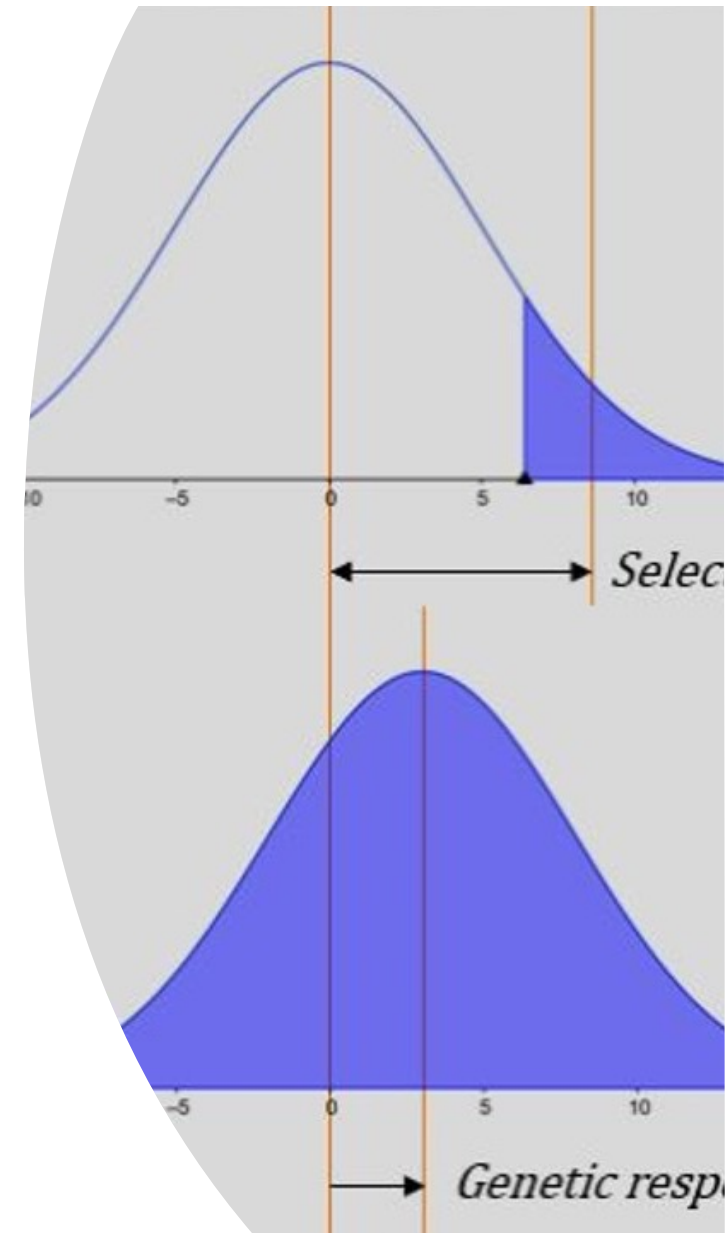


Mastering Monte Carlo based VCE

Martin Lidauer, Anna-Kaisa Ylitalo, Hongding Gao

MiX99 course on genomic prediction

COURSE DAY 2, March 11th, 2026



Contents

- [Balancing computation time and accuracy](#)
 - Optimizing the solving of the BLUP models
 - Number of Monte Carlo samples
 - Assessing convergence of VC (trace plots)
- [Calculation of standard errors](#)
- [Fixing variance components](#)



Balancing computation time and accuracy

Monte Carlo REML computation

- Monte Carlo (MC) REML algorithm avoids the inversion of MME's coefficient matrix
- This comes at the cost of solving the MME several times in each REML round
- Solving the mixed model equations accounts for almost all computing time in MC REML

Main computation time per REML round

- Solving the MME of the model for the real data
 - In each REML round, the MME solutions from the previous REML round are used as starting values
 - → solving time decreases with increasing REML rounds
- Solving the MME of the model for the sampled data
 - Each data sample requires to solve a sample-specific MME
 - The simulated model effects for generating a specific data sample are used as starting values



Optimizing the solving of the BLUP models

- DATA
 - Avoid unnecessary large data samples
 - Animal contemporary groups should be of reasonable size
- PEDIGREE
 - Pedigree pruning for variance component analysis (e.g., “[prune varcom](#)” option in RelaX2)
 - Limiting number of ancestor generations may considerably reduce size of pedigree (e.g., “[prune varcom 3](#)”)
- SOLVING OPTIONS
 - Optimal preconditioner: that should result shortest computing time (often, “[block-diagonal](#)” preconditioner for all effects)
 - Optimal stopping parameters



Optimal stopping parameters

- Make sure that max. number of PCG iterations is large enough
- Different convergence values can be specified for PCG iteration on real data, and PCG iteration on sampled data
- For more complex models, up to 3000 EM REML rounds are needed
- The convergence criterion specified in **STOP** option line also will be used for the PCG iteration on sampled data

Solver options file

```

#-----
# RAM: RAM options
#       H
# STOP: Max.iter,PCG-Conv.valREAL-DATA,Criterion,Enforce
#       5000      1.0e-5      d      f
# RESID: Residuals calculation
#       N
# VALID: Model validation
#       N
# VAROPT: Variance options VCE, PEV, HV
#       E
# STOPE: MaxREMLrnd,Samples,VC-Conv.val,[PCG-Conv.valSAMPLE-DATA]
#       3000      2      1.0e-10      3.2e-5
# SEED: Type of seed for the random number generator
#       R
# MIX99PATH:
#       /data/projects/animalgenetics/software/MiX99/bin_beta
# SOLTYP: Solution file options
#       Y
#-----

```

Convergence criterion for variance component parameter estimates

- Defining a suitable convergence criterion is difficult for MC REML estimates
- The criterion should be robust against MC noise in the VC estimates
- [Applied criterion for MC EM REML in MiX99](#)
 - Relative change between VC parameter estimates from REML round (n) and ($n-1$)
 - **Reducing MC noise in the criterion**
 1. Linear regression on the latest half of the estimates is fitted (estimates from round $1/2n$ to n)
 2. Variance component estimates of the latest REML round (n) are predicted
 3. Changes in predicted VC estimates of consecutive REML rounds are used for the criterion
- The criterion works satisfying when MC noise is small (with reasonable size of data)
- When MC noise is large, REML iterations may continue, even VC estimates have converged (analysing small data often requires larger number of samples)

Suitable convergence value

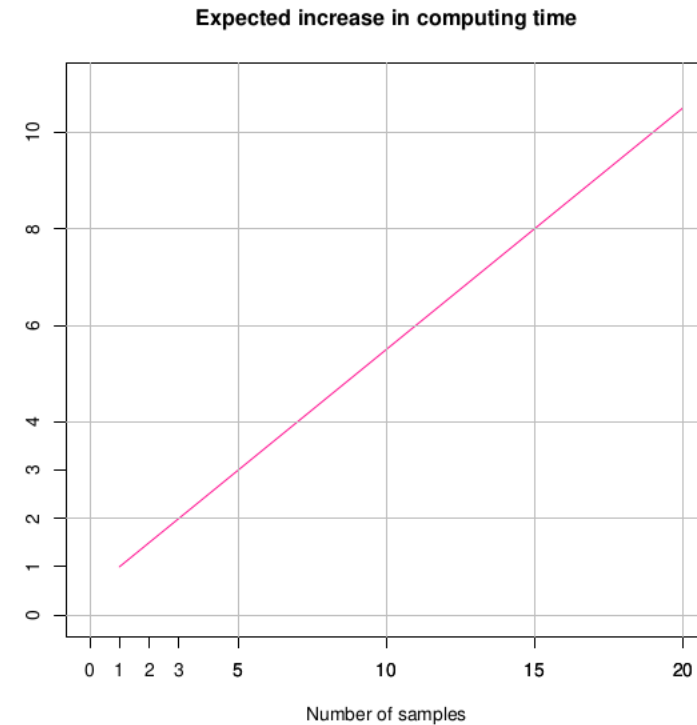
Convergence values developed from extensive VC analyses using a complex field data sample with 15 traits

Criterion		CR	CD
Convergence value for variance component parameters			
	Model with easy to estimate parameters		<1.0e-8
	Model with difficult to estimate parameters		<1.0e-10
Convergence value for PCG iterations on real data			
	Model with easy to estimate parameters	<1.0e-6	<3.2e-5
	Model with difficult to estimate parameters	<1.0e-7	<1.0e-5
Convergence value for PCG iteration on sampled data			
	Model with easy to estimate parameters	<1.0e-5	<1.0e-4
	Model with difficult to estimate parameters	<1.0e-6	<3.2e-5



Number of Monte Carlo samples

- Increasing number of MC sample reduces MC noise in the parameter estimates
- However, this comes at the cost of longer computing time
- For instance, increasing number of samples from 1 to 5 increases computing time 3-fold



Assessing convergence of VC parameter estimates

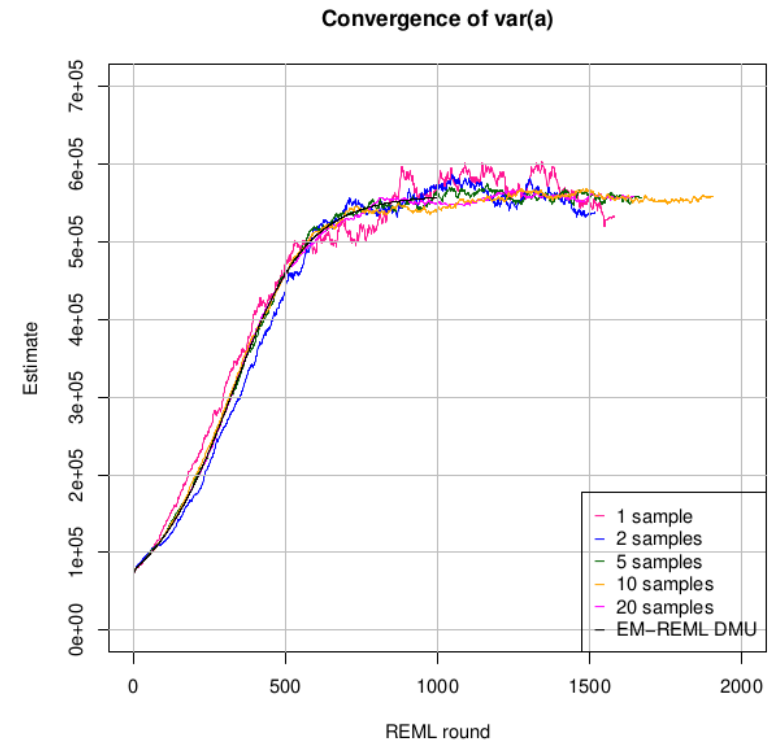
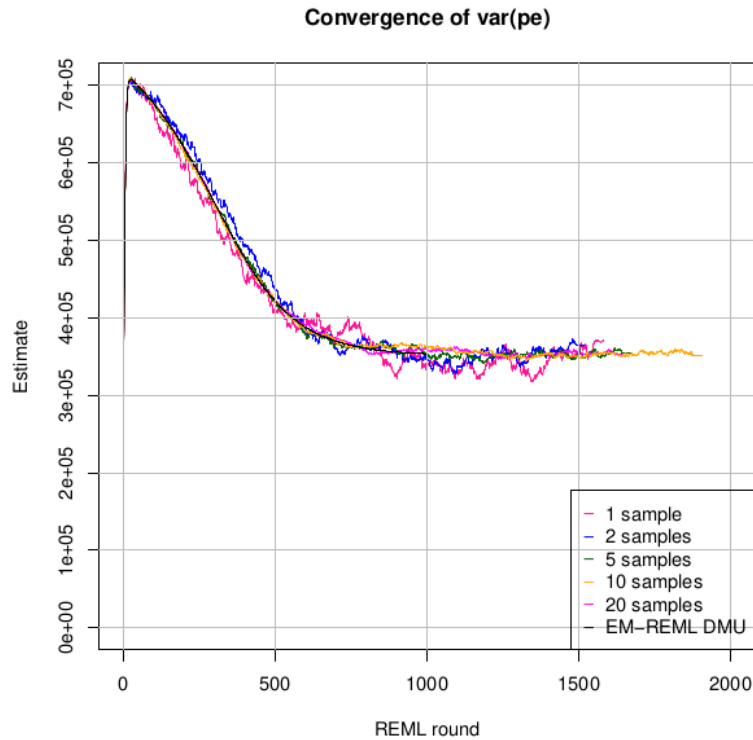
Case study I

- Single trait repeatability model
- Large data: 18,235 cows with repeated 305-day milk yield observations, 44,508 animals in the pedigree
- 41,795 records
- Parameters to be estimated: $\text{var}(pe)$, $\text{var}(a)$, $\text{var}(e)$
- MC EM REML (MiX99)
 - STOP option: 5000 1.0e-5 d f
 - STOPE option: 5000 2 1.0e-10 5.0e-5
 - Number of samples: 1, 2, 5, 10, 20
- Reference analysis: EM REML (DMU)



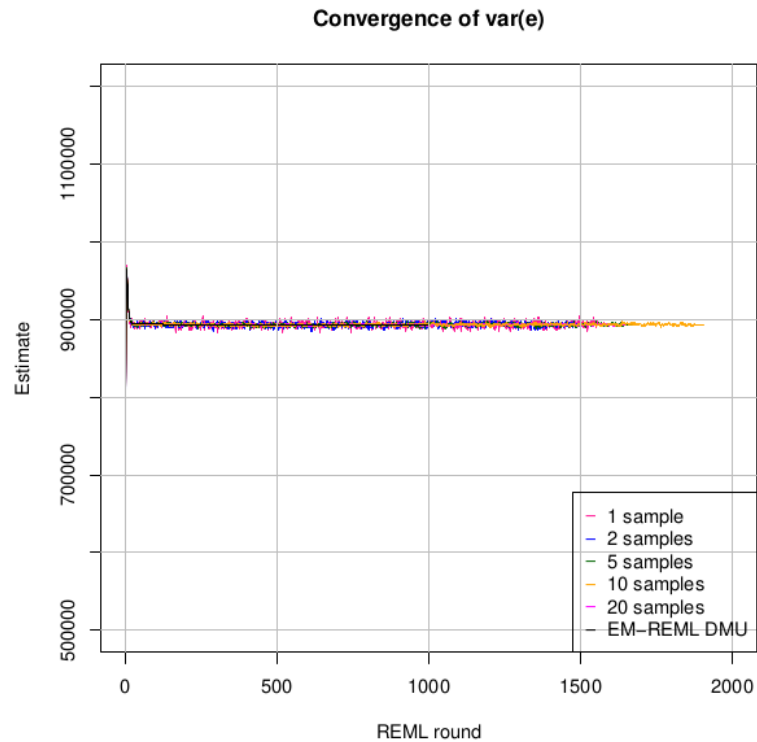
Assessing convergence of VC parameter estimates

Trace plots Case study I



Assessing convergence of VC parameter estimates

Trace plots **Case study I**



Runtime statistics for number of Monte Carlo samples

Samples	REML rounds	Time		h2
		REML round	Total	
1	1583	1.1 s	28.5 min	0.296
2	1519	1.9 s	47.3 min	0.299
5	1672	4.2 s	118.0 min	0.309
10	1907	20.0 s	381.1 min	0.309
20	1646	17.4 s	477.7 min	0.309
DMU	-	-	-	0.309

Assessing convergence of VC parameter estimates

Trace plots Case study I

mix99s standard output

```

116  0.1219E-05  0.4521E-07  0.4534E-06  0.1347E-04  0.1322
117  0.1047E-05  0.4032E-07  0.3895E-06  0.1173E-04  0.1153
118  0.9232E-06  0.3379E-07  0.3388E-06  0.1010E-04  0.1032
119  0.7926E-06  0.2827E-07  0.2977E-06  0.9033E-05  0.9369E-01
120  0.7064E-06  0.2549E-07  0.2617E-06  0.7995E-05  0.8427E-01
Stopping criterion CD < 0.1E-4 achieved in 120 iterations.
NOTE: CD criterion must be met by two consecutive iterations!

Solutions have converged according to CD criterion of the last iteration.

-----
MiX99_SOLVE: End of PCG Iteration                               Time: 12:20:45.4  08.03.2026
-----

114  0.9218E-05  0.6136E-05  0.6447E-05  0.3810E-04  0.1553E-03
111  0.1508E-04  0.2033E-04  0.1104E-04  0.4218E-04  0.2392E-03
110  0.1545E-04  0.5873E-05  0.1057E-04  0.4165E-04  -0.4336E-03
111  0.1393E-04  0.2744E-04  0.1018E-04  0.3757E-04  -0.2159E-03
111  0.1477E-04  0.1197E-04  0.1108E-04  0.3907E-04  0.2694E-03

REML ROUND      1 CONV Cd not yet computed

```

after 1. REML round,

after last REML round

```

16   0.1407E-06  0.1590E-07  0.4987E-07  0.1020E-05  -0.8619E-02
17   0.1394E-06  0.9030E-08  0.4453E-07  0.9143E-06  -0.7968E-02
18   0.1225E-06  0.1068E-07  0.4076E-07  0.8989E-06  -0.8232E-02
19   0.1047E-06  0.9310E-08  0.3407E-07  0.7328E-06  -0.6828E-02
20   0.8859E-07  0.6351E-08  0.2980E-07  0.6758E-06  -0.6360E-02
Stopping criterion CD < 0.1E-4 achieved in 20 iterations.
NOTE: CD criterion must be met by two consecutive iterations!

Solutions have converged according to CD criterion of the last iteration.

-----
MiX99_SOLVE: End of PCG Iteration                               Time: 14:18:41.4  08.03.2026
-----

103  0.2036E-04  0.8468E-05  0.1255E-04  0.4222E-04  -0.1850
103  0.2109E-04  0.1427E-04  0.1400E-04  0.3687E-04  0.1692
102  0.2081E-04  0.5483E-05  0.1201E-04  0.3919E-04  0.1849
104  0.2353E-04  0.9285E-05  0.1436E-04  0.4444E-04  0.2125
92   0.2622E-04  0.3811E-04  0.1807E-04  0.3689E-04  0.2219

REML ROUND 1671 CONV Cd 0.5959E-10(1641+30/30 rounds)

-----
MiX99_SOLVE: End of MC EM REML iteration                       Time: 14:18:45.2  08.03.2026
-----

REML convergence criterion (0.1E-9) was achieved in 1641 rounds
and additional 30 rounds were performed to reduce variation in variance
component estimates.

First 10 MC EM REML estimates:
-----
1   1   1  354973.79
2   1   1  557124.83
1   1   1  893319.89

```

Updating variance components in the 2nd last REML round had very little effect on real data solutions (using old solutions, MiX99 does at least 20 PCG iterations)



Assessing convergence of VC parameter estimates

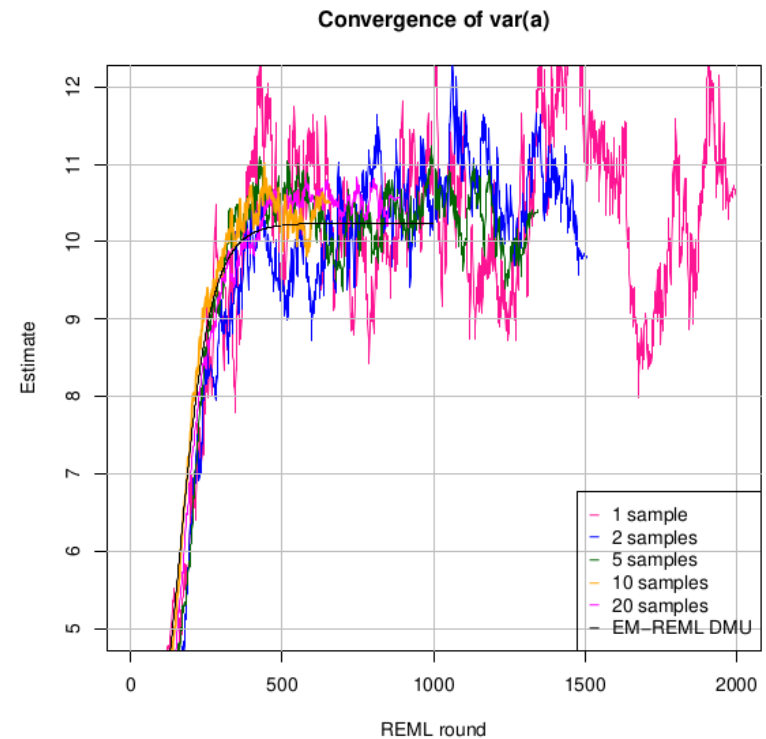
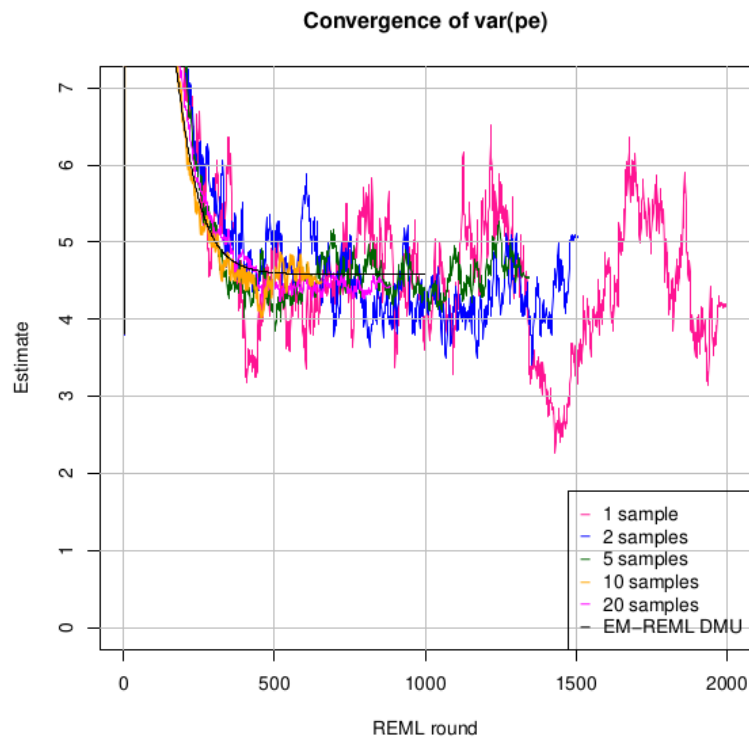
Case study II

- Single trait repeatability model
- Small data: 607 cows with repeated test-day milk yield observations, 2,751 animals in the pedigree
- 22,301 records from 1 herd (Luke research farm)
- Parameters to be estimated: $\text{var}(pe)$, $\text{var}(a)$, $\text{var}(\epsilon)$
- MC EM REML (MiX99)
 - STOP option: 5000 1.0e-5 d f
 - STOPE option: 5000 2 1.0e-11 5.0e-5
 - Number of samples: 1, 2, 5, 10, 20
- Reference analysis: EM REML (DMU)



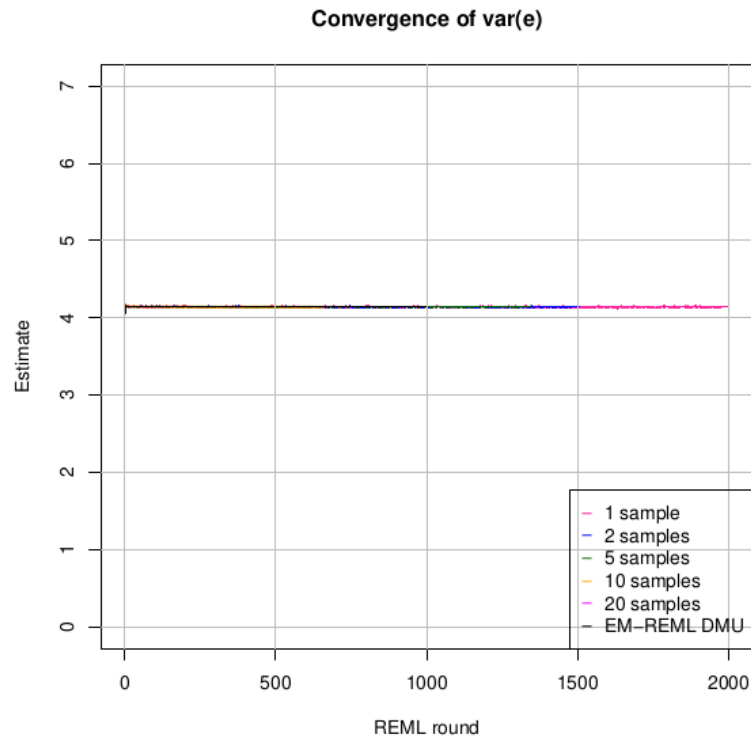
Assessing convergence of VC parameter estimates

Trace plots Case study II



Assessing convergence of VC parameter estimates

Trace plots **Case study II**



Runtime statistics for number of Monte Carlo samples

Samples	REML rounds	Time		h2
		REML round	Total	
1	2002	0.96 s	31.9 min	0.559
2	1508	1.51 s	38.0 min	0.516
5	1346	3.14 s	70.5 min	0.545
10	656	5.70 s	62.4 min	0.549
20	884	11.99 s	176.7 min	0.552
DMU	-	-	-	0.540

Standard errors and fixing VC parameter estimates

Standard errors

- Estimation of SE requires the information matrix

$$SE(\hat{\theta}_i) = \sqrt{\mathcal{J}_{ii}}$$
- With MC EM REML, the information matrix can be approximated
- Large number of MS samples needed
- MS sample number must be larger than number of VC parameters
- Recommended: $>N$ parameters + 50

Calculating standard errors

- Restarting *mix99s* with a modified solver option file after REML analysis has finished
- `$> mix99s < solver_options_SE.txt`

```
#-----
# RAM: RAM options
# H
# STOP: Max.iter,PCG-Conv.valREAL-DATA,Criterion,Enforce
# 5000 1.0e-5 d f
# RESID: Residuals calculation
# N
# VALID: Model validation
# N
# VAROPT: Variance options VCE, PEV, HV
# E
# STOPE: MaxREMLrnd,Samples,VC-Conv.val,[PCG-Conv.valSAMPLE-DATA]
# 1 50 1.0e-10 3.2e-5
# SEED: Type of seed for the random number generator
# R
# MIX99PATH:
# /data/projects/animalgenetics/software/MiX99/bin_beta
# SOLTYP: Solution file options
# Y
#-----
```

1 REML round

50 samples

Standard error output files

File with standard errors: **vceSE**

```
vi-int1:~/MiX99course2026/L11mt1> cat vceSE
1 1 1 1.52142
2 1 1 1.88193
3 1 1 0.354153E-01
```

File with information matrix: **vceI**

```
vi-int1:~/MiX99course2026/L11mt1> cat vceI
1 1 1 1 1 1 2.31471
2 1 1 1 1 1 -2.57801
2 1 1 2 1 1 3.54166
3 1 1 1 1 1 0.459747E-02
3 1 1 2 1 1 -0.498668E-02
3 1 1 3 1 1 0.125425E-02
```

 coordinates of information matrix parameter J_{ij}

No need of restarting *mix99s* with newest MiX99 versions

Number of samples for SE is given on **STOPE** option line

```
# STOPE: MaxREMLrnd,Samples,VC-Conv.val,[PCG-Conv.valSAMPLE-DATA,SEnSamples]
          3000      2      1.0e-10      3.2e-5      50
# SEED: Type of seed for the random number generator
```

Luke

mix99s standard output

```
REML convergence criterion (0.1E-10) was achieved in 854 rounds
and additional 30 rounds were performed to reduce variation in variance
component estimates.
```

First 10 MC EM REML estimates:

```
-----
1 1 1 4.4316960
2 1 1 10.573479
1 1 1 4.1402820
```

```
Number of (co)variance components :      3
Number of Monte Carlo samples      :      20
```

First 10 approximated standard errors for MC EM REML estimates:

```
-----
1 1 1 2.31438
2 1 1 2.65928
3 1 1 0.501035E-01
```

First 10 approximated covariances for MC EM REML estimates:

```
-----
1 1 1 1 1 1 5.35635
2 1 1 1 1 1 -5.79630
2 1 1 2 1 1 7.07176
3 1 1 1 1 1 -0.381792E-01
3 1 1 2 1 1 0.489995E-01
3 1 1 3 1 1 0.251036E-02
```

```
*****
NOTE: The number of MC samples may be too small for the
approximation of standard errors to work.
At least 50 MC samples more than the number of
(co)variance components is recommended.
*****
```

Fixing variance components

- MiX99 allows to fix any number of variance components to the values provided at the start of the REML analysis
- The number of fixed variance components is given on the **VAROPT** line, followed by same number of lines with the parfile integer values (random number, and i j values) of the parameters to be fixed

Set residual covariance to zero

```
vi-int1:~/MiX99course2026/L13vc> cat startingVC
1 1 1 5.6161964
1 2 1 0.19968117
1 2 2 0.73903890E-02
2 1 1 9.8365002
2 2 1 0.18377238
2 2 2 0.48884253E-02
3 1 1 4.1547128
3 2 1 0.0
3 2 2 0.62144568E-02
```

Solver options file

```
N
# VAROPT: Variance options VCE, PEV, HV
E f 1
3 2 1
# STOPE: REMLrounds, Samples, Conv.value VCE,
1000 3 1.0e-10
```

Set a variance parameter to a certain value

```
vi-int1:~/MiX99course2026/L13vc> cat startingVC
1 1 1 1.0
1 2 1 0.0
1 2 2 1.0
2 1 1 1.0
2 2 1 0.0
2 2 2 1.0
3 1 1 0.01
3 2 1 0.0
3 2 2 1.0
```

Solver options file

```
N
# VAROPT: Variance options VCE, PEV, HV
E f 1
3 1 1
# STOPE: REMLrounds, Samples, Conv.value VCE,
1000 3 1.0e-10
```



Thank you!

MiX99 course on genomic prediction, Tuusula, 9–11 March 2026