

# Simulation study on heterogeneous variance adjustment for observations with different measurement error variance

Pitkänen, T. Mäntysaari, E. A., Nielsen, U. S.,  
Aamand, G. P., Madsen, P. and Lidauer, M. H.



**NAV**



**MTT**

# Background

- In the current Nordic yield evaluation models heterogeneity of variance is corrected with a multiplicative mixed model approach (Meuwissen et al., 1996)
- Currently correction does not take into account differences in residual variance between automated and conventional milking systems (AMS / CMS)
- In this simulation study we applied models which take differences in residual variance into account in HV correction
- Objective is to compare performance of different HV correction models and propose calibration approach for HV correction

# Simulation of data

- As a basis for the simulation study a real data set was sampled from the Danish Holstein yield evaluation
- Sampled data were from the years 2001 – 2011
- Observations for first lactation milk, protein and fat yield were used
- 600 herds were randomly sampled from herds which fulfilled certain criteria
- 240 herds were considered to be AMS and 360 CMS herds

# Structure of data

	AMS	CMS	Total
Herds	240	360	600
Animals	136 002	109 021	245 023
Records	1 102 550	905 032	2 007 582
N milk	1 094 497	899 015	1 993 512
N prot	1 094 497	899 015	1 993 512
N fat	1 093 469	898 192	1 991 661

# Model used to simulate data

- Data was simulated in three steps
  1. Test-day observations for milk, protein and fat yields were simulated applying a random regression test-day model
    - $Y^* = Xb + Zu + e$
  2. Heterogeneity of variance for stratum  $i$  (herd x test-month) was simulated with model
    - $\xi_i = b_{i1} + b_{i2} + \varepsilon_i$
    - $\lambda_i = c_m \exp(-0.5 \xi_i)$ , where  $b_{i1}$  and  $b_{i2}$  are heterogeneity factors and  $c_m$  is a scaling factor specific to the milking system
  3. Observations within stratum  $i$  were then obtained by
    - $Y_i = Y_i^*/\lambda_i$
- 10 independent data samples were simulated

# Model for test-day observations $Y^*$

- Fixed effects
  - Herd x year, year x month x milking system, lactation curve x calving year x season
- Random effects
  - Herd-test-day, non-genetic and genetic animal effect
  - All random effects were correlated over traits
  - Same variance components were applied for both milking systems
- Residual
  - Different residual (co)variance matrices for AMS and CMS observations
- Variance components used in simulation were estimated from Holstein data (presented Interbull 2012)

# Model for heterogeneity observations $\xi$

- Multiple-trait model
- Within a herd-test-month stratum variance was assumed to be homogeneous
- Fixed effect year-month x milking system
- Random effects herd-year and residual
- Herd-year had AR(1) correlation structure within trait and traits were correlated

# Solving multiplicative mixed effect model

- Multiplicative mixed effect model can be formulated as

$$Y_i \lambda_i = X_i b + Z_i u + e_i \quad [1]$$

$$\xi_i = b_{i1} + b_{i2} + \varepsilon_i \quad [2]$$

- Values for  $\lambda_i$  are updated
$$\lambda_i = \exp(-0.5(\hat{b}_{i1} + \hat{b}_{i2}))$$
- In each cycle mean model [1] and variance model [2] are iterated
- and whole process is repeated until convergence.
- Observations for the variance model are obtained from the residuals of the mean model



# Different models fitted to simulated data

- Control model for observations  $Y^*$
- Model with HV correction but without milking system interaction (HVnoMS, represents current evaluation model)
- Model with MS interaction in mean model and variance model (HVMS)
- Previous model with residual variance calibration
- To mimic real evaluation life we used different variance components for HV models than in simulated model

# Calibrated HV correction model

- Same model as HVMS but applied residual variances, needed for calculating heterogeneity observations, are calibrated
- After multiplicative mixed effect model is solved genetic variances for AMS and CMS cows are calculated with a full model sampling approach
- Ratio of genetic variances is used to calibrate residual variances applied in HV so that it yields same genetic variance for both milking systems
- Process is repeated until calibrated residual variances yield same genetic variances

# Results

- All results are calculated for cows from latest two year birth class having more than 5 observations
- Number of cows is 11420 for AMS 8109 for CMS
- Following results are presented
  - Ratio of 305 day genetic variances compared to true simulated variance
  - Correlations between true and estimated 305 day breeding values
  - Percentage of AMS cows in top1000 list

# Results, ratio of simulated and estimated 305 day genetic variances (mean of 10 replicates)

MODEL	MILK		PROTEIN		FAT	
	AMS	CMS	AMS	CMS	AMS	CMS
Control (Y*)	1.01	1.01	1.00	1.02	1.00	1.01
HVnoMS	1.14	0.81	1.02	0.78	0.75	0.89
HVMS	0.70	0.78	0.69	0.76	0.90	0.91
Calibrated HVMS	0.69	0.69	0.69	0.69	0.91	0.90

Control model is fitted to observations Y\* which do not have heterogeneity simulated

# Results, correlations between true and estimated breeding values

	MILK			PROTEIN			FAT		
	AMS	CMS	ALL	AMS	CMS	ALL	AMS	CMS	ALL
Control (Y*)	0.79	0.79	0.79	0.77	0.77	0.77	0.74	0.75	0.75
HVnoMS	0.77	0.77	0.77	0.75	0.75	0.75	0.72	0.74	0.72
HVMS	0.77	0.77	0.77	0.75	0.76	0.75	0.74	0.74	0.74
Calibrated HVMS	0.77	0.77	0.77	0.75	0.76	0.75	0.74	0.74	0.74

# Results, percentage of AMS cows in top1000 list (mean of 10 replicates)

	MILK	PROTEIN	FAT
SIMULATED	58	59	58
Control (Y*)	58	59	58
HVnoMS	59	55	40
HVMS	58	61	63
Calibrated HVMS	59	60	59

Control model is fitted to observations Y\* which do not have heterogeneity simulated

# Conclusions

- Differences in measurement error variances needs to be taken into account also in HV correction
- The applied multiplicative mixed effect model for HV adjustment was sensitive to incorrect assumptions about the residual variances
- Not accounting for differences in residual variances affected on the genetic variances and the ranking of superior animals, but effect on predictability was very little