

## BLENDING OF TEST DAY AND LACTATION RECORDS USING A MULTITRAIT RANDOM REGRESSION MODEL

T.M. Villumsen<sup>1</sup>, P. Madsen<sup>1</sup>, J. Jensen<sup>1</sup> and J.H. Jakobsen<sup>2</sup>

<sup>1</sup>Danish Institute of Agricultural Sciences. Department of Animal Breeding and Genetics, P.O. BOX 50, DK-8830 Tjele, Denmark

<sup>2</sup>Danish Agricultural Advisory Centre, Udkaersvej 15, Skejby, DK-8200 Aarhus N, Denmark

### INTRODUCTION

Genetic evaluation for yield traits in dairy cattle is in Denmark currently based on an Animal Model (AM) for 305-d lactation records. Implementation of a random regression (RR) test day (TD) model is in preparation. The existing lactation model uses lactation records from 1984 and onwards; TD records are only available from 1990 and onwards. In order to utilize data from the same time span in a TD model as in the present lactation model, a blending procedure including both TD and lactation records is needed. Furthermore, a planned joint genetic evaluation of dairy cattle in the Nordic countries must be based on a model that is able to handle both lactation and TD records. This is because Finland and soon Denmark will use a TD model, while Sweden and Norway will continue using a lactation model for genetic evaluation. The main objective for this study was to develop a model using both TD records and lactation records. To test the approach, first lactation protein yields for Holsteins in a region of Denmark were used. A multitrait model for TD records and 305-d records was developed. Different scenarios illustrating the situations where blending will be needed were investigated and correlations between predicted breeding values (BV) in different scenarios were computed.

### MATERIAL AND METHODS

**Model.** A blending model was developed. It is a multitrait model where test day records are modeled by a random regression AM and the 305-day lactation records are modeled as a traditional AM. In matrix notation the bivariate blending model can be written as:

$$\begin{bmatrix} y_{td} \\ y_l \end{bmatrix} = \begin{bmatrix} X_{td}b_{td} \\ X_l b_l \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^n Z_{tdi} a_{tdi} \\ Z_l a_l \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^m W_{tdi} pe_{tdi} \\ W_l pe_l \end{bmatrix} + \begin{bmatrix} e_{td} \\ e_l \end{bmatrix},$$

where subscript td and l refers to test day and lactation records. The  $y$ 's are vectors of observations,  $b$ 's,  $a$ 's,  $pe$ 's, and  $e$ 's are vectors of fixed, random additive genetic, random permanent environmental, and random residual effects.  $X$ 's,  $Z$ 's, and  $W$ 's are the corresponding design matrices, and  $n$  and  $m$  are the order of fit of the additive genetic and permanent environmental (PE) covariance functions (CF) respectively, used for the TD model.

Variances are:

$$\text{var} \begin{bmatrix} a_{td} \\ a_l \end{bmatrix} = G_0 \otimes A, \quad \text{var} \begin{bmatrix} pe_{td} \\ pe_l \end{bmatrix} = P_0 \otimes I_{NA}, \quad \text{var} \begin{bmatrix} e_{td} \\ e_l \end{bmatrix} = R_0 \otimes I_N,$$

where  $\mathbf{A}$  is the relationship matrix,  $\mathbf{G}_0$  is the additive genetic (co)-variance matrix of dimension  $(n+1) \times (n+1)$ ,  $\mathbf{I}_{NA}$  is an identity matrix of dimension equal to number of animal with records,  $\mathbf{P}_0$  is the permanent environmental (co)-variance matrix of dimension  $(m+1) \times (m+1)$ ,  $\mathbf{I}_N$  is an identity matrix of size equal to the number of observations,  $\mathbf{R}_0$  is a  $2 \times 2$  residual (co)-variance matrix, the residual covariance between TD-records and 305-d records is assumed to be zero due to computational constraints.

A 305-d lactation record is a function of TD records. Consequently, the (co)-variance components for 305-d yield can be derived from the CF for TD records, which leads to non positive definite (co)-variance matrices. Positive definiteness can be obtained by bending (Hayes & Hill, 1981).

**Data.** To estimate (co)-variance components and to test the bivariate blending model, data for first lactation protein yield for 42,978 Danish Holsteins calving from January 1990 to December 2000, were extracted from the Danish Central Cattle Database. Cows were 19-40 month old at calving. Data included 379,671 TD records measured between 4 days in milk (DIM) to 365 DIM and 35,759 records for accumulated 305-d yield. The cow's average number of TD records was 8.8. All lactation records were based on minimum 280 DIM (i.e. short lactations were excluded). Two pedigree files were build, one for use in AM and one for use in sire models (SM). The pedigree files were traced on sire-dam relationship, the number of animals in the files was 117,657 animals for use in AM, and 9,442 animals for use in SM.

**Estimation of CF.** The two-step procedure of Kirkpatrick *et al.* (1990) was used. Step one involves estimation of (co)-variance matrices in a multitrait model, step two involves fitting of CF's. In step one, five "traits" were defined based on stage of lactation. The following periods were used: DIM 5-20, DIM 31-60, DIM 121-150, DIM 211-240, and DIM 301-330. Numbers of observations for the five traits were 20,965, 37,593, 34,685, 30,961, and 17,255. Variance components were estimated in bivariate AM. Covariances were estimated in SM and transformed to AM parameters based on SM correlations and AM variances. This is to account for the downwards bias in SM estimates, which is assumed to have an equal effect on variances and covariances. All analyses included period 2 (peak yield), in order to account for selection bias. Fixed class effects in the analyses were age at calving and herd, both nested within 5-year periods. Fixed regressions were breed proportion (Holstein), heterozygosity, and stage of gestation for the two last periods. Random effects were herd $\times$ test day, animal/sire, and residual. The estimates were obtained by the AI-REML algorithm (Jensen *et al.*, 1996) using the DMU-package (Madsen & Jensen, 2000).

The complete (co)-variance matrices were obtained by the iterative summing procedure of Mäntysaari (1999) on estimated submatrices. In step two, CF's were estimated using least squares techniques. DIM was defined within the interval 1 to 365 DIM and standardized from -1 to 1. A fifth order normalized Legendre polynomial was fitted to the genetic matrix. The residual (co)-variance matrix was separated in a PE covariance matrix and a homogeneous measurement error using least squares. A fourth order normalized Legendre polynomial was fitted to PE. The CF's were then expanded with the 305-day trait, defined as the sum of 305 test days, and bent to obtain positive definite CF's.

**Scenarios.** To test the accuracy of the multitrait approach, prediction of BV's were performed in five scenarios to simulate situations where blending will be necessary. Correlations between breeding values predicted in the scenarios (SC) below would give a direct measure of degree of association. Tested scenarios for prediction of BV's were:

- SC1 Prediction of BV based on TD records only
- SC2 Prediction of BV based on lactation records only
- SC3 Prediction of BV based on all TD records and 305-d lactation records
- SC4 Prediction of BV based on TD records for cows from herds with even herd numbers and 305-d records for cows from herds with uneven herd numbers (joint breeding evaluation between countries using either test day or lactation models)
- SC5 Prediction of BV based on lactation records for cows with first calving up to 1993, cows with first calving after 1993 had both lactation and TD records (illustrating a situation where old TD records are missing)

For prediction of BV's the following effects were included to describe TD protein yield. Fixed class effects were herd×test day, DIM, herd within period (5-year), calving year-month, calving age (months) within period. Fixed regressions were, breed proportion (Holstein), and heterozygosity. Random effects were five RR coefficients describing the additive genetic effect, four RR coefficients describing the PE effect, and a measurement error. The model for lactation yield included the fixed class effects management group, calving year-month within period, and calving age (months) within period. Fixed regressions were current calving interval, breed proportion (Holstein), and heterozygosity. Random effects were additive genetic effect, PE effect, and a measurement error. SC3, SC4, and SC5 gave two predictions of 305-d BV's, one based TD solutions and one based on the solutions for the 305-d trait.

## RESULTS AND DISCUSSION

Correlations between predicted 305-d BV's for cows in the five scenarios were estimated within year of birth. The results are presented in Table 1, where data source specifies whether the 305-d BV's are based on solutions to the TD model or solutions to the lactation model. The numbers of TD and lactation records for each scenario are given in parentheses.

The correlation between SC1 and SC2 was 0.798, which was the lowest found for the five scenarios. This was expected since SC1 only includes TD records and SC2 only includes lactation records. This indicates that changing the evaluation model does lead to considerable changes in BV's. Correlations between SC1 and SC3 where all information was utilized indicate that the inclusion of 305-d lactation records in the model leads to almost identical BV's –both when BV's were based on 305-d solutions and calculated from TD solutions. The correlations between SC1 and SC4 or SC5 were all close to 0.95, for both lactation and TD solutions indicating that replacement of some TD records with lactation records would not lead to drastic changes in BV's. All correlations between SC2 and other scenarios were in the range 0.80-0.85, showing that inclusion of TD records in the blending model does lead to large changes in BV's, which was expected due to the different approaches for the lactation and TD model. For SC3 to SC5, it was seen that within scenario, the correlation between BV's based

on TD and lactation solutions, respectively, were close to 1. The high within-scenario correlations causes that the between scenario correlations were very similar, independent of the BV solution source. The BV correlations between SC3 and SC4 or SC5 respectively, were about 0.95. Correlations for BV's based on TD solutions between SC3 and SC4 or SC5 were very similar to correlations between SC1 and SC4 or SC5, indicating that the inclusion of lactation records in SC3 had a minimal influence on the BV's.

**Table 1. Correlations within year for predicted 305-d BV's. The correlations are based on predicted breeding values for cows**

Scenario	Data source (no of records)	1		2		3		4		5		
		TD	Lact	TD	Lact	TD	Lact	TD	Lact	TD	Lact	
		<b>(379.671)</b>										
2	Lact (35.759)	0.798										
3	TD (379.671)	0.995	0.807									
	Lact (35.759)	0.986	0.819	0.996								
4	TD (248.786)	0.952	0.839	0.951	0.947							
	Lact (12.189)	0.949	0.844	0.949	0.946	0.995						
5	TD (231.163)	0.956	0.842	0.961	0.961	0.938	0.936					
	Lact (35.759)	0.948	0.854	0.958	0.965	0.934	0.934	0.996				

## CONCLUSION

The approach to combine TD records and lactation records in a multitrait random regression model worked well. The high correlation between predicted BV's in SC1 and SC3 indicated that the test day model as expected could explain almost all of the genetic variation in lactation records. Correlations between SC2 and the other scenarios were in the range 0.80-0.85. The lower correlations between the blending models to a model based solely on lactation records were expected due to the different evaluation models. The high correlations between SC1 and SC4 or SC5, where about 1/3 of all TD records were omitted, indicated that a genetic evaluation based on solutions from the bivariate blending model yielded accurate results. Therefore, genetic evaluations based on all available Danish TD and lactation records, or a joint breeding evaluation between two countries using the blending approach is expected to optimally use all available information.

## REFERENCES

- Hayes, J.F., and Hill, W.G., (1981) *Biometrics* **37** : 483-493.  
 Jensen, J., Mäntysaari, E.A., Madsen, P., and Thompson, R., (1996-97) *J. Ind. Soc. Ag. Stat.* **49** : 215-236.  
 Kirkpatrick, M., Lofsvold, D., and Bulmer, M., (1990) *Genetics* **124** : 979-993.  
 Madsen, P., and Jensen, J., (2000) *A User's Guide to DMU version 6*, release 4, DIAS.  
 Mäntysaari, E.A., (1999) *50<sup>th</sup> Ann. Meeting of the EAAP* **22** : 1-4.