Measurement Error Variance of Test-Day Observations from Automatic Milking Systems

Pitkänen, T.[1], Mäntysaari, E. A.[1], Nielsen, U. S.[2], Aamand, G. P[3]., Madsen[4], P. and Lidauer, M. H.[1]

[1] *Genetic Research, MTT Agrifood Research Finland, Jokioinen*
[2] *The Knowledge Centre for Agriculture, Cattle, Denmark*
[3] *Nordic Cattle Genetic Evaluation, Denmark*
[4] *Centre for Quantitative Genetics and Genomics, Aarhus University, Denmark*

Abstract

Automated milking systems (AMS) are becoming more popular in dairy farms. In this paper we present an approach for estimation of residual error covariance matrices for AMS and conventional milking system (CMS) observations. The variances for other random effects are kept as defined in the evaluation model. AMS residual variances were found to be 16 to 37 percent smaller for milk and protein yield and 42 to 47 percent larger for fat yield compared to CMS.

Key words: Residual variance, genetic evaluation model, Nordic Holstein

Introduction

Automated milking systems (AMS) are becoming more popular in dairy farms. However, in the current Nordic test-day model (TDM) the variance components are estimated from conventional milking system (CMS) observations and the same variance components are used for AMS herds. Measurement errors associated with test-day observations from herds with AMS are different due to the different measuring and sampling procedure. Under CMS the test-day observation for milk yield is the sum of morning and evening milking and under AMS it is the weekly sum divided by 7. Protein and fat content is measured from one milk sample in both milking systems.

In the variance component analyses for the Nordic TDM, separate residual covariance matrices for milk, protein and fat were estimated for 12 intervals within each lactation. The estimated variance components were used to derive covariance functions (CF) across traits and lactations. The CFs were formed in a way that differences in the 12 residual covariance matrices are explained by the CF for the non-genetic animal effect and by one residual covariance matrix for the measurement error part.

In this paper we present an approach to estimate separate measurement error covariance matrices for AMS and CMS observations.

Materials and Methods

Datasets

We sampled two Danish Holstein data sets of different size from data used in the Nordic TDM evaluation. A small data set (data 1) including 16 AMS and 24 CMS, and a large data set (data 2) including 40 AMS and 60 CMS randomly selected herds. Herds were required to have on average at least nine primiparious cows every year. From the sampled herds all observations recorded between 2001 and 2010 were included in the analyses. A herd was defined as AMS herd if it has started to use automatic milking system before 2010. Therefore, the majority of the sampled AMS herds were actually CMS herds in the beginning of the sampling years. Thus, the desired ratio of AMS/CMS herds is achieved at the end of the sampling years. The data sets are described in the Table 1.

A) Approach for estimation of residual (co)variances for automatic and conventional milking systems

Let's, in a simplified way, describe the model components for the estimation of variance components to be

$$Y = Xb + htd + \Phi p + \Phi a + e_{vce} \quad (1)$$

and the model components for the Nordic

TDM be

$$Y = Xb + htd + S_p \pi + S_a \alpha + e_{cf}\ , \quad (2)$$

where **X** is an incidence matrix for fixed effects **b**, **htd** are random herd test-day effects, $\boldsymbol{\Phi}_p$ and $\boldsymbol{\Phi}_a$ are matrices associating non-genetic animal effects **p** and genetic animal effects **a** to an observation, respectively and $\mathbf{e}_{vce}$ is the random measurement error vector. Elements of $\boldsymbol{S}_p\boldsymbol{\pi}$ and $\boldsymbol{S}_a\boldsymbol{\alpha}$ are covariance functions CF_p and CF_a for non-genetic, and genetic animal effects, respectively, which were derived from variance components estimated by model (1), and where var($\boldsymbol{\pi}$) = **P**, var($\boldsymbol{\alpha}$) = A⊗**G**, and var($\mathbf{e}_{cf}$) = $\mathbf{E}_{cf}$

Assuming that the differences between residual variances for different milking systems over the lactation is constant, and that there is no milking system interaction between the other variance components in the model, then there is no need to re-estimate all variance components for all random effects in model (1), even earlier variance components estimation by model (1) was based on CMS observations only. See Mulder et al. (2004) for discussion about Genotype×Environment interaction. Hence, only the measurement error covariance matrices for AMS ($\mathbf{E}_{\mathbf{AMS}}$) and CMS ($\mathbf{E}_{\mathbf{CMS}}$) have to be re-estimated while keeping already available $\boldsymbol{S}_\mathbf{p}$, $\boldsymbol{S}_\mathbf{a}$, **P** and **G** as fixed.

Updating the routine model with a (co)variance matrix for AMS observations can be achieve with minimum changes by keeping the original $\mathbf{E}_{cf}$ for CMS observations unchanged, but modifying the estimate $\mathbf{E}_{ams}$ covariance matrix by preserving the estimated correlations between traits and scaling the variances and covariances to have same ratios between updated CMS and AMS matrices as those obtained by the re-estimation.

B) Evaluation of estimation approach

The two data sets were used to evaluate the estimation approach. Variance components for the first lactation were estimated by applying model (1) where traits were milk, protein and fat yield. The fixed effects in ***b*** were age at calving, days carried calf and fixed lactation curve nested within 2-year calving period. The random effects were modeled as described in Lidauer *et al.* (2009) but with the differences that here residual covariance matrices were estimated for AMS observations as well. Residual covariance matrices were nested within 12 days in milk (DIM) intervals yielding 24 3×3 matrices to be estimated. Variance components were estimated separately for both data sets using a MC-EM REML algorithm (Matilainen *et al.*, 2012).

After the variance component estimation, covariance functions for both data sets and milking systems were fitted by a procedure described in Koivula *et al.* (2004). The CF for AMS were fitted using the AMS residual covariance matrices and for CMS the CMS residual covariance matrices were used. The CFs fitted for CMS observations were considered to be "true" underlying CF and were utilized to obtain $\mathbf{S}_{p1,}$ $\mathbf{S}_{a1}$, $\mathbf{S}_{p2}$ and $\mathbf{S}_{a2}$ for data sets 1 and 2 respectively. Also the fitted measurement error covariance matrices $\mathbf{E}_{AMS1}$, $\mathbf{E}_{CMS1}$, $\mathbf{E}_{AMS2}$ and $\mathbf{E}_{CMS2}$ were obtained for both data sets during the covariance function fitting. From here onwards the measurement error covariance matrices $\mathbf{E}_{AMS1}$ and $\mathbf{E}_{AMS2}$ are named "true-fitted" measurement error covariance matrix for AMS observations of data set 1 and 2, respectively.

Then the variance component estimation was carried out a second time by swapping **S** matrices between data sets, thus by using $\mathbf{S}_{p2}$ and $\mathbf{S}_{a2}$ as random effects for data 1 and S_{p1} and $\mathbf{S}_{a1}$ for data 2 and keeping corresponding variance parameters as fixed during the variance component estimation. Only the measurement error covariance matrices were re-estimated and are denoted by $\hat{\mathbf{E}}_{AMS1}$, $\hat{\mathbf{E}}_{CMS1}$,

$\hat{\mathbf{E}}_{AMS2}$ and $\hat{\mathbf{E}}_{CMS2}$. The element wise ratios of re-estimated covariance matrix components between milking systems were calculated and compared to the "true-fitted" element wise ratios of $\mathbf{E}_{AMS1}$ and $\mathbf{E}_{CMS1}$ (E_{R1}) and $\mathbf{E}_{AMS2}$, $\mathbf{E}_{CMS2}$ (E_{R2})

C)Variance component estimation

The proposed approach in A) for the estimation of AMS and CMS measurement covariance matrices was applied on data 2 for all three lactations. All other variance components in the model were kept fixed and were the same as in the Nordic Holstein routine evaluation model, which were originally estimated from Swedish Holstein data.

Results & Discussion

Evaluation of estimation approach

The "true-fitted" measurement error covariance matrices $\mathbf{E}_{AMS1}$, $\mathbf{E}_{CMS1}$, $\mathbf{E}_{AMS2}$ and $\mathbf{E}_{CMS2}$ and correlations are presented in Table 2 and were obtained from fitting CFs to the originally estimated variance components, which included 12 residual covariance matrices for each type of observations. Rank of the fitted CFs was reduced to seven for both, non-genetic ($\mathbf{S}_p\mathbf{P}\mathbf{S}_p'$) and additive genetic animal ($\mathbf{S}_a\mathbf{G}\mathbf{S}_a'$) effect.

Differences in estimates were found between data sets and milking systems. In the both data sets AMS had smaller measurement error variances for milk and protein and higher variance for fat compared to CMS. Also the correlations between traits were smaller in AMS than in CMS. All measurement error variance components and correlations for milk and protein were higher based on the analysis of the large data compared to the small data. However, for fat the estimates from both data sets were similar.

The difference in measurement error variances affected daily heritability curves (Figure 1). The daily heritability was higher in AMS for milk and protein and lower for fat. The AMS heritability curve based on CMS $\mathbf{S}_{p2}$, $\mathbf{S}_{a2}$ and $\mathbf{E}_{AMS2}$ deviated slightly from the "true" curve at the beginning of lactation. This originated from non constant difference of residual variances between milking systems at the early stages of lactation. The difference was minor and the estimation approach can still be used. A similar pattern was found in protein and fat, although in protein there was virtually no difference.

The ratios of re-estimated measurement error variance components were in good agreement with the original ones (Table 3) although the $\mathbf{S}_p$ and $\mathbf{S}_a$ from the other data were used. This shows that the measurement error variance ratio of AMS and CMS was not sensitive to $\mathbf{S}_p$ and $\mathbf{S}_a$ matrices used. and the presented approach can be used even if the elements of those matrices are based on covariance functions derived from different data set.

Variance component estimation

The AMS measurement error variance components were estimated to be 18 to 34 percent smaller in all three lactations for milk and protein yield compared to CMS (Table 4). For fat the AMS measurement error variance was 40 to 51 percent higher than in CMS. The results for the first lactation were consistent with the original estimates (Table 3) although the applied routine evaluation model CF produces somewhat larger differences (smaller ratios) compared to those of the "true-fitted" estimates.

The 305d heritabilities for AMS and CMS were practically the same in all three lactations and traits (Table 5) even there was a clear difference in the daily heritabilities (Figure 1).

Conclusions

We presented an approach to estimate measurement error variances for AMS when certain conditions hold. The approach was validated by using two different data sets. The results showed that the estimation approach was suitable.

The results also indicated that there is a considerable difference in measurement error variances between observations from AMS and CMS. As automated milking systems are becoming more popular there is need to account for these differences in genetic evaluation models.

Acknowledgements

We thank NAV for providing the data for the analyses.

References

Koivula, M., Negussie, E. & Mäntysaari, E.A. 2004. Genetic parameters for test-day somatic cell count at different lactation stages of Finnish dairy cattle. *Livestock Production Science 90*, 145-157.

Lidauer, M.H., Madsen, P., Matilainen, K., Mäntysaari, E.A., Strandén, I., Thompson, R., Pösö, J., Pedersen, J., Nielsen, U.S., Eriksson, J.-Å., Johansson, K. & Aamand, G.P. 2009. Estimation of variance components for Nordic red cattle test-day model: Bayesian Gibbs sampler vs. Monte Carlo EM REML. *Interbull Bulletin 40*, 37-41.

Matilainen, K., Mäntysaari, E.A., Lidauer, M.H., Strandén, I. & Thompson, R. 2012. Employing a Monte Carlo algorithm in expectation maximization restricted maximum likelihood estimation of the linear mixed model. *Journal of Animal Breeding and Genetics* (IN PRESS), doi:10.1111/j.1439-0388.2012.01000.x

Mulder, H.A., Groen, A.F., de Jong, G. & Bijma, P. 2004 Genotype × Environment Interaction for Yield and Somatic Cell Score with Automatic and Conventional Milking Systems. *Journal of Dairy Science. 80*, 1487-1495.

Table 1. Descriptive statistics for observations from first lactation**.**

	Data 1			Data 2		
	AMS	**CMS**	**Total**	**AMS**	**CMS**	**total**
N herds	16	24	40	40	60	100
N animals	5183	15998	20620	12267	38084	49145
N obs	38228	131717		91839	320596	
Mean						
Milk kg	27.0	26.5		28.2	26.8	
Protein kg	0.91	0.89		0.95	0.89	
Fat kg	1.09	1.09		1.12	1.09	
Sd						
Milk kg	5.97	5.97		6.24	5.98	
Protein kg	0.18	0.18		0.19	0.18	
Fat kg	0.24	0.24		0.25	0.24	

Table 2. Measurement error variances (diagonal) covariances (upper triangle) and correlations (lower triangle) for the "true-fitted" matrices E_{AMS1}, E_{CMS1}, E_{AMS2} and E_{CMS2} which were obtained from fitting covariance functions to the estimated variance components for first lactation milk (kg), protein (kg) and fat (kg).

	Data I						Data II					
	E_{AMS1}			E_{CMS1}			E_{AMS2}			E_{CMS2}		
	Milk	**Protein**	**Fat**	**Milk**	**Protein**	**Fat**	**Milk**	**Protein**	**Fat**	**Milk**	**Protein**	**Fat**
Milk	3.21	0.100	0.103	5.08	0.167	0.180	3.96	0.126	0.128	5.39	0.176	0.189
Protein	0.78	0.005	0.005	0.91	0.007	0.007	0.84	0.006	0.005	0.92	0.068	0.007
Fat	0.39	0.45	0.021	0.67	0.67	0.145	0.44	0.48	0.021	0.66	0.67	0.015

Table 3. The element wise ratios of "true-fitted" measurement error covariance matrices E_{AMS1}, E_{CMS1} (E_{R1}), E_{AMS2}, E_{CMS2} (E_{R2}) and re-estimated measurement error covariance matrices $\hat{E}_{AMS1}$, $\hat{E}_{CMS1}$ ($\hat{E}_{R1}$) and $\hat{E}_{AMS2}$, $\hat{E}_{CMS2}$ ($\hat{E}_{R2}$) for first lactation milk, protein and fat.

	Data I						Data II					
	E_{R1}			$\hat{E}_{R1}$			E_{R2}			$\hat{E}_{R2}$		
	Milk	**Protein**	**Fat**	**Milk**	**Protein**	**Fat**	**Milk**	**Protein**	**Fat**	**Milk**	**Protein**	**Fat**
Milk	0.63	0.60	0.57	0.62	0.59	0.52	0.73	0.72	0.68	0.72	0.69	0.64
Protein		0.77	0.72		0.78	0.68		0.84	0.78		0.83	0.75
Fat			1.47			1.45			1.42			1.41

Table 4. Ratios of estimated AMS and CMS measurement error covariance parameters from data II for milk, protein and fat of all three lactations, when applying covariance functions of the routine evaluation model and keeping variance components of other random effects as fixed.

	Lactation 1			Lactation 2			Lactation 3		
	Milk	**Prot**	**Fat**	**Milk**	**Prot**	**Fat**	**Milk**	**Prot**	**Fat**
Milk	0.71	0.69	0.65	0.71	0.69	0.65	0.66	0.65	0.61
Protein		0.82	0.76		0.81	0.76		0.77	0.69
Fat			1.40			1.51			1.45

Table 5. 305d heritabilities for AMS and CMS based on re-estimated measurement error variance components and on covariance functions applied in the Nordic Holstein routine evaluation model.

	Milk		Protein		Fat	
	ams	**cms**	**ams**	**cms**	**ams**	**cms**
Lact1	0.39	0.39	0.35	0.35	0.38	0.39
Lact2	0.29	0.29	0.29	0.28	0.32	0.33
Lact3	0.25	0.25	0.26	0.26	0.28	0.29

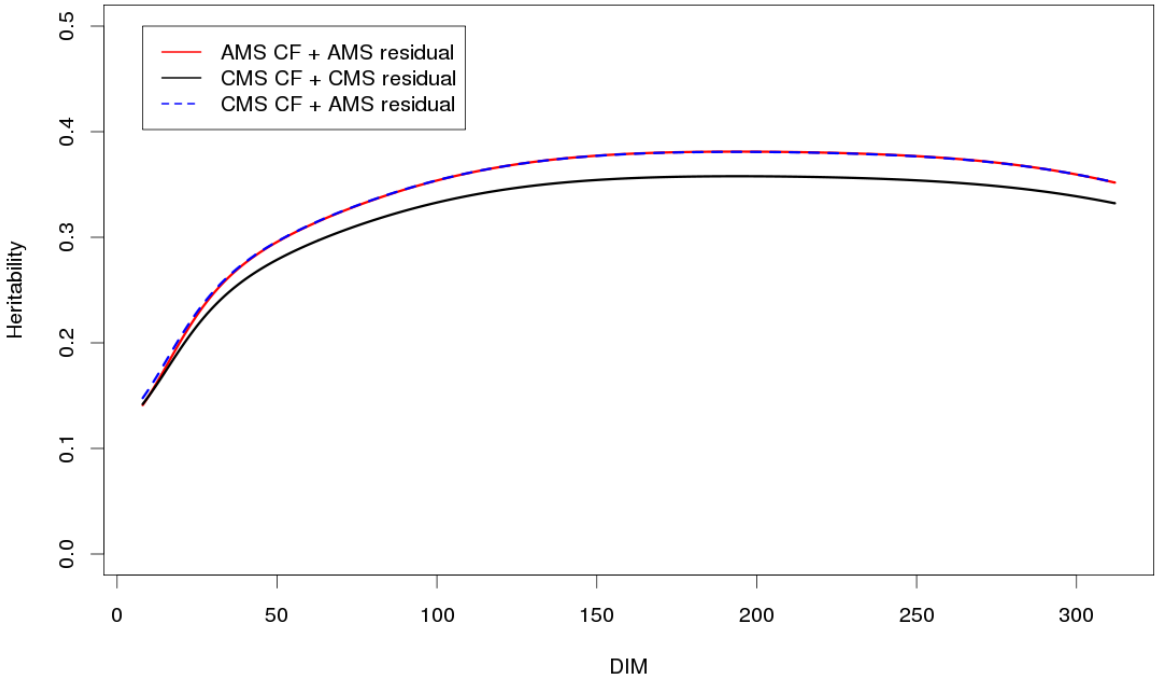

Figure 1. Daily heritability curves for first lactation milk for AMS (red), CMS (black) obtained from fitting covariance functions to original variance component estimates, and when AMS measurement error covariances were re-estimated by keeping CMS covariance functions ($\mathbf{S}_p$ and $\mathbf{S}_a$ and variance parameters) fixed for other random effects in the model (dashed blue).