V of the random effects was:

$$\mathbf{V} = \begin{pmatrix} \mathbf{G}\otimes\mathbf{A} & 0 & 0 & 0 \\ 0 & \mathbf{P}\otimes\mathbf{I} & 0 & 0 \\ 0 & 0 & \mathbf{H}\otimes\mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{R} \end{pmatrix}$$

where **G**, **P** and **H** were the first three lactations random regression covariance matrices initially modelled by parity for the genetic, permanent environment and herd-year of calving-parity effects. The matrices were estimated previously (Druet *et al.*, 2005). Their ranks were reduced from six to four for G and P and from nine to six for H. For the six traits, the four eigenvectors explained between 98.4 and 99.7% of the total genetic variation in the three lactations, and between 87.6 and 91.3% for the permanent environment variance. For the herd-year of calving-parity variance, the first six eigenvectors explained from 96.6 to 99.7% of the total variance. **A** was the additive genetic relationship matrix, and **R** was the residual variance changing continuously with DIM (Druet *et al.*, 2005). Other sources of residual variance heterogeneity, e.g due to herd-year effects were not accounted for, although they will be modelled in the national genetic evaluation. Phenotypic variance was ensured to remain the same before and after rank reduction by adding to **R**, the loss of variance in **G**, **P** and **H** due to rank reduction.

Practical approach

To reduce computational requirements and computing time, data were first pre-adjusted for effects related to the shape of the lactation curve with no year interaction (parity class of calving month, parity class of dry period length and parity class of calving age with 6-knot regression splines each and gestation effect with 4-knot regression splines, i.e. 22 splines coefficients) running a model including only fixed effects. In a second step, the model included the HTD effect, the three fixed effects including year interaction (ymonth, ydry and yage) and random additive genetic, permanent environment and herd-year of calving-parity regression coefficients, i.e. 18 effects instead of 40 in a 1-step procedure.

For practical implementation of the pre-adjustment, one element of interest was the stability of regression splines over years and with data files including more and more observations. If splines coefficients really do not depend on year, lactation curves estimated should be consistent over time. This property was checked using the complete “Jura” file, hereafter referred to “Jura 2005” and a subset of it, truncated 4 years earlier (test-day data collected between September 1988 and December 2001), called “Jura 2001”. In addition to correlations between splines coefficients, both the genetic and permanent environment values estimated from the Jura 2005 data file pre-adjusted for lactation curves estimated from Jura 2001 on the one hand and Jura 2005 on the other hand were compared (Figure 1).

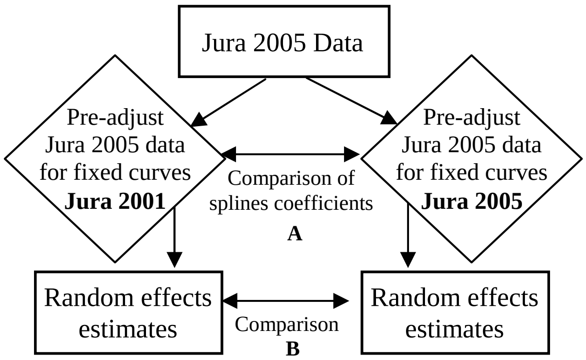

Figure 1. Approach implemented to check the impact of splines coefficients estimated at a 4-year interval.

Moreover, it was important to know whether the data pre-adjustment for lactation curve induces a variation of the random effects solutions of the model. Therefore genetic and permanent environment solutions obtained with a 1-step procedure were compared to the ones obtained with the 2-step procedure for the Jura 2005 data file (Figure 2).

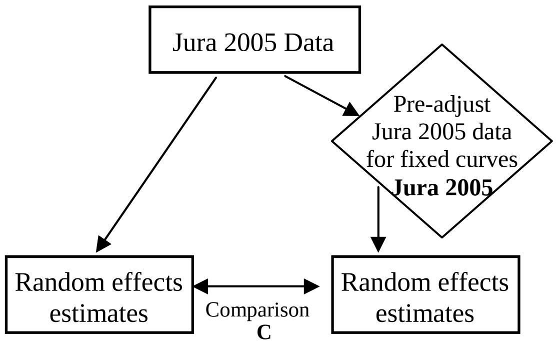

Figure 1. Approach implemented to assess the impact of a 2-step *vs* a 1-step evaluation.

Results and discussion

Impact of splines coefficients estimated at different dates

The comparison of the splines coefficients obtained from Jura 2001 on the one hand and from Jura 2005 in the other hand (comparison A in Figure 1) did not always give very high correlations (between 0.768 and 0.999, with a median at 0.986), When two corresponding splines were plotted, the shapes were not exactly the same, but it was difficult to determine if the differences were really significant over the whole lactation.

However, what actually matters are the sum of these curves and its effect on estimated random effects. The comparison of the solutions obtained for genetic and permanent environmental effects with the second step of the model on Jura 2005 data pre-adjusted for time independent fixed effects (shape of the lactation curve) from Jura 2001 on the one hand and from Jura 2005 on the other hand (comparison B in Figure 1) gave much better results. Correlations for each of the four genetic effects were between 0.99869 and 0.99998. and were between 0.99916 and 0.99993 for each of the four permanent environment effects (results not shown). When the solutions were brought back to a lactation value (one to three), all the correlations were between 0.99988 and 0.99997 for genetic effect and between 0.99948 and 0.99996 for permanent environment effect (Table 1). To sum up, genetic and permanent environment values were not modified when the data were pre-adjusted with splines coefficients calculated 4 years before.

Table 1. Within lactation correlations for genetic effects and permanent environment effects estimated with data pre-adjusted with splines coefficients obtained from two data files at a 4-year interval (Jura 2001 and Jura 2005) for Milk, Fat % and content, Protein % and content.

	Genetic		Permanent environment	
	Min	Max	Min	Max
Milk	0.99991	**0.99997**	0.99969	0.99994
Fat %	0.99997	**0.99997**	0.99978	0.99993
Fat	0.99988	0.99995	0.99985	**0.99996**
Pro %	0.99992	0.99994	0.99984	0.99992
Pro	**0.99988**	0.99993	**0.99948**	0.99988

Impact of 2-step evaluation vs 1-step evaluation

Very high correlations were obtained for both genetic and permanent environment effects (comparison C in Figure 2). Correlations for each of the four random effects were between 0.9981 and 0.9999 for genetic effect and between 0.9988 and 0.9999 for permanent environment effect (results not shown). When the solutions were brought back to lactation values (one to three), all the correlations were between 0.9984 and 0.9999 for genetic effect and 0.9974 and 0.9999 for permanent environment effect (Table 2). Furthermore, similar correlations were obtained for milk yield analysed with the complete Montbéliarde data file (i.e., with 21.1 millions of TD) with a minimum lactation value of 0.9996.

Table 2. Within lactation correlations between genetic effects and between permanent environment effects estimated with a 1-step and a 2-step evaluation on Jura 2005 for Milk, Fat % and content, Protein % and content.

	Genetic		Permanent environment	
	Min	Max	Min	Max
Milk	**0.9984**	0.9992	0.9981	0.9991
Fat %	0.9999	**0.9999**	0.9998	**0.9999**
Fat	0.9989	0.9995	0.9995	0.9997
Pro %	0.9995	0.9996	**0.9974**	0.9992
Pro	0.9992	0.9996	0.9981	0.9994

Conclusion

This study confirmed that it was possible to pre-adjust test-day records for time-independent fixed effects, i.e. lactation curves, in a first step, and to estimate random effects and year-dependent fixed effects in a second step. Genetic and permanent environment effects obtained for the two-step procedure were virtually identical to those from a one-step procedure. Two main advantages can be put forward from this approach: it required less computing time (here reduced by a factor four) and less memory (26 and 18 effects for first and second step respectively instead of 40 effects in a 1-step procedure). Moreover, the first step (computation of the splines coefficients for fixed effects) could be performed at regular intervals instead of at each evaluation..

Acknowledgements

The research was partly funded by a grant from the French Ministry of Agriculture (Action Innovante UTILEG). This study was realized within the framework of the Unité Mixte Technologique INRA – Institut de l'Elevage dealing with cattle genetic evaluation.

References

Druet, T., Jaffrézic, F., Boichard, D. & Ducrocq, V. 2003. Modeling Lactation Curves and Estimations of Genetic Parameters for First Lactation Test-Day Records of French Holstein Cows. *J. Dairy Sci. 86,* 2480-2490.

Druet, T., Jaffrézic, F. & Ducrocq, V. 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. *Genet Sel. Evol. 37,* 257-271.

Jensen, J. 2001. Genetic evaluation of dairy cattle using test-day models. *J. Dairy. Sci. 84,* 2803-2812.

Leclerc, H., Duclos, D. & Ducrocq, V. 2007. Environmental effects on lactation curves includes in a test-day model genetic evaluation. *Proc. 58th EAAP*. 10:34.

Mayeres, P., Stoll, J., Bormann, J., Reents, R. & Gengler, N. 2004. Prediction of daily milk, fat and protein production by a random regression test-day model. *J. Dairy. Sci. 87,* 1925-1933.

Swalve, H.H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. *J. Dairy. Sci. 83,* 1115-1124.

White, I.M.S., Thompson, R. & Brotherstone, S. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci. 82***,** 632-638.