fabiolacanavesi@anafi.it

Improving test day model genetic evaluation for the Holstein breed in Italy.

Fabiola Canavesi[1], S. Biffani[1], F. Biscarin[1]i, M.arusi[1] and G. Civati[1]
1 – Associazione Nazionale Allevatori Frisona Italiana, Cremona, Italy

Abstract
Genetic evaluation for production traits in the Holstein breed in Italy is based on a Random Regression Test Day Model (RRTDM) since November 2004. More specifically the model is a multiple lactation, multiple traits RRTDM, similar to the model used in Canada for official genetic evaluation. Fixed regression curve effect include time, region, age at calving, parity and season of calving. Last changes in the model included a new definition of the proof scale and of the genetic base. Production and somatic cells data from the February 2005 official evaluation were used to test a different definition of fixed regression curves in the model including year effect, days open effect and a combination of the two. Number of fixed curves increased from 456 to 17024 levels. Simple correlations among proofs ranged from 0.99 to 0.98. Rank correlation among proofs varied from 0.99 to 0.97. The effect of the year of production on the fixed curves showed that there was an increase in production level from 1988 to 1998 and then a slow decrease. The effect of days-open was generally toward the end of lactation. Mean absolute difference between observed and predicted valued was analysed. Research is still ongoing in order to determine which of the four models better predict breeding values.

Introduction
The work towards the development of the genetic evaluation for production traits and somatic cell score using a multiple trait multiple lactation Random Regression Test Day Model (RRTDM) based on Legendre polynomials with the same approach and programs of the Canadian Test Day Model (Jamrozik et al., 1997; Jamrozik et al., 1998; Schaeffer et al., 1999; Kistemaker, 2003) started in 2001. The RRTDM is official in Italy since November 2004.
Among the advantages of a test day model in comparison to a lactation model there are: more efficient use of data as it is collected in the field, a genetic model that more closely defines the true biology of a dairy cow, better estimation of environmental effects especially at the herd level and more accurate cow indexes which can improve the precision of estimated breeding values for bulls (Schaeffer et al., 2000). A random regression test day model allows for estimation of genetic effects throughout lactation and produces persistency of lactation as a by-product, which provides an additional tool to select cows that are easier to manage, have fewer fertility problems and less production stress.
From that day onward a big amount of time has been devoted to meet farmers and industry people to explain the advantages of the new system and of all the additional information that can be used to better select more profitable bulls. Since its first release research has started in order to improve the RRTDM system and to address the many questions that users are raising while getting acquainted with the new system.
One important difference between the old lactation Repeatability Animal Model and the RRTDM is that the latter assumes a more dynamic way of expressing genetic superiority, along the lactation and across lactations, that brings with itself more variation over time in bull proofs. This is not something the users easily get adjusted to and they challenged the system to show its corrected ness.

The aim of this paper is to present some of the work that has been done on the Italian RRTDM in order to verify and to improve:
a) the accuracy of the evaluation;
b) the ability to predict future lactations for all animals in the populations..

Material and methods
Data were test day yields of milk, fat and protein yield and SCS (on log2 scale) from the first three lactations of Italian Holsteins collected from 1988 to 2006 used for the official genetic evaluation of February 2006. Days in milk (DIM) ranged from 5 to 305. Cows were not required to have first lactation data to be included in the evaluation. Nineteen parity-age classes were created; 8 in first lactation, 6 in second lactation and 5 in third lactation. Two seasons of calving were defined: April to September and October to March. Four regions were created by grouping 96 provinces into similar geographical areas. Contemporary groups were defined by herd-test-day-parity.
The model that was used is the one applied for official genetic evaluation and for trait r (milk, fat, protein or SCS) in lactation p (first, second or third) was:

$$y_{ijkprt} = HTDP_{ipr} + \sum_{m=0}^{4} \beta_{kmpr} z_{tm} + \sum_{m=0}^{4} \alpha_{jmpr} z_{tm} + \sum_{m=0}^{4} \rho_{jmpr} z_{tm} + e_{ijkprt}$$

where y_{ijkprt} was the record on trait r of cow j in lactation p on DIM t, within herd-test day effect i and in the subclass k for time-region-season-parity-age of calving; $HTDP_{ipr}$ was fixed herd-test day-parity effect; β_{kmpr} were fixed regression coefficients specific to subclass k of time-region-season-parity-age of calving; α_{jmpr} were random additive genetic coefficients specific to cow j; ρ_{jmpr} were random permanent environmental (PE) coefficients specific for cow j; $e_{ijklprt}$ was the residual effect for each observation, and z_{tm} were covariates. The same function (Legendre polynomial of order 4) was used for all fixed and random regressions.

In matrix notation, the model can be written as

$y = Xb + Za + Wp + e,$

where y is the vector of observations (ordered as traits within cow within lactation), b includes fixed effects, a includes random genetic regression coefficients, p includes random PE regression coefficients, e is the vector of residual effects, and X, Z and W are the incidence matrices. Assume that

$y \mid b,a,p, R \sim N\,(Xb + Za + Wp, R),$

and

$$Var[g'p'e'] = \begin{pmatrix} G \otimes A & 0 & 0 \\ 0 & I \otimes P & 0 \\ 0 & 0 & R \end{pmatrix}$$

where G is the is the 60x60 covariance matrix of the additive genetic regression coefficients, A is the additive genetic covariance matrix among all animals, P is the is the 60x60 covariance matrix of the permanent environmental regression coefficients. is a block diagonal matrix (4x4) of residual covariances between traits with elements that depend on lactation (p) and the interval of days in milk (s). Four intervals of DIM were defined: when DIM are between 5 and 45 d, then s=1; s=2 when DIM are between 46 and 115d; s=3 when DIM are between 116 and 265 d; s=4 when DIM are between 266 and 305 d. Residual covariances between traits on the same test day are allowed to be different from zero and residual covariances were the same within a given interval within parity. Covariances between residuals for records made on different DIM are assumed to be zero in this model (OFF).
Table 1 report 305days genetic and permanent environmental correlations that used for the evaluation.
Three other genetic evaluations were run on the same data changing the structure of the fixed effect and more specifically:

- defining time as year of and not as a group of five years (YEAR), with 17 levels;
- including the effect of days open classes (DO), with 7 classes;
- defining time as year in combination with days open classes (DOY).

Results and discussion
Table 2 reports some statistics related to the data set used for the analysis. The number of classes for fixed effects increased dramatically, especially for the combination of year with days open classes.
Figures 1 to 3 reports solutions for the fixed effects curves of first parity cow of 28month of age, calving in region 1 from April to September from OFF, YEAR and DO evaluation runs, respectively.

Table 2 – General statistics on data used for the analysis

	Total number
Test day records	71,603,358
Cows with records	4,540,499
HTDP levels	5,120,584
Fixed curves (OFF)	456
Fixed curves (YEAR)	2584
Fixed Curves (DO)	3192
Fixed curves (DOY)	18,088

Looking at data from the official evaluation it is possible to observe that t production over time has decreased in the last 5 years compared to years 1993-1997, especially in the first part of the lactation (Figure 1).

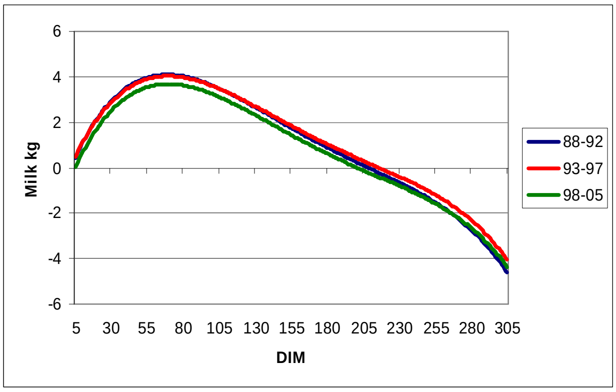

Figure 1 – Fixed regression curves (OFF) for milk yield for the time effect for cows of 28 months of age, calving from April to September, in Region 1.

From the solutions for the fixed effects curves when the year effect (YEAR) is considered in their structure, it is observed that the average production reached a peak around 1998 and then started to decrease perhaps due to change in selection strategies or more probably due to the introduction of milk quota (Figure 2).
Solutions of fixed effect when days open classes are included in the model are in Figure 3 suggesting that their effect is largely on the last part of the lactation as expected.

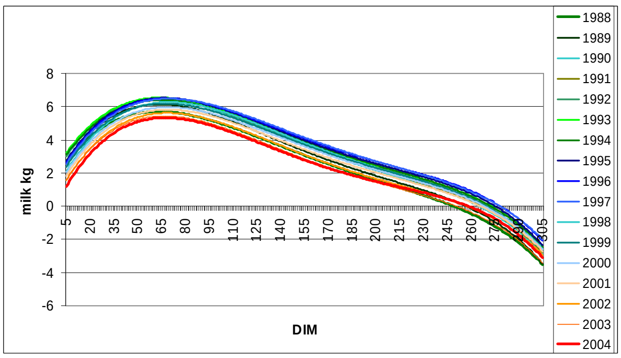

Figure 2 – Fixed regression curves for the year effect (YEAR) for milk yield for cows of 28 months of age, calving from April to September, in Region 1.

fabiolacanavesi@anafi.it

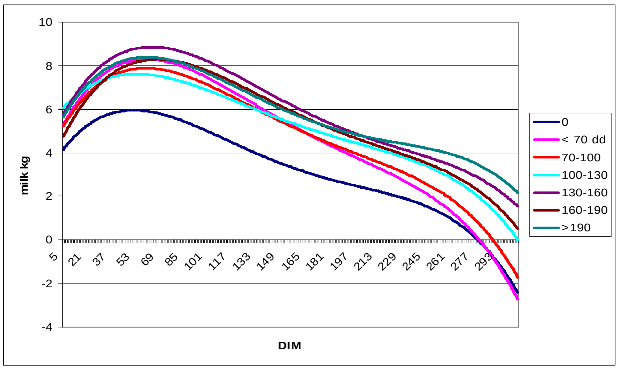

Figure 3 - Fixed regression curves for milk yield in different days open classes (DO) for cows of 28 months of age, calving from April to September, in Region 1

Solutions of fixed effects from the model including year and days open (DOY), not presented, show similar patterns to the ones observed.
Correlations among proofs from the different runs are very high and range from 0.98 to 0.99. Rank correlations for bull proofs ranged from 0.99 to 0.97 and showed the largest variation when both year and days open classes were included in the structure of fixed effects.
Successive analysis were performed on residuals that were normally distributed around zero for all the models. Looking more in details to residual averages according to different fixed effects, no significant differences were observed between OFF, YEAR and DO, but a great reduction in the size of average residual values was observed when year affect and days open were both included in the model.
For simplicity here we report only average residual values for the year effect (Figure 4 and 5) but a similar pattern was observed for other categories that were analysed, like age at calving and province of production.
Average residual values for days open classes clearly show a pattern in the official model which was expected since the official model does not take into account days open effect (Figure 6).

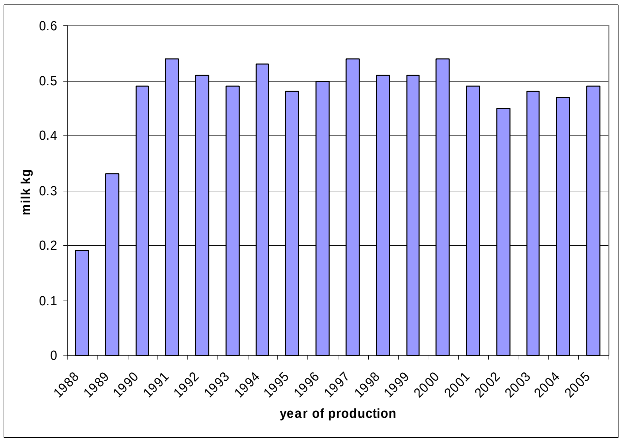

Figure 4 – Average residual level by year effect in the official model (OFF).

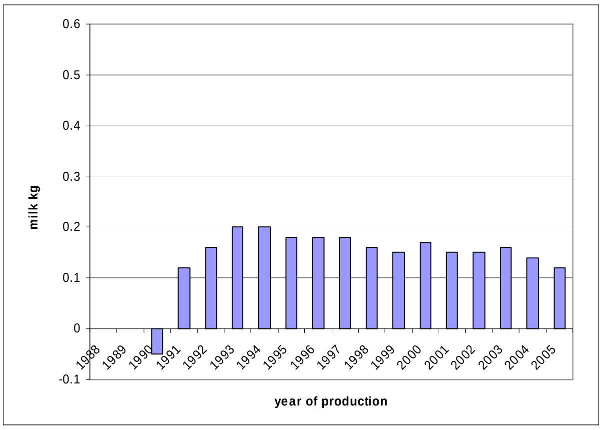

Figure 5 – Average residual level by year effect in the official model (OFF).

The pattern disappears when days open effect are considered in the model, in Figure 7 results from DOY analysis are reported.

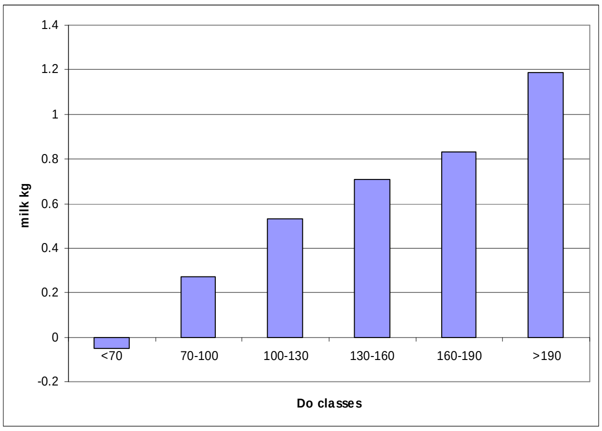

Figure 6 – Average residual level by days open classes in the official model (OFF).

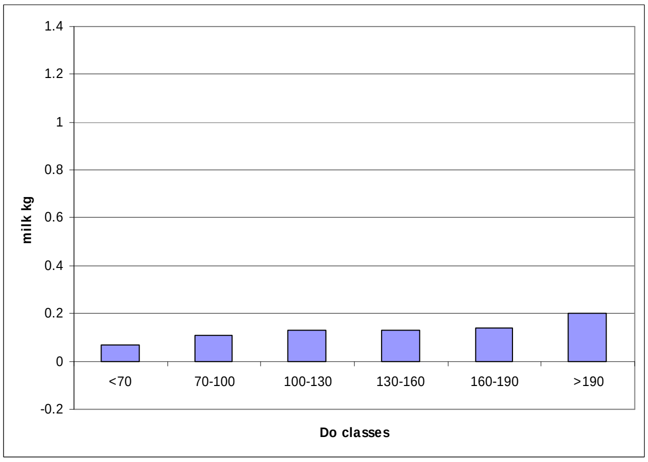

Figure 7 – Average residual level by days open classes in the model including year and days open classes effect(DOY).

fabiolacanavesi@anafi.it

Mean absolute differences between observed and predicted values were the smallest for the model including both year and days open effects.

Conclusions

Based on the results obtained from this study it was decided that in order to improve the accuracy of the model, both year and days open effects should be included in the model. The strategy of including days open as such seems questionable in its approach given the fact that days open information may change from one run of official genetic evaluation to the next. Cows that will change class from run to run will change their estimated breeding value and will affect the proofs of their sire as well thus leading to possible instability of the proofs. Alternative approaches such as accounting for days pregnancy will be investigated in the future.

Acknowledgement

Thank to Prof. Georgios Banos for its useful comments to this work.

References

Jamrozik, J., L.R. Schaeffer and F. Grignola. 1998. Genetic parameter for production traits and somatic cell score of Canadian Holsteins with a multiple trait random regression model. 6WCGALP 23:303-306.

Jamrozik, J., L.R. Schaeffer, Z. Liu and G. Jansen. 1997. Multiple trait random regression test day model for production traits. Interbull Bulletin No. 16:43.

Jamrozik, J., L.R. Schaeffer, and G.B. Jansen. 2000. Approximate accuracies of prediction from random regression models. Livestock Prod. Sci. 66:85-92.

Kistemaker G.J., 2003. The Canadian Test Day Model using Legendre Polynomials. Interbull bulletin 31: 202-204.

Liu Z., F. Reinhardt, and R. Reents. The multiple trait effective2002. 7thWCGALP 32: 553-556.

Liu Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and Calculation of Approximate Reliabilities and Daughter Yield-Deviations of a Random Regression Test-day model for genetic evaluation of Dairy Cattle. J. Dairy Sci. 87: 1896-1907.

Muir B., 2003. Personal communication.

Schaeffer, L.R., J. Jamrozik, G.J. Kistemaker, and B.J. Van Doormal. 2000. Experience with a test day model. J. Dairy Sci. 83:1135-1144

Strandén I., E.A. Mäntysaari. 2004. Comparison of two methods to approximate reliabilities of breeding values from test-day models. 55th EAAP Annual Meeting: 90.

.

Table 1 – 305days parameters for the Italian RRTDM, heritability (average daily) on the diagonal, 305 days genetic correlations below the diagonal, 305days PE correlations above the diagonal (Muir, 2003).

	m1	f1	p1	sc1	m2	f2	p2	sc2	m3	f3	p3	sc3
m1	**.30**	.86	.97	-.18	.48	.37	.47	-.02	.36	.30	.36	.05
f1	.51	**.27**	.88	-.15	.40	.50	.44	-.02	.27	.40	.33	.02
p1	.88	.62	**.28**	-.15	.47	.41	.50	-.01	.35	.33	.39	.06
sc1	.12	-.04	.12	**.17**	-.09	-.08	-.09	.36	-.07	-.05	-.06	.26
m2	.79	.42	.70	.01	**.30**	.88	.97	-.24	.42	.39	.45	.02
f2	.40	.82	.49	-.09	.63	**.29**	.90	-.27	.33	.48	.40	-.01
p2	.67	.54	.79	.03	.90	.73	**.30**	-.22	.43	.44	.49	.03
sc2	.13	.00	.13	.49	-.03	-.09	-.01	**.21**	-.14	-.16	-.15	.44
m3	.70	.35	.63	.05	.86	.51	.78	-.02	**.33**	.88	.97	-.23
f3	.37	.75	.47	-.03	.51	.84	.63	-.06	.66	**.31**	.91	-.25
p3	.57	.45	.69	.07	.74	.60	.85	-.01	.90	.75	**.33**	-.21
sc3	-.01	-.04	-.01	.43	-.17	-.14	-.16	.52	-.21	-.18	-.17	**.25**