Comparison of different test day models used for genetic evaluation of Polish Black-and-White cattle

E. Ptak and A. Żarnecki

Department of Genetics and Animal Breeding, Cracow Agricultural University
Al. Mickiewicza 24/28, 30-059 Kraków, Poland

ABSTRACT

A single-trait random regression (RR) model was used for a data set of 151,413 test day milk yields of 22,068 cows calving for the first time from 1992 through 1994. The fixed regressions were considered within 32 subclasses of genetic group (defined according to the proportion of HF genes) by age group by season of calving. Two different submodels were used for fixed and random regressions: regression model (RM) and Wilmink model (WM) as well as HYS or HTD fixed effects were incorporated, giving eight different models for comparison.

Differences in estimated breeding values (BVs) calculated on the basis of solutions from each of eight RR models indicated that the change of HYS to HTD effect had no significant influence on BVs. Residual variances were reduced by up to 23% when models with HTD effects were used, showing that they accounted for factors specific to each test day much better than did models with HYS. Correlations between BVs from test day models and the national 305d BVs were 0.71-0.79 for sires and were higher for models with HYS effect. Comparison of the correlations between BVs from different test day models indicated that the choice of covariates for random regressions was important. The use of the Wilmink function for those regressions is justified because it is accurate enought and computationally more efficient.

KEY WORDS: breeding value, test day milk yield, lactation curve, random regression model

INTRODUCTION

Accurate estimation of breeding values is of great importance in cattle breeding programmes. For many years genetic evaluation of dairy sires and cows has been based on analysis of 305d lactation milk production. Since 1991 there has

been growing interest in using test day (TD) yields instead of the traditional 305d lactation yields for that purpose. In the first TD model, fixed regressions within age and season of calving were included to account for the lactation curve of groups of cows (Ptak and Schaeffer, 1993). In the next TD model, random regressions were incorporated to model a separate lactation curve for every cow (Schaeffer and Dekkers, 1994). In recent years much work has been done to improve models for analysis of TD data (Jamrozik et al., 1997a,b, 1998; Van der Werf et al., 1998; Rekaya et al., 1999).

TD models account for genetic and environmental effects directly influencing milk yield at the day of recording. They use every piece of information available and there is no need for extension factors. No assumptions about the lactation course have to be made. Models for TD yields allow more accurate modeling of single yields and reduce the residual variance (Ptak and Schaeffer, 1993; Reents et al., 1995; Swalve, 1995; Ptak and Żarnecki, 1998).

In TD models the environmental effects are often represented by herd-test day (HTD) effects rather than herd-year-season (HYS) effects, resulting in remarkable reduction of residual variance (Ptak and Schaeffer, 1993; Ptak and Żarnecki, 1998). Ptak and Schaeffer (1993) compared TD models with HTD effects and with HYS effects finding the former more suitable than the latter for the Canadian dairy cattle population. In countries with small herd sizes, like Poland or Germany, the number of HTD subclasses is usually many times greater than that of HYS, adversely affecting the degree of connectedness and causing problems with estimability. Grünhagen and Swalve (1997) concluded that there was only a slight impact of the small size of HTD subclasses on the degree of connectedness and estimated breeding values. In order to increase the accuracy of genetic evaluation of dairy cattle, different submodels (as functions of covariates for fixed and random regressions) in TD models were tested (Jamrozik et al., 1997b; Guo and Swalve, 1997). Fewer covariates and simpler combinations of covariates are better for computations.

The objective of this study was to compare random regression (RR) models for milk yield with either HTD or HYS effects in a population with small herd sizes. The comparison also included the use of different submodels for fixed and random regressions.

MATERIAL AND METHODS

The data file consisted of 151,413 TD milk yields of 22,068 first lactation cows with an average of 7 tests per cow. Cows were calved from 1992 through 1994. There were 1,339 sires and 20,039 dams. The pedigree file contained 45,775 animals including seven phantom parent groups (three created for bulls and four

for dams). The data were grouped in 4,661 HYS subclasses or 27,903 HTD subclasses, almost 27% of them with single records. Only data from cows that calved between 18 and 48 months of age and TD yields recorded between 5 and 305d in milk were used.

The following single-trait random regression model (Jamrozik et al., 1997a) was used:

$$y_{ijkl} = H_i + \sum_{m=0}^{s} b_{km} x_{jlm} + \sum_{m=0}^{s} a_{jm} x_{jlm} + p_j + e_{ijkl}$$

where y_{ijkl} is the milk yield of a cow at t_{ji} day in milk (DIM), H_i is HYS or HTD effect, b_{km} are fixed regression coefficients within the k-th genetic group by age at calving by season of calving subclass, a_{jm} are random regression coefficients for the j-th animal with var(**p**)= **G**⊗**A**, p_j is the random permanent environmental effect with var (**p**) = $\mathbf{I}\sigma_p^2$, and e_{ijkl} is the random residual effect with var(**e**) = **R**. The vector of covariates, $\mathbf{x}_{ji}$, is defined as WM-Wilmink submodel $(x_{j10}, x_{j11}, x_{j12})'$ = $(1, t_{ji}, \exp(-0.05t_{ji}))'$ or RM-regression submodel $(x_{j10}, x_{j11}, x_{j12}, x_{j13}, x_{j14})'$ = $(1, t_{ji}/305,(t_{ji}/305)^2, \ln(305/t_{ji}), \ln^2(305/t_{ji}))'$. **G** and **R** denote the (co)variance matrices for additive genetic and residual effects, **A** is the numerator relationship matrix between animals, and **I** is the identity matrix. **R** is the diagonal matrix, with elements equal to estimates of residual variances i.e.**R** = diag (σ_e^2). These values (σ_e^2)were calculated for each TD record using regression equation with linear and quadratic terms. The regression equation was fitted to variances of errors $(\hat{e}_{ijkl})$ in each day in milk through lactation to smooth these values. The residuals were evaluated as:

$$\hat{e}_{ijkl} = y_{ijkl} - \hat{H}_i + \sum_{m=0}^{s} \hat{b}_{km} x_{jlm} + \sum_{m=0}^{s} \hat{a}_{jm} x_{jlm} - \hat{p}_j$$

The symbol ⊗ denotes the Kronecker product function (Searle, 1982).

The data were analyzed using different combinations of fixed effect H (HYS or HTD) and different combinations of covariates for fixed and random regressions (RM-regression submodel (Ali and Schaeffer, 1987) or WM-Wilmink submodel (Wilmink, 1987)). As a result, eight models were applied to the TD data:

(1)   HYS effect and RM for fixed regressions and RM for random regressions,
(2)   HYS effect and WM for fixed regressions and RM for random regressions,
(3)   HYS effect and RM for fixed regressions and WM for random regressions,
(4)   HYS effect and WM for fixed regressions and WM for random regressions,
(5)   HTD effect and RM for fixed regressions and RM for random regressions,
(6)   HTD effect and WM for fixed regressions and RM for random regressions,

(7)   HTD effect and RM for fixed regressions and WM for random regressions,
(8)   HTD effect and WM for fixed regressions and WM for random regressions.

The fixed regressions were considered within 32 subclasses of genetic group by age group by season of calving, with numbers of TD yields per subclass in Table 1. The four genetic groups were pure Polish Black-and-White (PBW), ith <50% HF genes, with <75% HF genes, and with up to 100% HF genes. Four age groups were also created: 18 to 23 mo, 24 to 29 mo, 30 to 35 mo, and 36 to 48 mo. Cows calved in two seasons: summer (April-September) and winter (October-March).

Number of test day records by age, genetic group and season of calving

Season	Genetic group	Age at calving; months				
		18-23	24-29	30-35	36-48	Total
April-	PBW*	672	4,078	1,730	961	7,441
September	<50% HF	1,109	11,587	4,787	1,226	18,709
	<75% HF	1,178	11,816	3,259	452	16,705
	up to 100% HF	366	4,947	1,537	111	6,961
October-	PBW*	1,169	10,192	5,321	1,961	18,643
March	<50% HF	2,548	19,177	11,516	3,116	36,357
	<75% HF	2,006	19,116	8,509	1,455	31,086
	up to 100% HF	871	9,318	4,776	546	15,511
Total		9,919	90,231	41,435	9,828	151,413

* PBW – Polish Black-and-Whites

The BV based on TD yields for the j-th animal on each DIM were calculated as $DBV_j(DIM) = a_{j0} + a_{j1}(DIM/305) + a_{j2}(DIM/305)^2 + a_{j3}\ln(305/DIM) + a_{j4}\ln^2(305/DIM)$ for models (1), (2), (5) and (6), or as $DBV_j(DIM) = a_{j0} + a_{j1}DIM + a_{j2}\exp(-0.05DIM)$ for models (3), (4), (7) and (8), where **a** is a vector of solutions for random regression coefficients. The BVs for a complete 305d lactation were obtained by summing $DBV_j(DIM)$ for DIM=1,305. These values were correlated with the national evaluations (BV(R)) based on first 305d lactation yields (Żarnecki et al., 1999).

Comparisons of models for daily milk yields were based upon the differences in estimated breeding values (BV), residual variances, and correlations between actual and predicted TD yields. Correlations of BVs between pairs of models were also considered.

RESULTS AND DISCUSSION

   Mean differences in BV for 305d milk yield between all pairs of models with HTD effect are plotted in Figure 1 for sires with more than 50 daughters. The mean differences between pairs of models with HYS effect were very similar to those shown in Figure 1, so for clarity they were not included in the plot, but they indicated that the change of HYS to HTD had not significant effect on BVs. The differences were smallest when the same submodel was used for both fixed and random regressions (model (5) and (8)) or when WM was used for fixed and RM for random regressions (model (6)). When RM was applied for fixed and WM for random regressions (model (7)), the differences from other submodels were very large, suggesting that this combination of functions used for regressions gave differently shaped lactation curves than did models (5), (6) and (8). In these cases the largest differences were observed. The standard deviations of the differences in BVs between models made up from 20% up to almost 850% of the mean. The smaller were mean of differences, the higher were values of the standard deviations of these differences. Jamrozik et al. (1997b) obtained similar results and concluded that changing the function used for the random regressions had a greater effect on the standard deviations of differences in BVs than did changing the function for fixed regressions.
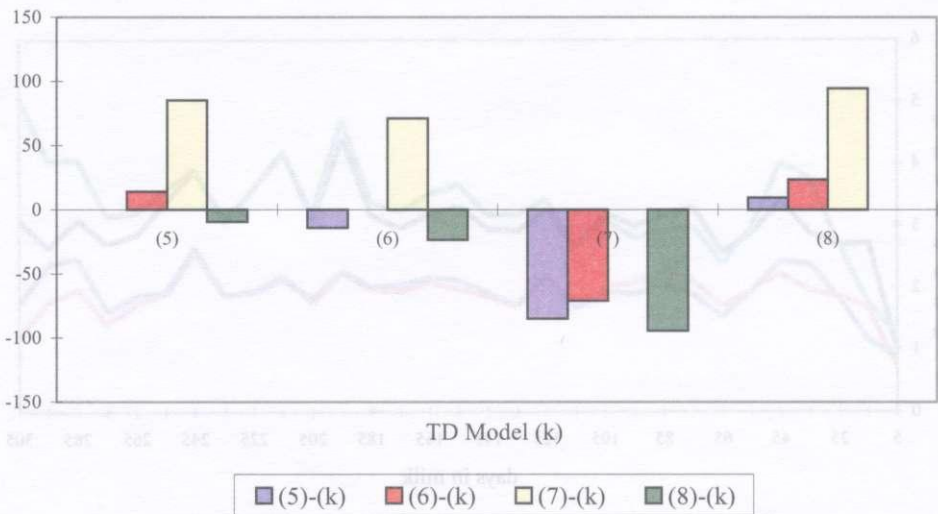

Figure 1. Mean differences in BV between pairs of TD models with fixed effect of HTD for sires with more than 50 daughters

Residual variance estimates calculated for each model are given in Table 2. Models with HTD effects accounted for factors specific to each test day much better than did those with HYS. Residual variances were reduced by 19 to 23% when models with HTD were used. Similar results were reported by Ptak and Schaeffer (1993). The use of regression (RM) or Wilmink (WM) submodels had no impact on the goodness of fit of models with either HYS or HTD effects.

Residual variance estimates for milk yields at every DIM through lactation for different models were calculated. For simplicity, only a few models were chosen to present the differences in residual variance estimates (Figure 2). Models with the same function for random regressions and different functions for fixed regressions followed almost the same pattern, indicating that the precision of the model is determined by the submodel chosen for random regressions (Jamrozik et al., 1997b; Ptak and Żarnecki, 1998). The differences in residual variance estimates

TABLE 2

Residual variance estimates and correlations between actual and predicted test day yields for TD models

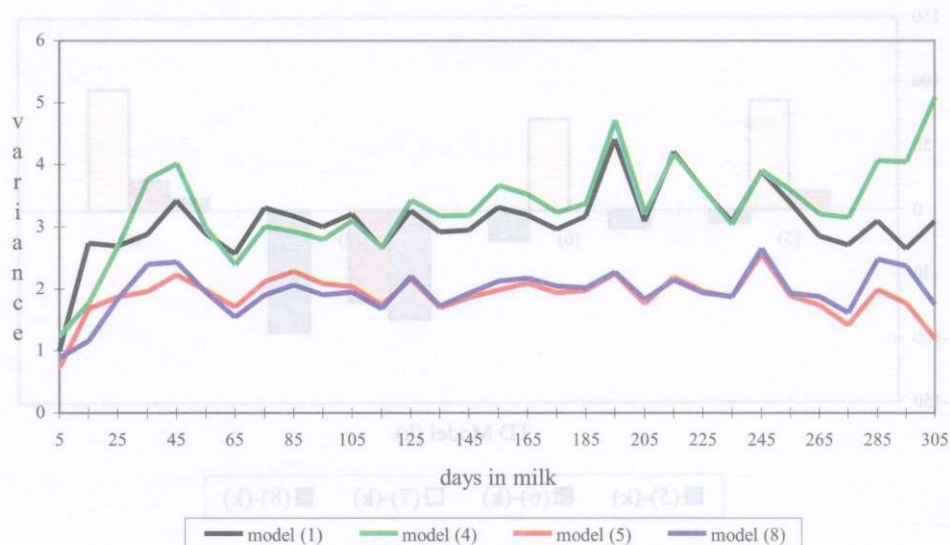| Criterion of comparison	Model							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Residual variance	4.273	4.292	4.064	4.244	3.448	3.452	3.186	3.280
Correlation estimates	0.937	0.936	0.935	0.934	0.960	0.959	0.960	0.960

Figure 2. Residual variance estimates of daily milk yields from some TD models

among models were negligible in 4, 7 and 9 months of lactation, whereas at the end of lactation they were large, especially for models with HYS. These variances ranged from 0.72 to 2.65 for models with HTD and from 0.96 to 5.08 for models with HYS (Figure 2).

The correlations between actual and predicted TD yields are presented in Table 2. The higher values, 0.96, were observed in models with HTD effects (models (5)-(8)). The correlations between actual and predicted TD yields within successive months of lactation were also calculated, and they ranged between 0.88 and 0.97. The fit of each model was better at the beginning of lactation (0.94-0.97); then it decreased, with the lowest values of correlations after 240 days in milk (0.88-0.95).

The correlations between BVs from TD models and the national 305d BVs were calculated for 79 sires with more than 50 daughters, and for 19,996 recorded cows (Table 3). In all cases, the correlations between BVs for bulls were higher than those between BVs for cows. Higher correlations between national BVs and TD BVs occured in models with HYS effects because these models were more similar to the model for national evaluation. In both types of models the HYS effect was defined as a common effect for all test day yields of all cows that calved in the same herd-year-season.

The correlations between BVs from TD models (1) and (2) as well as from models (3) and (4) was 0.99, indicating that the covariates for fixed regressions (RM or WM) gave the same solutions, whereas the correlations between BVs from other TD models were lower, and the choice of covariates for random regressions was important. This result agrees with those obtained by Jamrozik et al. (1997b). From a computational point of view it is better to choose the Wilmink submodel

TABLE 3

Correlations between breeding values (BV(1 to 8)) and 305d national breeding values (BV(R)) for bulls (above diagonal) and for cows with records (below diagonal)

	BV(1)	BV(2)	BV(3)	BV(4)	BV(R)
BV(1)		0.99	0.96	0.97	0.79
BV(2)	0.99		0.95	0.96	0.77
BV(3)	0.88	0.88		0.99	0.77
BV(4)	0.88	0.88	0.89		0.79
BV(R)	0.66	0.65	0.64	0.63	

	BV(5)	BV(6)	BV(7)	BV(8)	BV(R)
BV(5)		0.99	0.98	0.97	0.73
BV(6)	0.99		0.98	0.98	0.73
BV(7)	0.92	0.92		0.99	0.71
BV(8)	0.92	0.92	0.99		0.71
BV(R)	0.59	0.59	0.56	0.55	

(WM) because it has only three parameters to estimate. The use of simpler functions as submodels in TD models has already been proposed by other authors (Guo and Swalve; 1997, Jamrozik et al., 1997b).

Comparison of all eight TD models indicates that models with HTD effects are better, and that the use of the Wilmink submodel for random regressions is justified because it is as accurate as the regression submodel but computationally more efficient.

REFERENCES

Ali T.E., Schaeffer L.R., 1987. Accounting for covariances among test day milk yields in dairy cows. Can. J. Anim. Sci. 67, 637-644

Grünhagen E., Swalve H.H., 1997. Effect of cell size in test day models. Interbull Meeting, Vienna (Austria), Bull. 16, 66-70

Guo Z., Swalve H.H., 1997. Comparison of different lactation curve sub-models in test day models. Interbull Meeting, Vienna (Austria), Bull. 16, 75-79

Jamrozik J., Schaeffer L.R., Dekkers J.C.M., 1997a. Genetic evaluation of dairy cattle using test day yields and random regression model. J. Dairy Sci. 80, 1217-1226

Jamrozik J., Kistemaker G.J., Dekkers J.C.M., Schaeffer L.R., 1997b. Comparison of possible covariates for use in a random regression model for analyses of test day yields. J. Dairy Sci. 80, 2550-2556

Jamrozik J., Schaeffer L.R., Grignola F., 1998. Genetic parameters for production traits and somatic cell score of Canadian Holsteins with multiple trait random regression model. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale (Australia), 23, 303-306

Ptak E., Schaeffer L.R., 1993. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34, 23-34

Ptak E., Żarnecki A., 1998. Genetic evaluation of Polish Black and White cattle using different test day models. Abstracts of the 49th Meeting of EAAP, Warsaw (Poland), 4, 44

Reents R., Dekkers J.C.M., Schaeffer L.R., 1995. Genetic evaluation for somatic cell score with a test day model for multiple lactations. J. Dairy Sci. 78, 2858-2870

Rekaya R., Carabano M.J., Toro M.A., 1999. Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle. Livest. Prod. Sci. 57, 203-217

Schaeffer L.R., Dekkers J.C.M., 1994. Random regressions in animal models for test day production in dairy cattle. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Guelph (Canada), 18, 443-446

Searle S.R., 1982. Matrix Algebra Useful for Statistics. John Wiley and Sons, Inc., New York, NY

Swalve H.H., 1995. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. J. Dairy Sci. 78, 929-938

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

Wilmink J.B.M., 1987. Comparison of different methods of predicting 305d milk yield using means calculated from within herd lactation curves. Livest. Prod. Sci. 17, 1-17

Żarnecki A., Jagusiak W., Trela J., 1999. Sire evaluation for dairy production traits. Research Institute of Animal Production, Kraków (Poland), Vol. 19, ISSN 0867-1249

STRESZCZENIE

Porównanie różnych modeli oceny wartości hodowlanej polskiego bydła czarno-białego na podstawie próbnych udojów

Do obliczeń posłużyły dane z 151 413 próbnych udojów 22 068 krów-pierwiastek, cielących się w latach 1992-1994. Wartości hodowlane oszacowano przy pomocy jednocechowego modelu zwierzęcia z losowymi regresjami (RR model). W modelu tym uwzględniono regresje na liczbę dni doju (efekt stały) w obrębie 32 klas, które utworzono wg przynależności krowy do grupy genetycznej (w zależności od % udziału genów HF), grupy wiekowej oraz sezonu ocielenia. Regresje traktowane jako efekt losowy i efekt stały wyrażono przy pomocy dwóch modeli: tzw. regresyjnego (RM) i Wilminka (WM). W modelu liniowym uwzględniono stały efekt stada-roku-sezonu ocielenia (HYS) lub stały efekt stada-dnia próbnego udoju (HTD), uzyskując osiem różnych RR modeli do porównania.

Porównując różnice między wartościami hodowlanymi oszacowanymi na podstawie ośmiu RR modeli nie stwierdzono, aby zamiana efektu HYS na HTD wpłynęła istotnie na wielkość oszacowanych wartości hodowlanych. W modelach ze stałym efektem HTD wariancje resztowe natomiast zmniejszyły się nawet o 23%. Świadczy to o dokładniejszym modelowaniu próbnych udojów przez modele z efektem HTD, ponieważ uwzględnia się w ten sposób czynniki wpływające danego dnia, w danym stadzie, na wydajność mleczną krowy.

Współczynniki korelacji liniowej między wartościami hodowlanymi oszacowanymi przy pomocy modeli dla próbnych udojów i pochodzącymi z oficjalnej oceny opartej na 305-dniowych wydajnościach wynosiły 0,71-0,79 dla ojców i były większe, gdy zastosowano RR modele z efektem HYS. Porównanie wielkości współczynników korelacji liniowej między wartościami hodowlanymi obliczonymi na podstawie różnych modeli RR wskazuje, że wybór funkcji dla losowych regresji (RM lub WM) ma istotne znaczenie, przy czym zalecane jest użycie funkcji Wilminka (WM), gdyż oba modele dają prawie takie same wyniki oceny wartości hodowlanej, ale prostszy, z obliczeniowego punktu widzenia, jest model WM.