Impact of Application of a Random Regression Test Day Model to Production Trait Genetic Evaluations in Dairy Cattle

F. Reinhardt, Z. Liu, A. Bünger, L. Dopp and R. Reents
VIT, Heideweg 1, D-27280 Verden/Aller, Germany

Introduction

A genetic evaluation system using a **R**andom **R**egression **T**est **D**ay **M**odel (RRTDM) was developed for production trait evaluations for German and Austrian Holstein, Red and Jersey breeds (Liu *et al.*, 2001). In parallel to the current official genetic evaluations based on a **F**ixed **R**egression **T**est **D**ay **M**odel (FRTDM, Reents *et al.*, 1998), test runs of the RRTDM system were conducted using data from four consecutive routine evaluations. This paper summarises the changes and improvements introduced to the RRTDM system during the testing phase and compares stability of estimated breeding values (EBV) of both models. In addition, the efficiency of a **M**ultiple **T**rait **E**ffective **D**aughter **C**ontribution (MTEDC, Liu *et al.*, 2001) for approximating reliabilities of EBV is presented here as well.

Material and Methods

Model: A multiple lactation RRTDM was applied to 24-hour test day yields from first three lactations separately for milk, fat, protein yield or somatic cell scores:

$$y_{ijklo} = h_{il} + \sum_{m=1}^{3} \beta_{jlm} f_{jlm} + \sum_{m=1}^{3} b_{klm} a_{klm} + \sum_{m=1}^{3} b_{klm} p_{klm} + e_{ijklo}$$

where y_{ijklo} is 24-hour test day yield, adjusted for heterogeneous herd variance (Reents *et al.*, 1998), of the *o*-th test day of lactation *l* of cow *k*; h_{il} is fixed effects of the *i*-th herd-test-date-parity-milking-frequency (HTD) for lactation *l*; f_{jlm} represents the *m*-th regression coefficient for the *j*-th fixed lactation curve of lactation *l*; β_{jlm} is the *m*-th term of Wilmink function with $\beta_{..1} = 1, \beta_{..2} = d$ and $\beta_{..3} = e^{-0.05d}$ and *d* denoting days in milk (DIM); a_{klm} and p_{klm} are the *m*-th random regression coefficient of lactation *l* of cow *k* for genetic and permanent environmental effects, respectively; b_{klm} is the *m*-th term of the third-order Legendre polynomials with $b_{..1} = 1, b_{..2} = \sqrt{3}z$ and $b_{..3} = \frac{1}{2}\sqrt{5}(3z^2 - 1)$ and $z = (d-5)/150 - 1$; and e_{ijklmo} is error effect. Note that the mathematical function for modelling fixed lactation curves was changed from the third-order Legendre polynomials (Liu et al., 2001) to Wilmink function.

Data: Test day data from four consecutive routine genetic evaluations were analysed using the RRTDM for German and Austrian Holstein, Red and Jersey breeds, in parallel to the official evaluations based on the FRTDM. Table 1 shows the data used for May 2002 test evaluation. Test day records collected between DIM 305 and 330 for production traits were excluded in first three test runs of the RRTDM but included in May 2002 test run (see Table 2). These records represented 4.3% of total test day records.

In the first test run May 2001 it was found that top EBV ranking was dominated by the youngest group of bulls born in 1996, although these bulls had only daughters with short first lactation. A later investigation led to the change of mathematical function for modelling fixed lactation curves from the third-order Legendre polynomials to Wilmink function. Table 2 describes the changes made in the RRTDM genetic evaluation system during the test runs.

Table 1. Size of data set and mixed model equation system for May 2002 test run of production trait genetic evaluation using the RRTDM for German and Austrian Holstein, Red and Jersey breeds.

	Cows with records	Animals in pedigree	Lactations	Herd-test-date-parity-milking-frequency	Test day records	Equations per yield trait
Frequency	9,856,231	14,623,613	19,961,137	13,902,110	163,639,003	234,254,870

Table 2. Changes introduced to the production trait genetic evaluation system using the RRTDM during the testing phase.

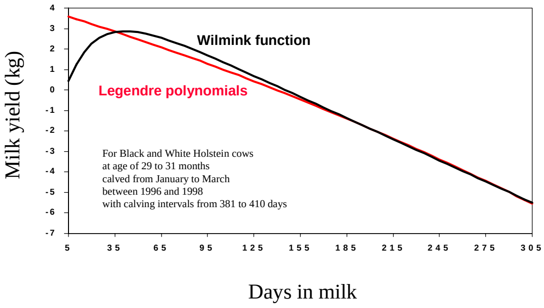
Test run	May 2001	August 2001	February 2002	May 2002
Function for modelling fixed lactation curves	Third-order Legendre polynomials	Third-order Legendre polynomials	Wilmink function	Wilmink function
Use of test day records between DIM 305 and 330	No	No	No	Yes

Results and Discussion

Influence of mathematical functions for fixed lactation curves: An unexpected high proportion of the youngest bulls born in 1996 in top EBV ranking of May 2001 test run led to conducting a residual analysis, besides other investigations, in which the inadequacy of the third-order Legendre polynomials was identified as the cause of the bias. It can be seen in Figure 1 that the third-order Legendre polynomials cannot fit the fast increase phase of lactation curve properly for the example fixed lactation curve, even though the polynomials have a quadratic term. In contrast, Wilmink function results in reasonable fit with a peak of milk yield between DIM 40 and 45. Because the areas beneath both curves are equal, the inappropriate lactation curve function had little effect on cows with complete lactations or bulls with daughters having predominantly complete lactations. However, the inadequacy of the Legendre polynomials clearly caused an unjustified large variation of EBV for the young bulls born in 1996 that had only daughters with incomplete first lactations at the time of genetic evaluation.

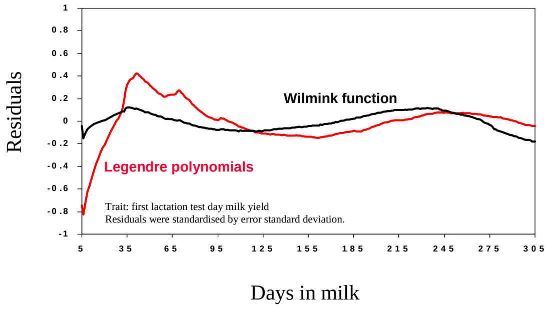
Figure 2 shows that part of lactation curve effects that was not properly accounted for by the third-order Legendre polynomials was left in the residuals of test day records. In contrary to this, average residuals resulting from the use of Wilmink function were close to zero in all lactation stages and varied within 10% error standard deviation with only exception at the very ends of lactation. The two figures demonstrated that the bias caused by the inadequacy of the Legendre polynomials was removed by using Wilmink function for modelling fixed lactation curves.

Figure 1. Estimated fixed lactation curves for first lactation test day milk yield using the third-order Legendre polynomials and Wilmink function



For Black and White Holstein cows
at age of 29 to 31 months
calved from January to March
between 1996 and 1998
with calving intervals from 381 to 410 days

Figure 2. Standardised average residuals using the third-order Legendre polynomials and Wilmink function for modelling fixed lactation curves



Trait: first lactation test day milk yield
Residuals were standardised by error standard deviation.

Stability of EBV of both test day models: An important criterion for assessing the quality of a genetic evaluation model is its predictive ability, particularly in case of missing information. A model with good predictive ability results in high stability of EBV over time, whereas instable EBV may indicate a poor predictive ability of the model. Young dairy bulls in progeny testing programme have almost always daughters with incomplete and/or missing lactations, which forms an ideal sample of data to examine the predicative ability of any model. In an unbiased genetic evaluation system correlation between EBV of the same animal in two evaluations is expected to be $\sqrt{R_e^2 / R_l^2}$ with R_e^2 and R_l^2 being true reliability of early and later evaluation. Approximated reliability values using MTEDC (Liu *et al.*, 2001) were treated as true reliabilities for computing the expected correlations for the RRTDM.

Table 3 shows correlations of EBV of bulls between February and May 2002 test runs using the RRTDM. Because of the difference in mathematical functions for fixed lactation curves, the first two test runs were excluded. For all 9,574 Holstein bulls with at least 35 daughters in at least 10 herds in February 2002 test run, the correlations between both test runs are around 0.995 for both individual lactation and combined lactation EBV of the yield traits. The observed correlations of combined lactation EBV are 0.002 lower than their expected values, which can be explained by the inclusion of test day records beyond DIM 305 in May 2002 test run and by possible difference between the approximated and true reliability values. Since the observed correlations were above 0.99 for bulls born before 1995, only bulls born between 1994 and 1997 are shown here. As daughters' performance data become less and less incomplete, in terms of number of daughters and number of complete lactations and number of test day records, observed correlations between both test runs decrease. Nevertheless, the observed correlations between first EBV in February 2002 and second EBV in May 2002 for the youngest bulls born in 1997 are above 0.96, which indicates fairly high stability of EBV of the RRTDM.

Table 3. Observed and expected correlations of lactation EBV between February and May 2002 test evaluations for Black and White Holstein bulls with at least 35 daughters in at least 10 herds in February 2002 test run using the RRTDM.

Birth	No. of		Observed correlation for lactation			Combined lactation	
year	bulls	Yield traits	1	2	3	Observed	Expected
1994	877	Milk	.998	.995	.993	.996	.999
		Fat	.997	.995	.993	.996	.999
		Protein	.997	.993	.990	.994	.999
1995	986	Milk	.997	.985	.986	.991	.996
		Fat	.996	.985	.985	.990	.996
		Protein	.996	.984	.983	.989	.995
1996	934	Milk	.985	.972	.973	.980	.988
		Fat	.982	.973	.973	.977	.988
		Protein	.979	.968	.969	.974	.986
1997	142	Milk	.963	.962	.964	.963	.977
		Fat	.960	.961	.962	.961	.971
		Protein	.958	.955	.959	.958	.969
All	9,574	Milk	.997	.995	.995	.996	.998
		Fat	.996	.993	.993	.995	.998
		Protein	.997	.994	.994	.995	.997

Similar to Table 3 for the RRTDM, Table 4 gives observed and expected correlations of EBV between February and May 2002 for the FRTDM. Note that test day records beyond DIM 305 were analysed in both test runs of the FRTDM system. Expected correlations were calculated only for protein yield and reliability values were approximated not by the MTEDC rather by an extended single trait reliability method that is less accurate than the MTEDC for test day models. Overall the FRTDM results in high stability of EBV as well, judging from the observed correlations for all bulls. Slightly lower correlations were found for the youngest bulls born in 1997 for the FRTDM than for the RRTDM, although the differences in the correlations are very small. The less accurate reliability method may be responsible for the larger differences between observed and expected correlations than the differences of the RRTDM.

Table 4. Observed and expected correlations of lactation EBV between February and May 2002 evaluations for Black and White Holstein bulls with at least 35 daughters in at least 10 herds in February 2002 run using the FRTDM.

Birth year	No. of bulls	Yield traits	Observed correlation for lactation			Combined lactation	
			1	2	3	Observed	Expected
1994	877	Milk	.999	.997	.994	.997	
		Fat	.998	.997	.996	.997	
		Protein	.998	.997	.990	.996	.999
1995	986	Milk	.998	.979	.984	.989	
		Fat	.996	.983	.983	.986	
		Protein	.997	.983	.976	.987	.995
1996	934	Milk	.986	.949	.964	.971	
		Fat	.984	.965	.966	.974	
		Protein	.982	.962	.960	.970	.988
1997	142	Milk	.967	.951	.955	.959	
		Fat	.958	.950	.951	.953	
		Protein	.961	.952	.951	.955	.986
All	9,574	Milk	.998	.994	.996	.997	
		Fat	.997	.994	.994	.995	
		Protein	.998	.995	.994	.996	.998

Correlations of EBV between both test day models: Table 5 gives correlations of milk yield lactation EBV between the FRTDM and RRTDM for Holstein bulls with at least 35 daughters in at least 10 herds in May 2002 evaluations. It can be seen that bulls' EBV are very highly correlated, specially for bulls born before 1994. The correlations decrease as daughters of bulls have more and more incomplete or missing lactation information. The decreased correlations are caused by the different correlation structures under the two models. Obviously, the two models make different projection of EBV for missing DIM and lactations.

Efficiency of the MTEDC for reliability approximation: After the MTEDC was developed for approximating reliabilities for general multiple trait models (Liu *et al.*, 2001), it has been applied in the test runs of the RRTDM. Both accuracy and efficiency of the MTEDC were studied (Liu *et al.*, 2002), in which one conclusion was made via simulation that the MTEDC does not overestimate reliabilities for young bulls with incomplete daughter performance information but extended single trait reliability methods do. Even without any fine tuning, the reliability computer program using the MTEDC required less RAM, c.a. 90% RAM of the iteration program which is already low in RAM usage due to the lactation based iteration scheme (Liu *et al.*, 2001). Table 6 demonstrates that the MTEDC needs much less CPU than its corresponding iteration program. Based on the RAM and CPU consumption, it is concluded that the MTEDC is feasible for very large populations.

Table 5. Correlations between milk yield lactation EBV of the FRTDM and RRTDM for Black and White Holstein AI bulls with at least 35 daughters in at least 10 herds in May 2002 genetic evaluations.

Birth year	No. of bulls	Lactation 1	Lactation 2	Lactation 3	Combined lactation
1987	400	.99	.97	.97	.98
1988	445	.99	.97	.97	.98
1989	446	.99	.98	.96	.98
1990	626	.99	.96	.95	.97
1991	719	.99	.96	.95	.97
1992	770	.98	.96	.95	.96
1993	831	.98	.96	.95	.97
1994	881	.98	.96	.94	.96
1995	989	.98	.92	.92	.94
1996	1,024	.96	.88	.88	.91
1997	357	.93	.88	.85	.89
All	9,478	.98	.95	.95	.96

Table 6. CPU consumption of the MTEDC method for approximating reliabilities of the RRTDM applied to the data set of May 2002 test run on a HP9000/L2000 unix server.

	Step of computing				
	Data contribution	Progeny contribution	Parental contribution	Reliability of lactation EBV	Total
CPU usage (minutes)	18	121	162	19	320
Equivalent number of rounds of iteration	1.3	8.4	11.3	1.3	22.3

Summary

A production trait genetic evaluation system using a random regression test day model was tested using data from routine evaluations for German and Austrian Holstein, Red and Jersey dairy cattle breeds. The inadequacy of the third-order Legendre polynomials for modelling fixed lactation curve effects was identified via residual analysis and this problem was fixed by replacing it with Wilmink function. Correlations of EBV between two test runs were fairly high for both fixed and random regression test day models. For young bulls having daughters with incomplete or missing lactations the random regression model resulted in slightly more stable EBV over time than the fixed regression model, though the difference in stability is marginal. EBV of both models are very highly correlated, and the correlations decrease as daughters' lactation information becomes less incomplete, indicating different projection of EBV for missing days in milk and lactations by the two models. Approximating reliabilities using the multiple trait effective daughter contribution method required less RAM and much less CPU resources than solving mixed model equations for the RRTDM. It is feasible to apply the multiple trait effective daughter contribution method to very large population for reliability approximation.

Literature Cited

Liu, Z., Reinhardt, F., Bünger, A., Dopp, L. & Reents, R. 2001. Application of a random regression model to genetic evaluations of test day yields and somatic cell scores in dairy cattle. *Interbull Bulletin 27*, 159-166.

Liu, Z., Reinhardt, F. & Reents, R. 2002. The multiple trait effective daughter contribution method applied to approximate reliabilities of estimated breeding values of a random regression test day model for genetic evaluation in dairy cattle. *The 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, France, August 19-23, 2002.

Reents, R., Dopp, L., Schmutz, M. & Reinhardt, F. 1998. Impact of application of a test day model to dairy production traits on genetic evaluations of cows. *Interbull Bulletin 17,* 49-54.