



## Estimation of variance components for Nordic red cattle test-day model: Bayesian Gibbs sampler vs. Monte Carlo EM REML

M.H. Lidauer<sup>1</sup>, P. Madsen<sup>2</sup>, K. Matilainen<sup>1</sup>, E.A. Mäntysaari<sup>1</sup>, I. Strandén<sup>1</sup>,  
R. Thompson<sup>3</sup>, J. Pösö<sup>4</sup>, J. Pedersen<sup>5</sup>, U.S. Nielsen<sup>5</sup>, J.-Å. Eriksson<sup>6</sup>,  
K. Johansson<sup>6</sup>, G.P. Aamand<sup>7</sup>

<sup>1</sup>*MTT Agrifood Research Finland*

<sup>2</sup>*Danish Institute of Agricultural Science*

<sup>3</sup>*Rothamsted Research, United Kingdom*

<sup>4</sup>*Faba Breeding, Finland*

<sup>5</sup>*The Danish Agricultural Advisory*

<sup>6</sup>*Swedish Dairy Association*

<sup>7</sup>*NAV Nordic Cattle Genetic Evaluation*



# Introduction

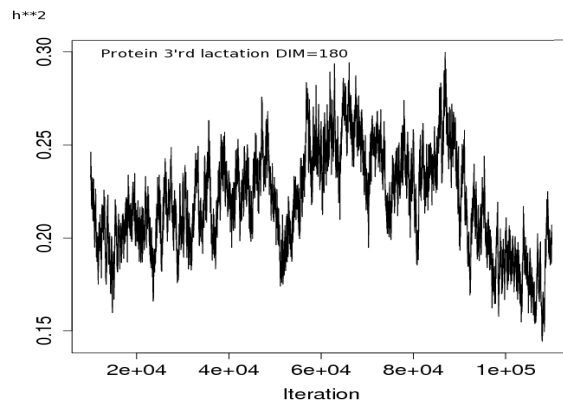
## Joint Nordic yield evaluation

- Countries: Denmark, Finland, Sweden
- Evaluation for each breed: Red Cattle, Holstein, Jersey
- Biological traits: milk, protein, fat
- Lactations: 1, 2, 3, (3+ for Finland)
- Model:  
reduced rank multiple-trait random regression test-day model

# Introduction

## Re-estimation of variance components

- Currently used variance components  
derived from multiple-trait REML analyses (estimated in 1998 and 2001)
- New variance components  
should be from random regression variance component analyses
- Bayesian analysis left us uncertain about quality of parameter estimates (Madsen et al., 2008)  
poor mixing properties for many variance components



Trace plot for daily heritability of 3<sup>rd</sup> lactation protein yield at days in milk 180

# Introduction

## REML by re-sampling

- Exact REML methods impossible for this size of analysis
- Re-sampling techniques for PEV calculation in REML algorithms (García-Cortés et al., 1992)
- Ongoing Ph. D. study on developing multivariate Monte Carlo REML algorithms (Kaarina Matilainen)
- A first beta-version is ready for testing

# Objective

**Estimation of random regression test-day model variance components by Monte Carlo EM REML, and compare estimates with those from Bayesian analysis**

# Data sets

	<b>Finnish Ayrshire*</b>	<b>Swedish Red Breed</b>	<b>Red Danish Cattle</b>
<b>Cows with observations</b>	<b>19 709</b>	<b>19 928</b>	<b>19 857</b>
<b>Number of herds</b>	<b>130</b>	<b>75</b>	<b>60</b>
<b>Number of TD records</b>	<b>374 009</b>	<b>353 340</b>	<b>320 881</b>
<b>Animals in pedigree</b>	<b>31 255</b>	<b>30 082</b>	<b>28 333</b>

\* For Finnish Ayrshire protein and fat yields were only available every second test-day

- **Test-day records on milk, protein and fat yield**
- **First three lactations**
- **Data size allows inference about differences in heritabilities for the three populations**

# Model for variance component analysis

- **9 traits**
  - milk, protein and fat yield of 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> lactation
- **Fixed effects**
  - herd×2-years calving period
  - calving age
  - days carried calf
  - Legendre polynomials + Wilmink term ( $e^{-0.04}$ )
    - 2-years calving period: *L0, L1, L2, L3 and W*
- **Random effects**
  - herd×test-day
  - Legendre polynomials + Wilmink term ( $e^{-0.04}$ )
    - herd×2-years calving period: *L1, L2 and W*
    - non-genetic animal effect: *L0, L1, L2 and W*
    - additive genetic animal effect: *L0, L1, L2 and W*
  - 12 residual classes (from DIM 8: 3× 2 weeks, 3× 3 weeks, 3× 7 weeks, 3× 5 weeks)
- **Variance components**
  - in total, 1971 variance components to be estimated for each breed

# Bayesian inference

- **Gibbs sampler implementation in the DMU package (Madsen & Jensen, 2008)**
- **Priors**
  - Fixed effects
    - flat priors
  - Random effects
    - Wishart distributions
    - prior values from analysis with records from ~1000 cows
    - proper priors: degree of belief was dimension of covariance matrix + 2
- **Chain length**
  - >170 000 samples



# Monte Carlo EM REML analysis

- **Monte Carlo EM REML algorithm based on ideas of García-Cortés et al. (1992)**
- **Algorithm developed for multivariate analysis**
  - Each REML round needs sums of squares from BLUP solutions of real and sampled data
  - 2 data sets were sampled in each REML round
  - BLUPs were solved by PCG (MiX99 package, Vuori et al., 2006)
- **Detailed description will be given by Matilainen et al. (2009) at the forthcoming EAAP meeting**

# RESULTS

# Post-Gibbs analysis

Effective sample size (Initial sequence estimators, Geyer 1992)

VC for Random Effect	Finnish Ayrshire	Swedish Red Breed	Red Danish Cattle
herd test-day	6855.8	6460.7	3164.6
herd curve	2659.5	2942.8	2345.6
non-genetic	290.3	164.9	225.7
genetic animal	57.2	42.8	42.6
Residual	7231.5	7117.5	4752.7

- **Available samples**

- Finnish Ayrshire: 269 720
- Swedish Red Breed: 246 840
- Red Danish Cattle: 189 800

- **Post-Gibbs analysis**

- First 70 000 samples discarded
- From remaining chain every 20<sup>th</sup> sample used
  - → Effective sample size
  - → posterior means

# Convergence of Monte Carlo EM REML

Number of EM REML rounds to reach convergence

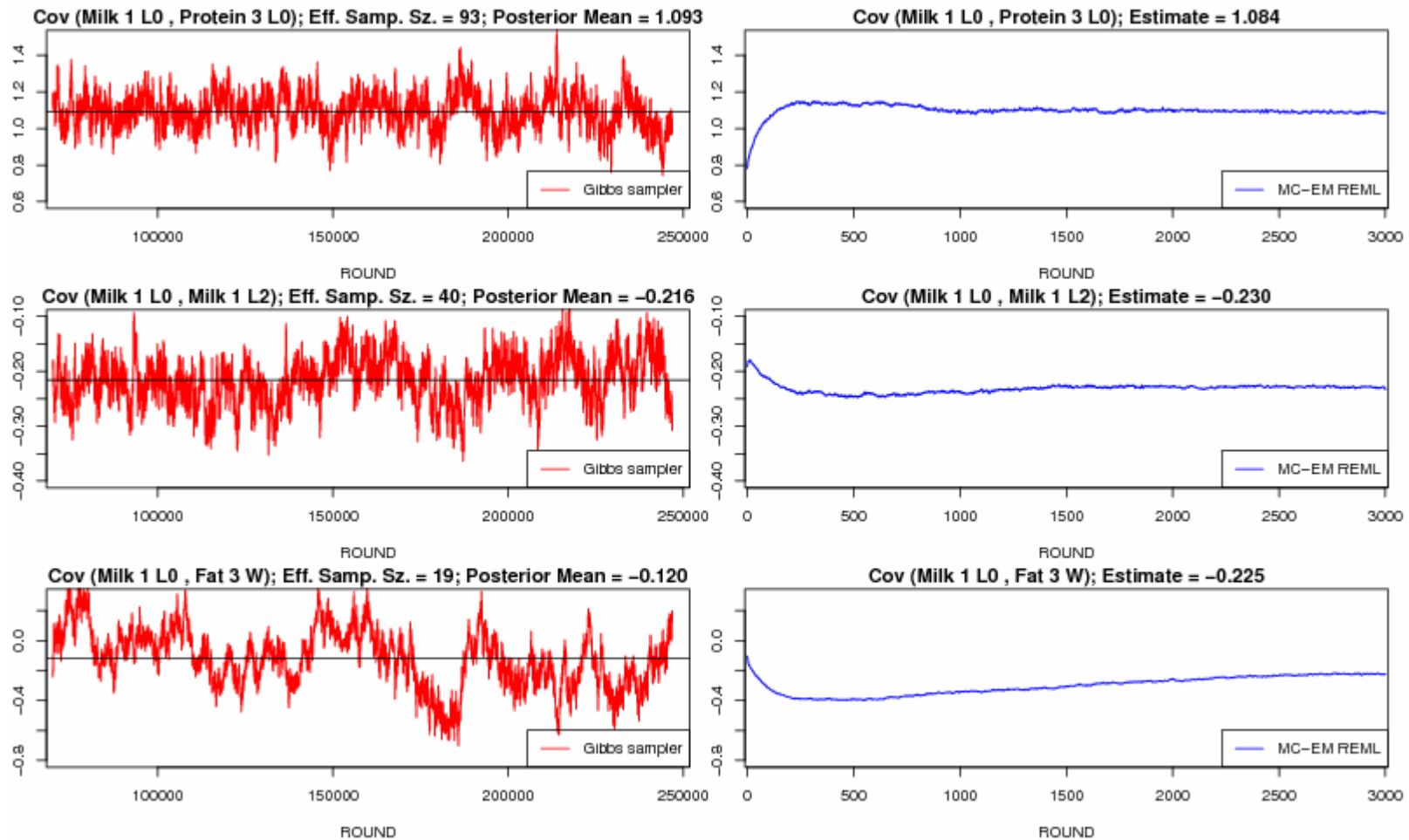
VC for Random Effect	Finnish Ayrshire	Swedish Red Breed	Red Danish Cattle
herd test-day	164	131	159
herd curve	466	566	451
non-genetic	2075	2535	3000*
genetic animal	1576	1954	2397
Residual	1652	609	462

\*converged to  $1.7 \times 10^{-8}$

- **Robust convergence behavior**
- **Estimates for co-variances between some additive genetic animal effect parameters converged slowly**
  - same parameters had low effective sample size in the Bayesian analysis

# 3 example parameters

Genetic co-variances between milk 1,  $L_0$ : & Protein 3,  $L_0$ ; & Milk 1,  $L_2$ ; & Fat 3,  $W$ .

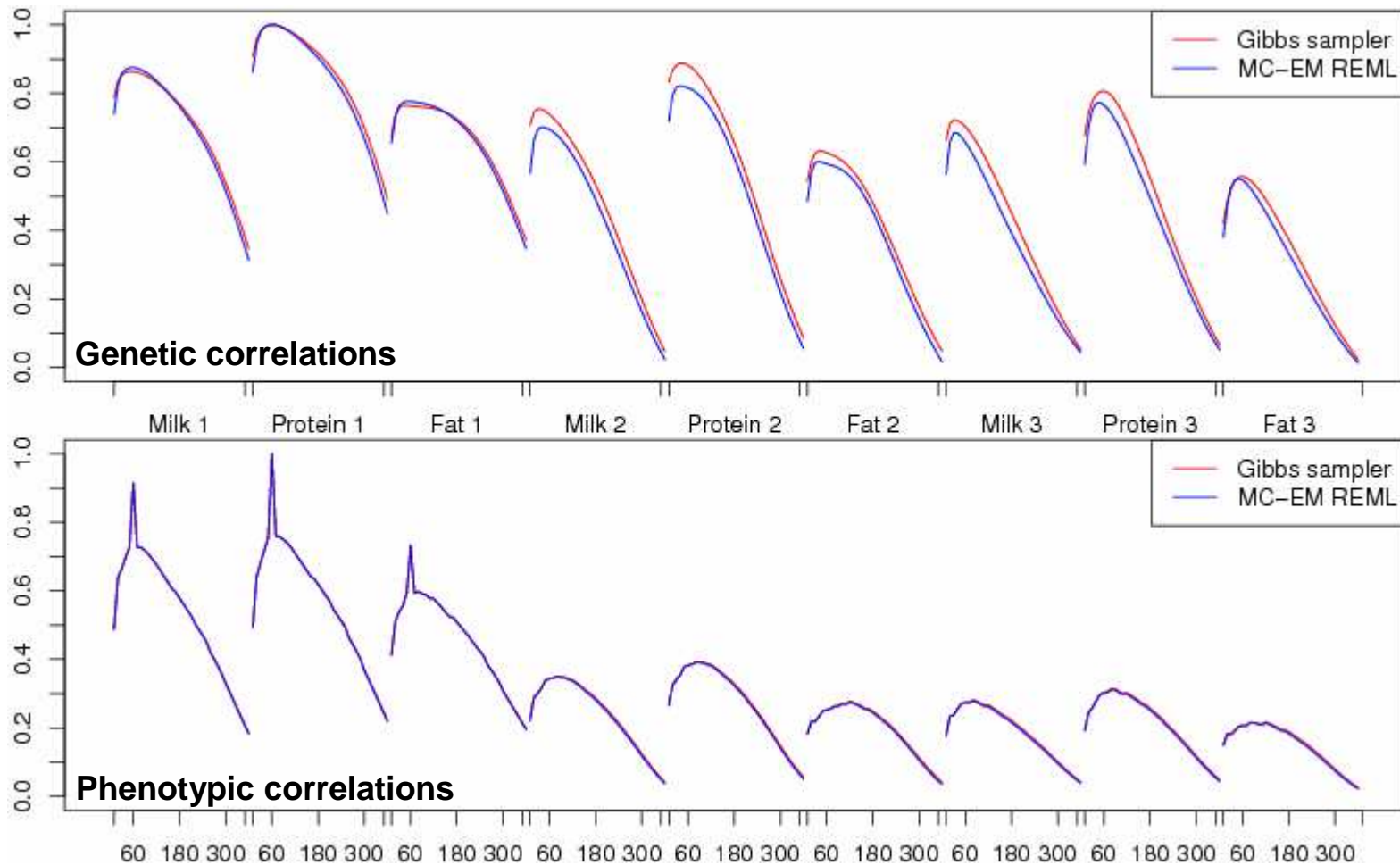


# Comparison of variance component estimates

- **VC estimates were in good agreement when effective sample size was high in post-Gibbs analysis**
- **Significant differences were found for genetic animal effect co-variances which included a Wilmink term**
- **Some significant differences were found for Finnish Ayrshire milk yield parameter estimates**

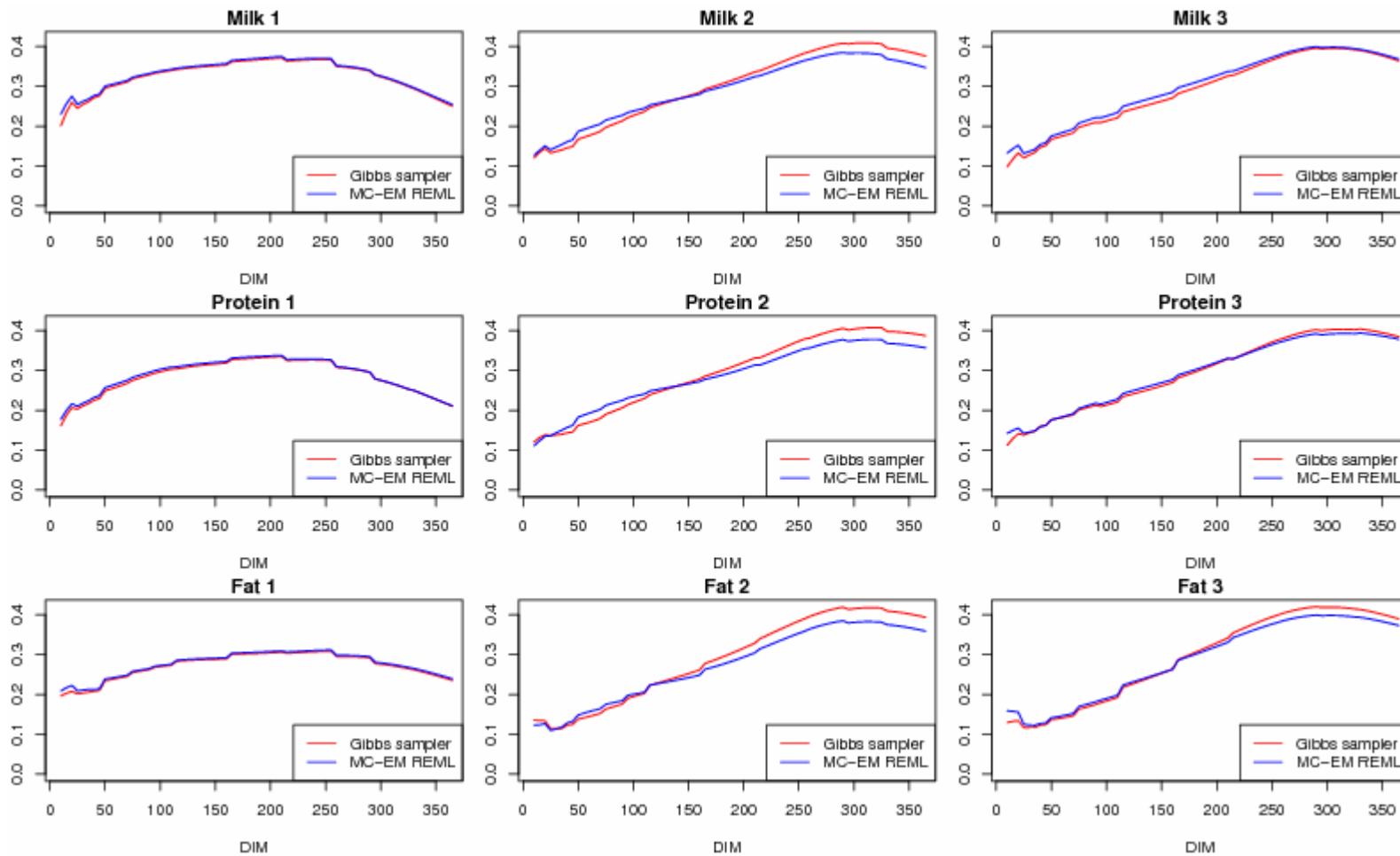
# Genetic and phenotypic correlations

Correlations between days in milk 60 of 1<sup>st</sup> lactation protein yield and all other days in milk and traits given for Swedish Red Breed



# Daily heritabilities

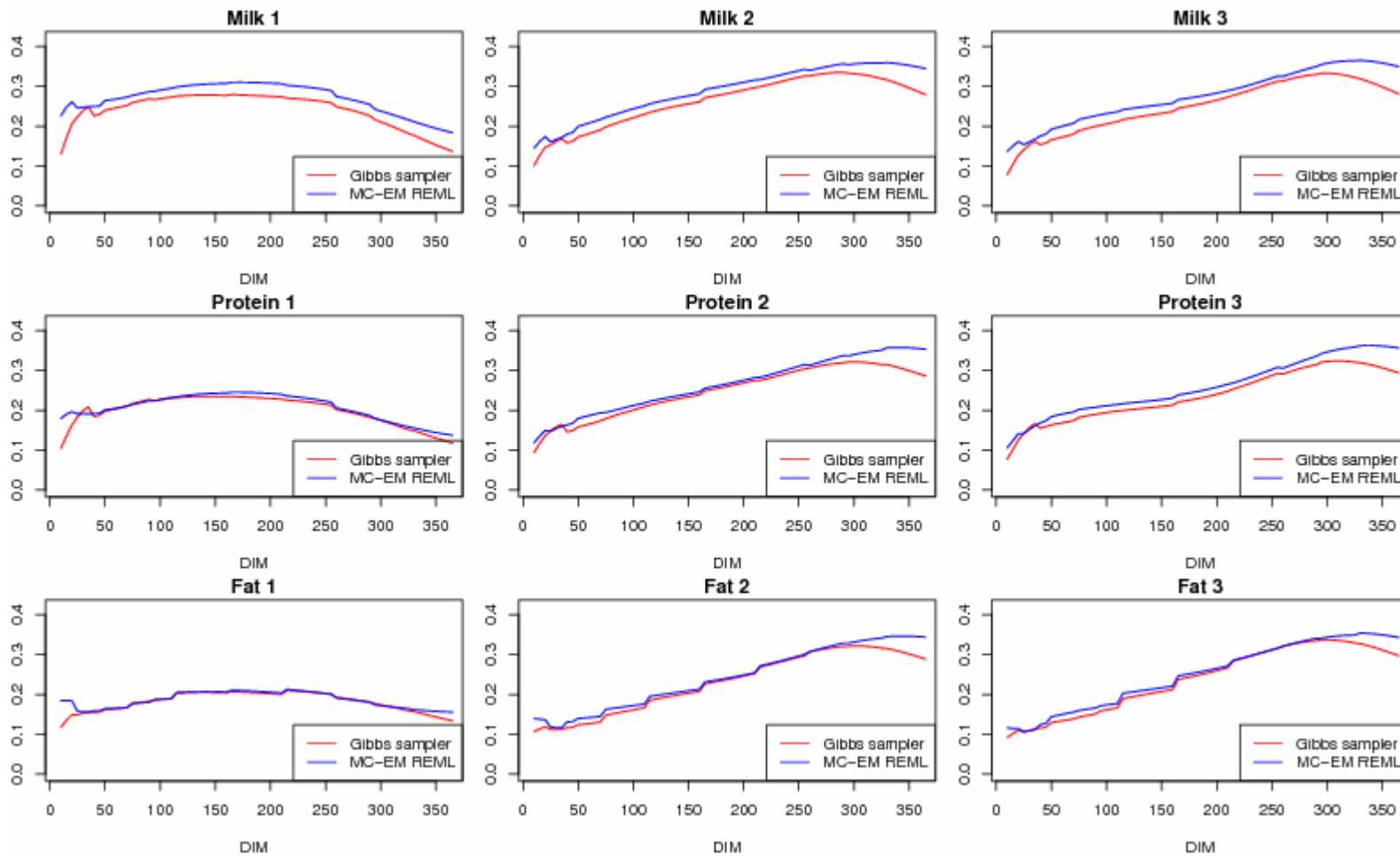
## Swedish Red Breed





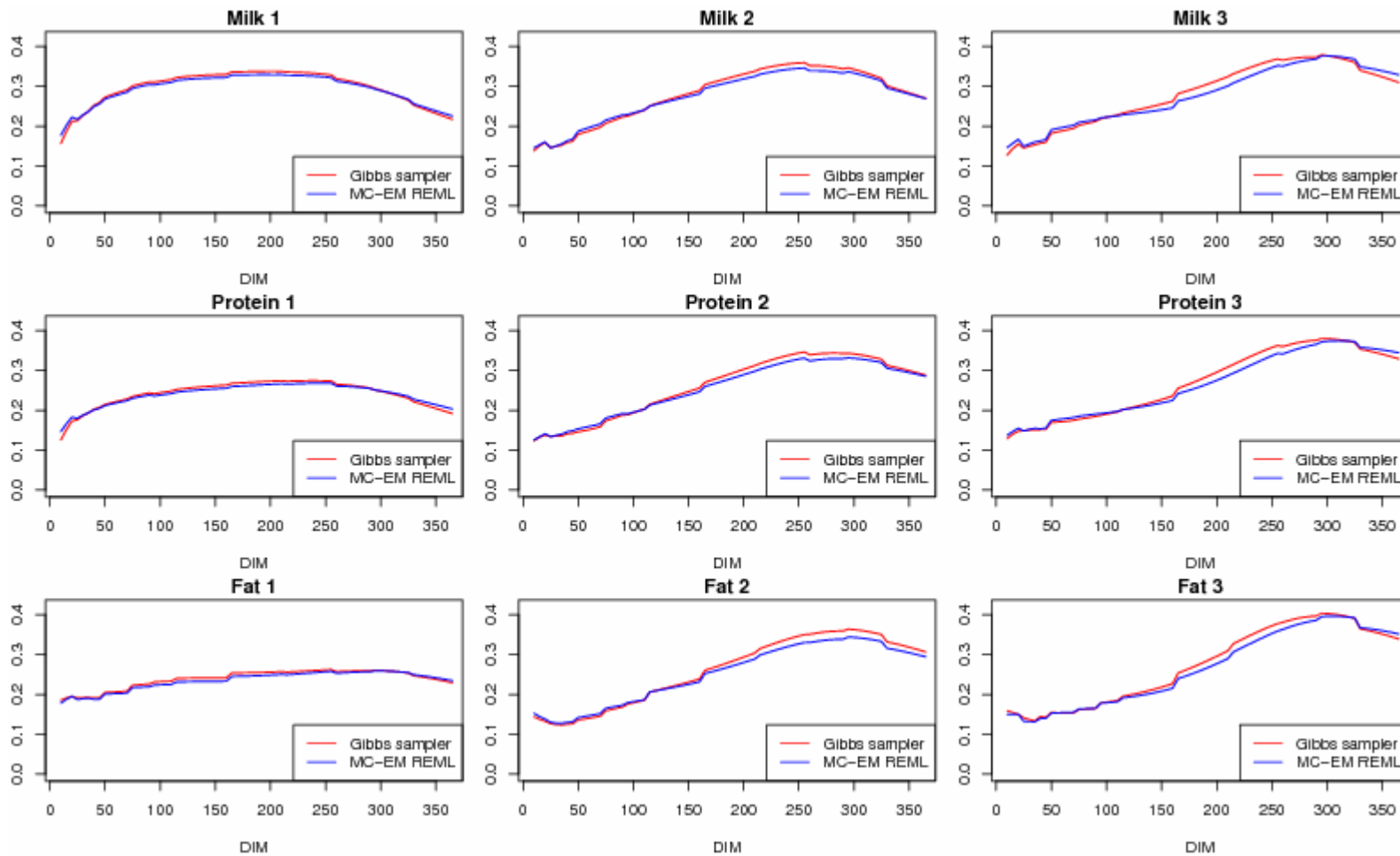
# Daily heritabilities

## Finnish Ayrshire



# Daily heritabilities

## Red Danish Cattle



# Computing time

- **1 MC-EM REML round ~ 11 Gibbs sampler rounds**
  - Based on software available at the time of study
- **Computing time for Finnish Ayrshire analysis**
  - 1 Gibbs sampler round 1.51 min.
    - **Total computing time for 269720 rounds: 282 days**
  - 1 MC-EM REML round 17.1 min.
    - **Total computing time for 3000 EM REML rounds: 35 days**

# Conclusions

- **Variance components estimated by Gibbs sampler or Monte Carlo EM REML were in good agreement**
- **Variance parameters having low effective sample size in the Gibbs sampler converged slowly in Monte Carlo EM REML**
- **Monte Carlo EM REML was superior in computing time and can become an alternative for large variance component analyses**

# Acknowledgement

**Further development and implementation of Monte Carlo EM REML is part of a Ph.D. study financed by Rothamsted Research and MTT**

