Strategy for Estimation of Variance Components for the Joint Nordic Yield Evaluation

P. Madsen[1], M. H. Lidauer[2] and E. A. Mäntysaari[2]
[1]Faculty of Agricultural Sciences, Genetics and Biotechnology, Foulum, Denmark;
[2]MTT Agrifood Research Finland, Biotechnology and Food Science, Biometrical Genetics, Jokioinen, Finland

1. Introduction

In 2006 Denmark, Finland and Sweden introduced a common genetic evaluation of yield traits for Red Cattle, Holstein and Jersey. The evaluation is based on test-day (TD) records from Denmark and Finland, and 305d records from Sweden. The applied meta-model accommodates the different data sources and consent to a unique set of breeding values across the three countries (Mäntysaari *et al.*, 2006). This is achieved by modelling a genetic correlation of unity across countries, but allowing different variances and heritabilities. Because in Finland content traits are measured bimonthly, the traits milk, protein, and fat yield are evaluated simultaneously. Considering the first three lactations as different traits, leads for the main breeds to a multiple-trait model with 27 traits. Currently the models are upgraded to replace the Swedish 305d records by TD records, which requires new variance components (VC) for the Swedish data. Moreover, used VC for the Finnish data were estimated in 1997, which made it obvious to update the VC for all the breeds and countries.

VC analysis (VCA) for random regression (RR) models confronts with highly over-parameterized models, which causes slow convergence of the models (Bohmanova *et al.*, 2007). Use of VC from such models may result in instability of estimated breeding values (Canavesi *et al.*, 2007). It would be sensible reducing the rank of the VC matrices while estimating the VC.

It was decided to apply Bayesian method via Gibbs sampling for the VCA. Rank reduction was not supported by the chosen software, and therefore large effort was made to design the VCA. The objectives of this study were finding 1) a suitable model, 2) the size and number of samples, and 3) performing a post-Gibbs analysis.

2. Material and methods

2.1 Chosen model

Based on extensive model comparisons on different data sets and breeds, the following multiple-trait RR animal model was chosen:

$$y_{ijkmnoqt} = H_{ijt} + A_{kt} + D_{mt} + \sum_{r=0}^{4} L_{jrt}\varphi_r(d) + h_{int} +$$
$$\sum_{r=1}^{3} c_{ijtr}\lambda_r(d) + \sum_{r=0}^{3} a_{otr}\lambda_r(d) + \sum_{r=0}^{3} p_{otr}\lambda_r(d) + e(s)_{ijkmnoqt}$$

Fixed effects were herd×2-years-calving-period (*H*), calving age (*A*), days carried calf (*D*) and regressions on days in milk (*d*) within 2-years-calving period (*L*). Random effects were herd-test-day (*h*), RR for herd×2-years-calving-period (*c*), additive genetic (*a*), and non-genetic (*p*) effects. The random error (*e*) was nested within 12 days in milk (DIM) classes (*s*) from DIM 8 to DIM 365 with different intervals: 3× 2 weeks, 3× 3 weeks, 3× 7 weeks, and 3× 5 weeks. Fixed, and random regressions were fitted with a 3rd, and 2nd order Legendre polynomial, respectively. In addition, the term *exp(-0.04d)* was added to all regression functions yielding φ of size five and λ of size four. For certain data samples, also regressions on heterosis and breed proportion were modelled.

2.2 Size and number of data samples

To make inference whether there is a true difference in heritabilities across countries, or differences are merely due to sampling, two reasonable large data sets were sampled for each country×breed VCA. For each data set as many herds were sampled as needed to comprise about 20,000 cows with records. It was required that a sampled herd had at least ten first calvers in each calving year from 1995 to 2006. Twelve data sets were sampled for the breeds Holstein, Red

Cattle and Jersey and the three countries. The data sets included between 320,881 and 383,119 recodes on milk, protein and fat yield from the first three lactations. The pedigrees compiled between 28,333 and 31,337 animals.

First lactation VC analyses. VCA on first lactation traits was deemed to be sufficient to make inference whether there is a difference in heritabilities across countries. Further, over-parameterization is less severe than when including all three lactations.

Three lactation VC analyses. The full three lactation model analyses (nine traits) were started with one sample per breed and country. Here we present the first results from the Danish Holstein sample only.

2. 3 Bayesian inference

All VCA were conducted by a Gibbs sampler implemented in the DMU-package (Madsen & Jensen, 2008). Flat priors were assumed for fixed effects and Wishart distributions for all random effects. Starting and prior values for VC were obtained form analysis of samples of TD records from ~1000 cows. To make the priors proper, the degree of belief was set to dimension of the covariance matrix + 2. A chain length of 110.000 was used, where the first 10.000 samples was discarded as burn-in.

Convergence of the Gibbs sampler was checked by the method of batching which also estimate the effective posterior sample size. The post Gibbs analysis was conducted both on the VC and on derived parameters such as heritabilities and correlations.

3. Results

3.1 Differences in VC across countries

Genetic correlations between different DIM and traits were highly consistent across the different analyses within the three main breeds. Average difference between two analysis was between -0.03 (Finnish Ayrshire *vs* Swedish Red Breed) and 0.07 (Finnish Ayrshire *vs* Red Danish Cattle).

Non-genetic animal effect correlations between different DIM and traits were very similar within samples from the same country (average difference smaller than 0.01) but somewhat different between samples from different countries (average difference between -0.05 and 0.07).

Heritabilities were different between countries. (Table 1). Differences were between 0.0% and 14% between samples within countries, but heritabilities from Danish and Swedish samples were on average 35% and 21% higher than those from Finnish samples, respectively.

First lactation analyses suggested that one sample per country×breed combination is sufficient for the VCA.

Table 1. Heritability estimates for first lactation traits on a 305d basis for Nordic Red Cattle by different data samples.

	Sample				
	Denmark	Finland		Sweden	
Trait	I*	I	II	I	II
Milk	0.48	0.35	0.36	0.41	0.42
Protein	0.44	0.31	0.31	0.38	0.41
Fat	0.43	0.34	0.33	0.37	0.42

* Results were very similar within country for Finland and Sweden, and therefore, only one sample was analysed for Denmark.

3.2 Post Gibbs analysis

Post Gibbs analysis for the VC showed poor mixing properties for several of the VC, and that a much longer burn-in is needed. As an example, trace plots of the 10 genetic VC for 1st lactation protein yield are shown in Figure 1.

Functions of VC such as heritabilities and correlations (non-genetic as well as genetic) computed for each sample of the VC had better mixing properties. This is illustrated in Figure 2, where trace plots of heritabilities for protein in 1st and 3rd lactation at DIM 30, 180 and 300 are shown. But also here a longer burn-in and chain is needed especially for 3rd lactation.

Parameters expressed on 305d basis shows even better mixing properties. This is illustrated by 305d heritabilities for protein yield in 1st, 2nd and 3rd lactation (Figure 3). Posterior means and SE for 305d heritabilities, phenotypic correlations and genetic correlations are shown in Table 2.

4. Discussion

The problem in obtaining convergence of the Gibbs sampler is probably due to a serious over-parameterization of the model. The dimension of the co-variance matrices for both the genetic and non-genetic animal effects was 36. Analysis on both co-variance matrices, obtained as the posterior means over the samples from round 10.000 to 110.000, showed that the first 12 eigenvalues explain more than 98% of the variation.

5. Conclusions

This study has shown that heritability for yield traits in 1st lactation for Nordic Red Cattle differs between the three countries. It has also shown problems in obtaining convergence of a Gibbs sampler for a 9-traits (3 biological traits × 3 lactations) model. The problem seems larger for individual VC than for functions such as heritabilities and correlations. A possible solution to the convergence problem could be a reduced rank model.

6. Acknowledgement

The authors are grateful to Nordic Cattle Genetic Evaluation for providing the data and for financial support.

References

Bohmanova, J., Jamrozik, J., Miglior, F., Misztal, I. & Sullivan, P.G. 2007. Legendre polynomials versus linear splines in the Canadian test-day model. *Interbull Bulletin 37*, 42-48.

Canavesi, F., Biffani, S. & Nicolazzi, E.L. 2007. Impact of new genetic parameters from a three trait model on the Italian Holstein genetic evaluation for production. *Interbull Bulletin 37*, 54-56.

Madsen, P. & Jensen, J. 2008. *A Users's Guide to DMU*. A package for analysing multivariate mixed models.Version 6, release 4.7. Faculty of Agricultural Sciences, University of Aarhus. Denmark. 33 pp.

Mäntysaari, E.A., Lidauer, M., Pösö, J., Madsen, P., Strandén, I., Pedersen, J., Nielsen, U.S., Johansson, K., Eriksson, J-Å. & Aamand, G.P. 2006. Joint Nordic test day model: variance components. *Interbull Bulletin*. *35*, 97-100.

Table 2. Posterior means and standard deviation (in brackets) for heritabilities, genetic correlations (above diagonal) and phenotypic correlations (below diagonal) expressed as 305d parameters for Danish Holstein.

Trait	h^2	ESS[a)]	Milk 1st	Prot. 1st	Fat 1st	Milk 2nd	Prot. 2nd	Fat 2nd	Milk 3rd	Prot. 3rd	Fat 3rd
Milk 1st	.42(.02)	90.2		.87(.01)	.48(.03)	.92(.02)	.77(.02)	.32(.04)	.89(.02)	.69(.04)	.26(.05)
Prot. 1st	.39(.02)	85.5	.92(.01)		.65(.02)	.76(.03)	.90(.02)	.48(.04)	.77(.03)	.86(.03)	.46(.05)
Fat 1st	.43(.02)	152.6	.72(.01)	.80(.01)		.41(.08)	.64(.03)	.92(.01)	.43(.03)	.63(.04)	.89(.02)
Milk 2nd	.30(.02)	60.5	.58(.01)	.53(.01)	.36(.01)		.81(.01)	.38(.04)	.93(.02)	.70(.04)	.28(.05)
Prot. 2nd	.27(.02)	60.1	.52(.01)	.60(.01)	.46(.01)	.92(.01)		.62(.03)	.79(.03)	.91(.03)	.55(.04)
Fat 2nd	.36(.02)	63.9	.34(.01)	.43(.01)	.62(.01)	.23(.01)	.83(.01)		.37(.05)	.59(.04)	.93(.02)
Milk 3rd	.29(.03)	30.5	.53(.01)	.48(.01)	.33(.01)	.57.01)	.54(.01)	.38(.01)		.81(.02)	.37(.05)
Prot. 3rd	.28(.03)	29.9	.44(.01)	.52(.01)	.41(.02)	.51(.01)	.61(.01)	.48(.02)	.93(.02)		.63(.03)
Fat 3rd	.35(.03)	35.7	.29(.02)	.38(.02)	.55(.01)	.24(.02)	.46(.01)	.61(.01)	.76(.01)	.85(.01)	

[a)] ESS = effective sample size for heritabilities.

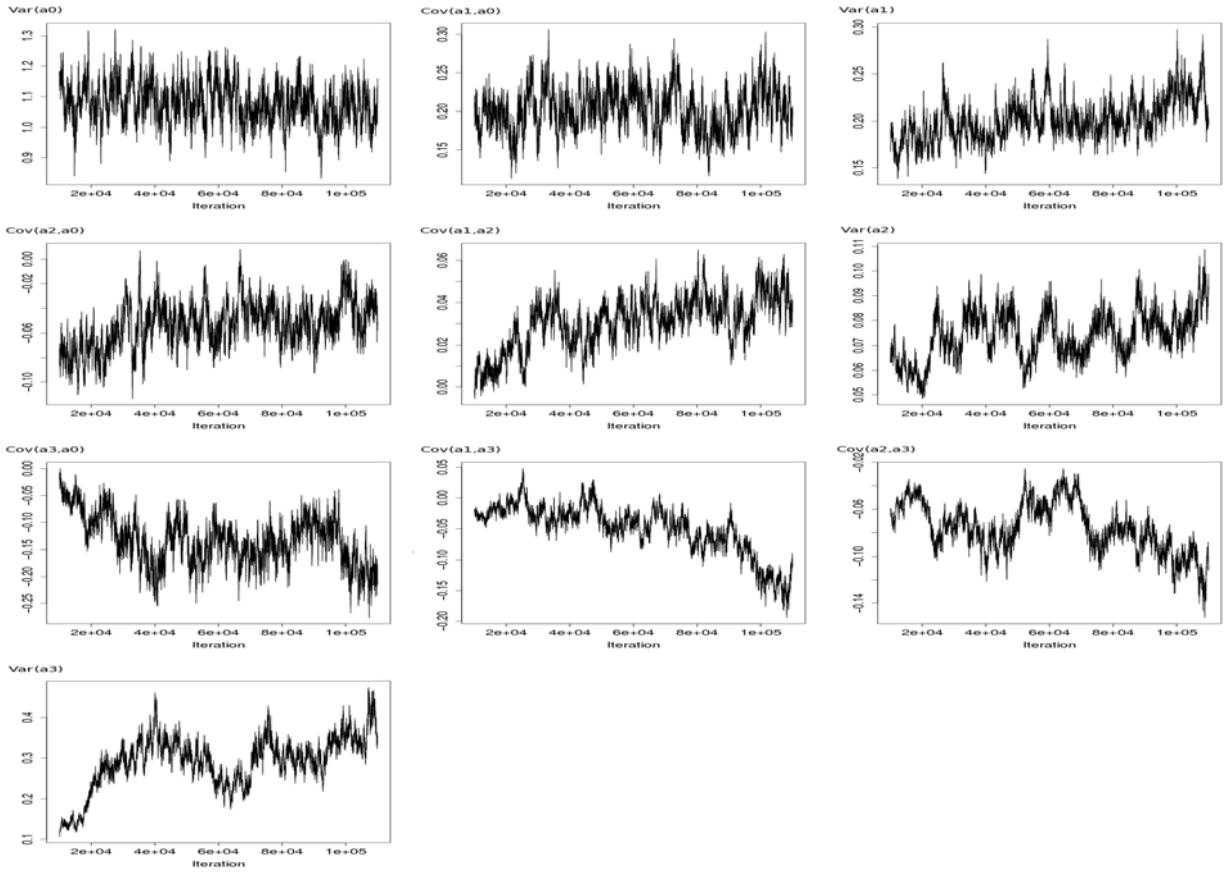

Figure 1. Trace plots of genetic (co)-variance components for 1st lactation protein yield.

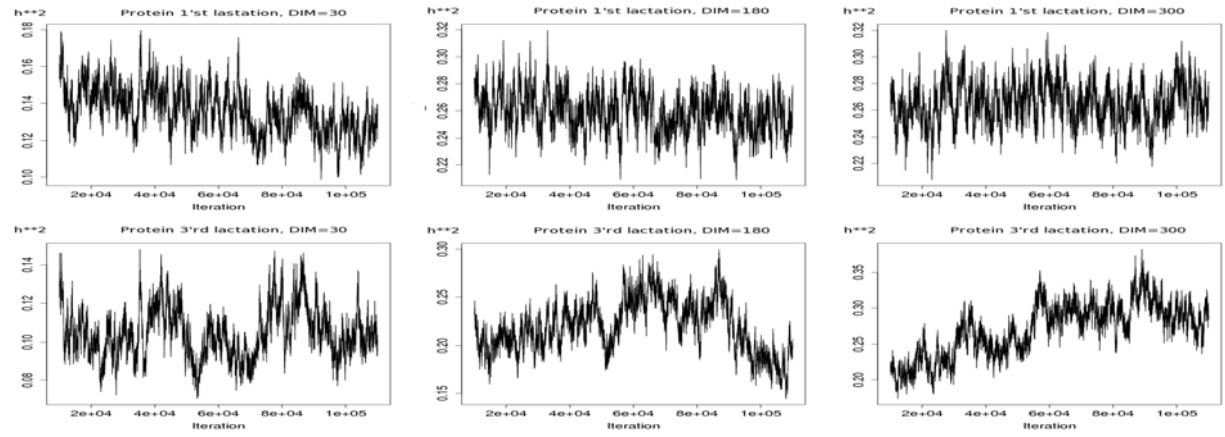

Figure 2. Trace plots of heritabilities for protein yield at DIM 30, 180 and 300 in 1st and 3rd lactation.

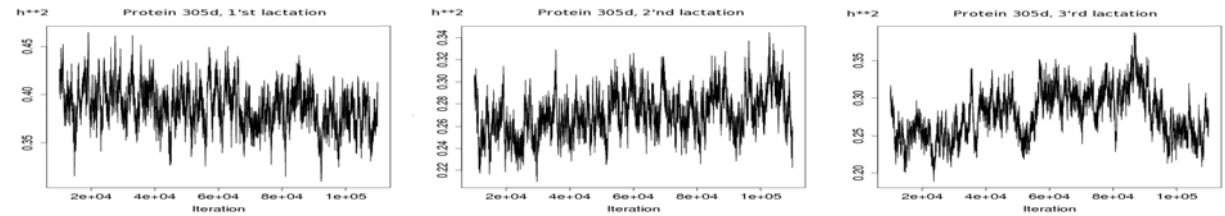

Figure 3. Trace plots of 305d heritabilities for protein yield in 1st, 2nd and 3rd lactation.