Fit of different functions to the individual deviations in random regression test day models for milk yield in dairy cattle

N. P. P. Macciotta[1], F. Miglior[2], A. Cappio-Borlino[1], L. R. Schaeffer[3]

[1] Dipartimento di Scienze Zootecniche, Università di Sassari, Italy
[2] Agriculture and Agri-Food Canada CDN. Guelph, Canada
[3] Department of Animal and Poultry Science, University of Guelph, Canada

Corresponding author: Nicolò P.P. Macciotta. Dipartimento di Scienze Zootecniche, Università degli studi di Sassari, Via De Nicola 9, 07100 Sassari, Italy - Tel. +39 079 229298 - Fax: +39 079 229302 - Email: macciott@uniss.it

ABSTRACT: The shape of individual deviations of milk yield for dairy cattle from the fixed part of a random regression test day model (RRTDM) was investigated. Data were 53,217 TD records for milk yield of 6,229 first lactation Canadian Holsteins in Ontario. Data were fitted with a model that included the fixed effects of herd-test-date, DIM interval nested within age and season of calving. Residuals of the model were then fitted with the following functions: Ali and Schaeffer 5 parameter model, fourth-order Legendre Polynomials, and cubic spline with three, four or five knots. Result confirm the great variability of shape that can be found when individual lactation are modeled. Cubic splines gave better fitting pe4rformances although together with a marked tendency to yield aberrant estimates at the edge of the lactation trajectory.

INTRODUCTION – Random regression models (RRM) are rapidly becoming the reference models for genetic evaluations of dairy cattle in several countries (Schaeffer, 2004). In RRM, a fixed part includes effects peculiar to all cows on the same test day and effects specific to a particular cow on a given test day, such as pregnancy or disease, plus a factor accounting for the general shape of the lactation curve (Ptak and Schaeffer, 1993). Individual deviations from the fixed part of the model are fitted by random regression coefficients (Jamrozik and Schaeffer, 1997). The choice of functions to model fixed and random lactation curves is a critical point in the construction of a RRM. Evaluation of different models is usually based on fit diagnostics, predictive ability, correlations and scale of parameters (Lopez-Romero *et al.*, 2003; Liu *et al.*, 2006; Misztal, 2006) . In this study, five of the most used functions used to in RRM were evaluated for their ability to fit individual deviations around the mean curve.

MATERIAL AND METHODS – Data were 53,217 TD records for milk yield of 6,229 first lactation Canadian Holsteins of the province of Ontario. Edits were on the number of TD records per cow (>7), calving year (>2001), days in milk (between 5 and 305), age at calving (between 22 and 31 months), milk yield (between 1.5 and 90 kg), DIM at first test (<50) and number of records within a herd test date (>14). Data were analyzed with a linear model that included the fixed effects of herd-test date (4,272) DIM interval (60 of 15 days each) nested within calving season (2 seasons) and age at calving (22-31 months). Individual patterns of residuals, treated as a new variable (Z), were fitted with the following models: the Ali and Schaeffer multiple regression (AS), a fourth order Legendre Polynomials (LP4), a cubic Spline regression model (CSPL) with three (CSPL3), four (CSLP4) or five (CSPL5) knots. The AS is a parametric model that has been used in the early RRMs whereas the more flexible Legendre orthogonal polynomials represent the reference model both for fixed and random curves. Finally, cubic splines have been recently suggested for their great flexibility and numerical properties. Goodness of fit of different models was assessed by examining the distribution of fits among five classes of adjusted R-squared (ADJRSQ). Moreover, different models were compared on the basis of the standard deviation of residuals.

RESULTS. The AS function classified about 90% of individual Z patterns into two shapes that are reported in figures 1a and b whereas LP4 resulted in a balanced distribution of curves among classes of different combinations of parameter sign.

Figure 1a. Average Z pattern detected in about 48% of cases by the AS function.
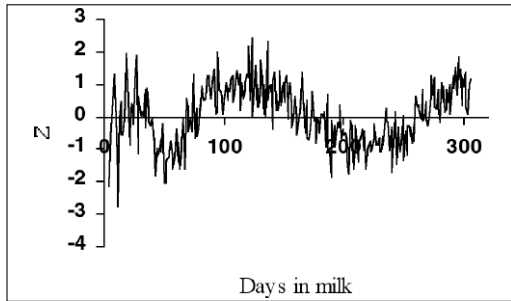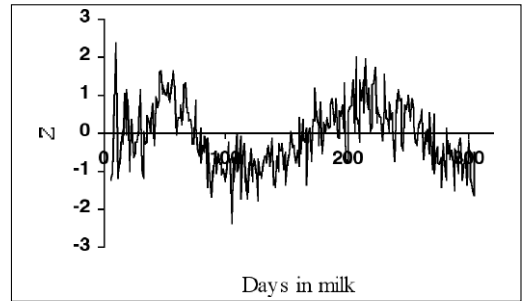
Figure 1b. Average Z pattern detected in about 43% of cases by the AS function.

Days in milk

Days in milk

The fitting of Z was rather poor with about 30-40% of curves showing an ADJRSQ lower than 0.20 in all the models. However, some differences between the functions can be highlighted by considering both the distributions of fits among ADJRSQ classes and the magnitude of standard deviation of residuals of the different models (Table 1). The number of curves showing an ADJRSQ>0.80 is almost double for all the three types splines in comparison with AS and LP4. This behavior was confirmed by a parallel reduction in the standard deviation of residuals. No substantial differences were found in adjusted R-squared values obtained by fitting Z with splines with three, four or five knots. This results could be ascribed to the data structure (Meyer, 2005): actually, the distribution of data across days in milk was very homogeneous, with an average of 176 records (standard deviation of 16) per day in milk.

Table 1. Relative frequencies (%) of fits among different classes of adjusted r-squared for Z for milk yield and standard deviation of residuals.

	Milk yield				
ADJRSQ[1]	AS	LEG4LP4	CSPL3	CSPL4	CSPL5
<0.20	34	35	29	29	29
0.20-0.40	14	14	9	9	9
0.40-0.60	17	16	11	10	11
0.60-0.80	18	19	15	16	17
>0.80	17	16	36	36	34
RSD	1.51	1.53	1.00	1.00	1.00

The superiority of CSPL, i.e. the ability to fit patterns characterised by marked oscillations, can be clearly observed in Figures 2a and 2b) where examples of individual curves for Z fitted with the Legendre Orthogonal polynomials of fourth order and cubic spline with three knots are reported. In any case, besides the greater flexibility, CSPL shows a marked tendency to produce very large estimates of Z at the extremes of the lactation trajectory.

Figure 2a. Example of an individual curve of Z for milk yield fitted with the Legendre orthogonal polynomials of fourth order.
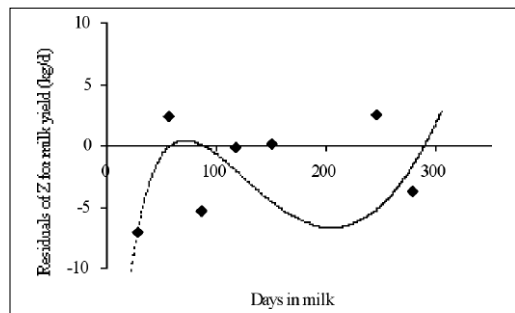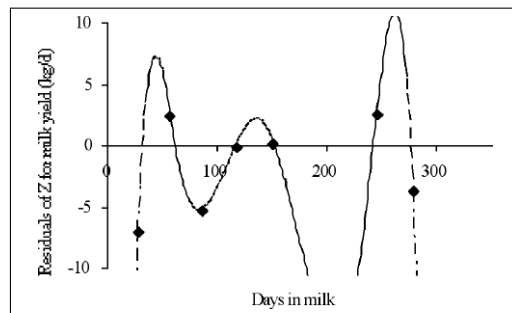
Figure 2b. Example of an individual curve of Z for milk yield fitted with a cubic spline with three knots placed at.

CONCLUSIONS – Results of the present study confirm the difficulty to adequately model individual patterns around the mean curve for milk yield due to great variability of shapes between cows. The AS model and the orthogonal polynomials were able to detect two basic shapes. Although the goodness of fit was in general rather poor for all models considered, a superiority of the cubic splines function to fit Z patterns was highlighted. No differences were found in the model adequacy by varying the number of knots. In any case the superiority of cubic splines over orthogonal polynomials should be tested with genetic models, by considering possible differences on the estimation of breeding values and genetic parameters. Moreover, several studies have already pointed out the tendency of flexible models to yield aberrant estimates at the edges of the space of the independent variable, that has been confirmed also in this work. Therefore, careful attention should be given to random fluctuations that could result in an undesirable mix of genetic and permanent environmental effects with temporary perturbations.

REFERENCES – **Jamrozik, J.,** and Schaeffer L.R.. 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. **Liu, Y. X.**, Zhang, J., L.R. Schaeffer, Yang, R.Q., and Zhang, W. L.. 2006. Optimal random regression models for milk production in dairy cattle. J. Dairy Sci.89:2233-2235. **Lopez-Romero P.,** and Carabano, M.J.. 2003. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein Friesian cattle. Livest. Prod. Sci. 82:81-98. **Meyer, K.** 2005. Random regression analyses using B-splines to model growth of Australian angus cattle. Genet. Sel. Evol. 37:473-500. **Misztal, I.** 2006. Properties of random regression models using linear splines. J. Anim Breed. Genet. 123:74-80. **Ptak, E**., and Schaeffer, L.R.. 1993. Use of test day yields for genetic evaluations of dairy sires and cows. Livest. Prod. Sci. 34:23-34.**Schaeffer, L. R.** 2004. Applications of Random Regression models in animal breeding. Livest. Prod. Sci. 86:35-45.