

MiX99 software suite

MiX99 course on genomic prediction

COURSE DAY 1, March 9th, 2026



Contents

- Overview of MiX99 and supporting programs
- MiX99 manuals
 - CLIM
 - Pre-processor
 - Solver
- New developments in latest release
- MiX99 user community group
- CLIM example for pedigree model
 - Data and pedigree for MiX99
 - Variances
 - Random regression
- CLIM examples for some new features
 - Specify variance components
 - REGMATRIX
- Macros for large models



MiX99 overview



Motivation – theory

- Models for genetic evaluation (linear mixed model):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

- \mathbf{y} is vector of phenotypes, \mathbf{b} is a vector of fixed effects
- \mathbf{u} is a vector of random genetic effects (animal, sire), \mathbf{e} is a vector of residuals
- \mathbf{X} and \mathbf{Z} are design matrices
- Assumptions

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{V}_u \otimes \mathbf{A})$$

- \mathbf{A} is pedigree relationship matrix, can be genomic \mathbf{G} or combined \mathbf{H}
- \mathbf{V}_u is the genetic covariance matrix

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{R} = \mathbf{V}_e \otimes \mathbf{I})$$

- \mathbf{V}_e is the residual covariance matrix



Mixed Model Equations

- The associated mixed model equations (MME) are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{V}_u \otimes \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

- Can be simplified into

$$\mathbf{C}\mathbf{a} = \mathbf{r}$$

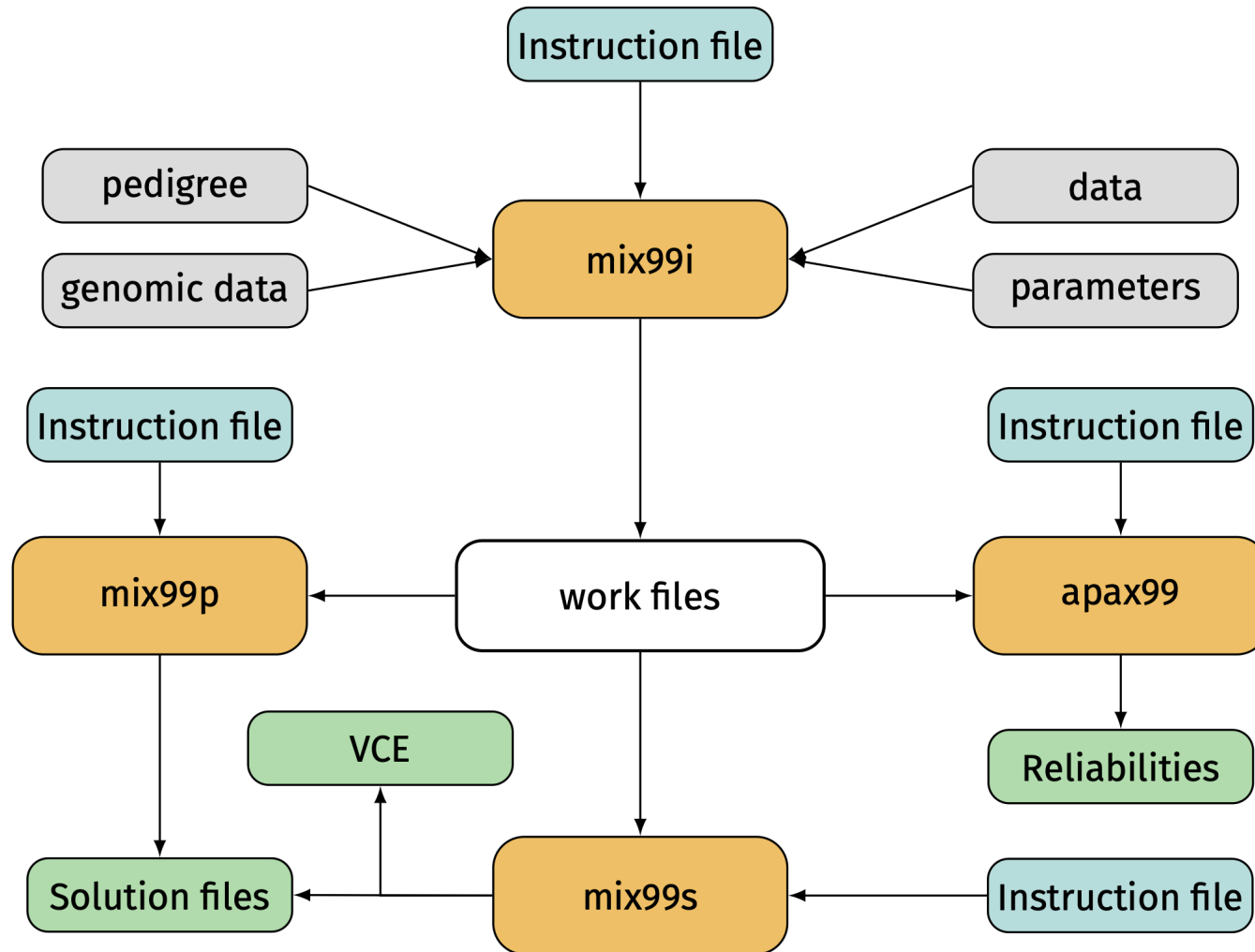
- C** is coefficient matrix (LHS), **r** is right hand side
- Solving the MME is necessary for estimation of breeding values
- For large models (as in genetic evaluations) **C** cannot be inverted directly
 - Cholesky decomposition is computationally expensive
- Instead solve iteratively
 - A few solvers (Jacobi, PCG)

MiX99 strategy for solving MME

- Iteration on data
 - Do not setup MME explicitly
- Preconditioned conjugate gradient algorithm
 - One of the best solving algorithms
 - No restrictions on the type of models we use in animal breeding
- Block ordering of data and equations
 - Given the MME a structure that favours fast solving
- Parallel computing
 - Utilizing many CPUs (cores, threads) at the same time
 - For distributed or shared memory hardware architecture



MiX99



Supporting programs

- Not all part of MiX99 but useful for analyses:
 - Sort and prune pedigree
 - Get reliabilities
 - GEBV
 - Standard errors (PEV)
- Make components for single-step
 - G inverse
 - Gamma matrix (metafounders)

mix99hv

exa99

RelaX2

snp_blup_rel

hginv

T48eig_make

Bpop

predict_GEBV

MiX99 manuals

CLIM, pre-processor, solver



Updated manuals

- Can be found at:

<https://www.luke.fi/en/services/mix99-solving-large-mixed-model-equations>

- Latest version (September 2025)
- Webpage has also additional resources
 - [Presentations](#)
 - [Examples](#)



On this page

Documentation

[Presentations](#)

[Examples](#)

[Gene resources](#)

[Livestock production](#)

[Agriculture](#)

MiX99 is a software suite for breeding value estimation of large-scale genetic and genomic evaluations. MiX99 can be used with wide variety of prediction models and data sets. The software is used worldwide in national and international evaluations for cattle but also for pigs, horses, sheep, goats, fish, foxes, poultry and barley.



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

CLIM manual

- **Command Language Interface for MiX99**
 - Simplified way to specify models (compared to directive file)
- Theory and notation
- Input files
- Describe models
- How to use CLIM for different models
 - Single and multiple trait
 - Genomic
- Options
 - MACROS
 - Special topics



Technical description for `mix99i` (pre-processor)

- How to run pre-processor
- How to edit input files
- MiX99 instruction file
- Options
 - Sequential
 - Parallel
 - Use of MACROS
- Output files
- Examples



Technical description for MiX99 solver (mix99s, mix99p)

- Description of computing methods used to solve MME
- How to run the solver
- Options
 - Command line options
 - Convergence
- Output files
- Reliabilities and related information
 - Daughter yield deviations
- Variance component estimation



New features



What's new?

- Re-written CLIM manual (better explanations)
 - [Single-step models](#)
- `mix99i` has more output
 - [Number of observations per effect](#)
- Maximum length for variable names increased
 - $8 \rightarrow 80$
- Alternative ways to specify variance components
 - [PARFILE options](#)
- Lifted limit for number of traits
- `mix99s` options
 - [Parent average and mendelian sampling](#)
- Improved pre-conditioner for ssSNPBLUP
- `mix99p`
 - [Fully component-wise ss-GTBLUP, ss-SNPBLUP](#)
- Variance component estimation
 - [GBLUP and REGMATRIX](#)
- COVFILE
 - [Lower-triangle dense format](#)
- Options for REGMATRIX
 - [REGINDEX and REGTRAITS](#)
- Single-trait reverse reliability approximation
 - [Parallelized within trait](#)
- Single-step SNPBLUP
 - [Marker-specific weights allowed](#)



New and development features

- **New features**
 - Available in official MiX99 release
- **Development features**
 - Not yet in official MiX99 release

A **MIXED** format variance components file using **SPARSE** and **LOWER** is

SPARSE

```
1 SPARSE
  1 1 3.0
  2 1 2.5
  2 2 2.5
2 SPARSE
  1 1 7.0
  2 1 2.0
  2 2 7.0
```

LOWER

```
1 LOWER
  3.0
  2.5 2.5
2 LOWER
  7.0
  2.0 7.0
```

Assuming the variance matrices were diagonal, the **MIXED** variance components file is

NEW

10.5.2 Marker-specific variances or weights in ssSNPBLUP

Marker-specific variances in the ssSNPBLUP model can be specified using command **SNPPARFILE**. The file format for the variances has the same structure as for heterogeneous variances in the **REGMATRIX** command.

In multi-trait models, this approach allows providing marker-specific and trait-specific (co)variances. For example, consider a two-trait model, extending the previous ssSNPBLUP model by including an additional trait. The CLIM commands can include marker-specific variance matrices:

DEV

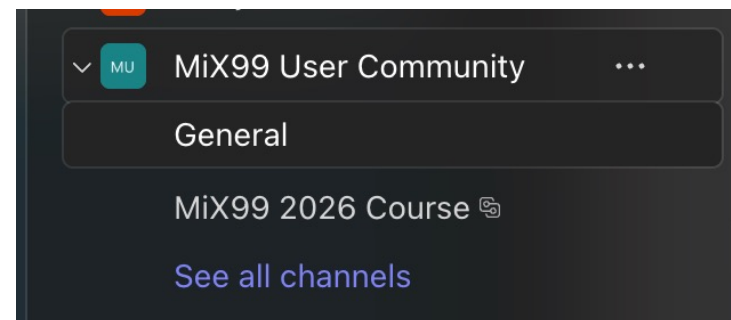


MiX99 user community group



Teams group

- You should have received an email inviting you to the group
 - [Need strong authentication](#)
- Intended for discussing issues related to using MiX99
 - [Error messages](#)
 - [Bugs](#)
 - [Convergence and implementation issues](#)
- Additional resources
 - [Resources related to 2026 Course](#)
- More “channels” will be added



Brief introduction to CLIM



CLIM philosophy

- Allows to specify the minimum commands needed
- Default values are used if no specific commands are given
- Specific commands to optimize performance of MiX99
- Specific commands for advanced models
- **Required commands**
 - Input file names
 - Variables and parameters in input files
 - Traits and applied model(s)
- **Some optional commands**
 - Data structure / ordering of data
 - Ordering of equations
 - Preconditioning
 - Trait groups
 - Correlation matrices
 - Matrices for genomic prediction
 - Parallel computing

CLIM syntax

Required commands

```

DATAFILE Data.dat # Name of data file
INTEGER ID mean # Integer variables
REAL x y # Real variables (x = regression variable, y = phenotype)
PEDFILE Ped.dat # Name of pedigree file
PEDIGREE ID am # Genetic effects associated with ID, am = animal model
PARFILE vc.dat # Name of the file with variance components
MODEL
    y = x mean ID # Model line
  
```

- No headers in data files
- Integer variables first then reals

In the model line:

1. Fixed effects
 - Regression
 - Class
2. Random effects
 - Non-genetic first
 - Genetic



CLIM syntax

- Special symbols
 - # start for comment
 - & symbol for line continuation
 - " " string in between the apostrophes is read unchanged
 - ! options follow (on the model line)
 - () parenthesis used on model line(s)
- **Important!** CLIM commands and options need to be specified as in the manual (in CAPITAL letters)



CLIM syntax

Optional commands

```

DATASORT    BLOCK=<block in INTEGER> PEDIGREECODE=<code in INTEGER>
MISSING     <value for missing real number data>
NORANSOL    <random effect numbers without solution file>
PARFILE     <filename>
RANDOM       <random effect names>
REGFILE     <filename>
REGPARFILE  <filename>
REGMATRIX   <type> <name> [ID=<column>] FIRST=<column> [LAST=<column>]
             [REGINDEX=<INTEGER column name>] [IMPUTE=<missing>]
             [SCALE={<real value>|<filename>|2pq|m|m2}] [FORMAT=<file format>]
             [CENTER[={<real value>|<filename>}]] [PRECON=<preconditioner>]
RESIDFILE   <filename>
RESIDUAL    <INTEGER column name of the residual variance number>
SSGBLUP     [LOWER|MIXED] <filename>
TABLEFILE   <filename>
TABLEINDEX  <table index INTEGER column name>
TITLE       <title of analysis>
TMPDIR      <directory>
TRAITGROUP  <trait group INTEGER column name>
WITHINBLOCKORDER <Effect names in the order>
DEFINE      <macro name> <replacement text>

```



CLIM examples



Simple animal repeatability model

- Model

$$y = herd \cdot year + c + a + e$$

- c = random permanent environmental effect
 - a = random additive genetic effect
- Two effects related to the same subject (animal ID) but which is c , which is a ?

```

INTEGER ID mean herd_year
REAL    Y
PEDIGREE ID am
MODEL
  y = mean herd_year ID ID
  
```

← Incorrect!

- Use of component names is required

Simple animal repeatability model

- Model

$$y = herd \cdot year + c + a + e$$

- c = random permanent environmental effect
- a = random additive genetic effect
- Two effects related to the same subject (animal ID) but which is c , which is a ?
- Define **G** as a component name
- Add a `RANDOM` command to specify permanent environmental effect

```
INTEGER ID mean herd_year
REAL y
PEDIGREE G am
RANDOM PE
MODEL
  y = mean herd_year PE(ID) G(ID)
```

Random regression

- Model

$$SCS_{ijklm} = A_i + \delta_H F_{Hm} + \delta_R F_{Rm} + \delta_I F_{Im} + \sum_{p=0}^4 \phi_{pt} \beta_p + \eta_l + \sum_{p=0}^3 \phi_{pt} c_{pm} + \sum_{p=0}^3 \phi_{pt} u_{pk} + e_{ijklmp}$$

```

INTEGER      ID sire dim cymc htm hy_scs age
REAL         scs het recl inbr
RANDOM        htm PE G
TABLEFILE    Leg.cov
TABLEINDEX   dim
MODEL
scs = het recl inbr hy_scs age LAcurve(t1 t2 t3 t4 t5 |cymc) htm PE(t1 t2 t3 t5 |ID) G(t1 t2 t3 t5 |sire)

```

Leg.cov

8	1.00000	-1.22474	1.58114	-1.87083	0.72615
9	1.00000	-1.21788	1.55464	-1.80838	0.69768
10	1.00000	-1.21102	1.52829	-1.74681	0.67032
11	1.00000	-1.20416	1.50209	-1.68611	0.64404
12	1.00000	-1.19730	1.47603	-1.62628	0.61878
13	1.00000	-1.19044	1.45013	-1.56731	0.59452
14	1.00000	-1.18358	1.42438	-1.50919	0.57121
15	1.00000	-1.17672	1.39877	-1.45193	0.54881
16	1.00000	-1.16985	1.37331	-1.39551	0.52729
17	1.00000	-1.16299	1.34800	-1.33993	0.50662

Long model line,
we can simplify
it

Use of MACROS and range expansion

- Some model parts are replaced by a short name
- Less errors because same macro used in different traits
- Large models look simpler
 - Error messages may be difficult to interpret
- Not a name already used (by CLIM)
 - PEDIGREE, MODEL, ..., etc
 - Component name
- Has to be defined before being used
- Range expansions can help further decrease model size

- Usage

```
DEFINE <name> <what this name means>
```

- In command line:

```
mix99i --usemacros rmodel_wmacros.clm
```

- `rmodel_wmacros.clm` is CLIM file with macros defined



Random regression model with MACROS

- Define MACROS for fixed, permanent environmental and genetic effects
- Get a simplified model

```

INTEGER    ID sire dim cymc htm hy_scs age
REAL       scs het recl inbr
RANDOM      htm PE G
TABLEFILE  Leg.cov
TABLEINDEX dim
MODEL
scs = het recl inbr hy_scs age LAcurve(t1 t2 t3 t4 t5 |
DEFINE     F_curve LAcurve(t1 t2 t3 t4 t5 | cymc,
DEFINE     P_eff   PE(t1 t2 t3 t5 | ID)
DEFINE     G_eff   G(t1 t2 t3 t5 | sire)
MODEL
scs = F_eff F_curve htm P_eff G_eff

```

```

INTEGER    ID sire dim cymc htm hy_scs age
REAL       scs het recl inbr
RANDOM      htm PE G
TABLEFILE  Leg.cov
TABLEINDEX dim
DEFINE     F_eff   het recl inbr hy_scs age sire)
DEFINE     F_curve LAcurve(t1:5 | cymc)
DEFINE     P_eff   PE(t1:3 t5 | ID)
DEFINE     G_eff   G(t1:3 t5 | sire)
MODEL
scs = F_eff F_curve htm P_eff G_eff

```



Examples of some new features



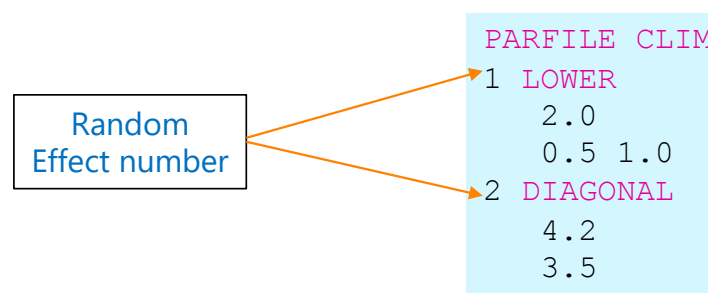
Variance components

- Two trait model with two effects (genetic and residual)
- Parameter file

```
1 1 1 2.0
1 2 1 0.5
1 2 2 1.0
2 1 1 4.2
2 2 1 0.0
2 2 2 3.5
```

Random Effect number	Row number	Column number	Covariance parameter
----------------------	------------	---------------	----------------------

- Alternatively in new MiX99:



- Special option 1: PARFILE IDENTITY
 - All variance matrices set to identity matrices
- Special option 2: PARFILE MIXED <parfile>
 - parfile has the LOWER/DIAGONAL ... commands

Animal model with genetic groups – 2 ways

- Simple model
 - Mean and animal effect
 - Genetic groups (random)
- CLIM file

```
DATAFILE QuaasData.dat
INTEGER ID mean
REAL y
PEDFILE QuaasPed.dat
PEDIGREE G am+p 0.5
PARFILE vc.dat
MODEL
  y = mean G(ID)
```

QuaasPed.dat

1	-1	-1
2	-2	-2
3	1	2
4	-2	-2
5	-1	-1
6	-2	-2
7	3	4
8	5	6
9	7	8

vc.dat

1	1	1	20.0
2	1	1	40.0

QuaasData.dat

3	1	5.50
7	1	6.50
8	1	4.75
9	1	2.25

Animal model with genetic groups – 2 ways

- Alternative way:
 - Pass covariates using REGMATRIX

```

DATAFILE  QuaasData.dat
INTEGER  ID mean
REAL     y
DATASORT PEDIGREECODE = ID
PEDFILE  QuaasPedNoGG.dat
PEDIGREE G am
PARFILE  vc.dat
→ REGMATRIX RANDOM GGCOV ID = 1 REGINDEX = ID FIRST = 2 LAST = 3
→ REGFILE  QuaasCov.dat
→ REGPARFILE Covpar.dat
MODEL
  y = mean G(ID)

```

QuaasPedNoGG.dat

1	0	0
2	0	0
3	1	2
4	0	0
5	0	0
6	0	0
7	3	4
8	5	6
9	7	8

QuaasCov.dat

1	1.000	0.000
2	0.000	1.000
3	0.500	0.500
4	0.000	1.000
5	1.000	0.000
6	0.000	1.000
7	0.250	0.750
8	0.500	0.500
9	0.375	0.625

vc.dat

1	1	1	20.0
2	1	1	40.0

Covpar.dat

1	1	1	40.0
---	---	---	------

← 40=20.0×(1/0.5)

Summary

- Introduced MiX99 suite of programs
 - Strategy for solving equations
 - Genetic and genomic evaluation
 - Accompanying programs
- Updated manuals
 - New features
- Brief introduction to CLIM
 - Basic models
 - Component names
 - Random regression – table file
 - MACROS and range expansions
- New features
 - PARFILE
 - REGMATRIX



Thank you!



luke.fi



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

© Luonnonvarakeskus
© Natural Resources Institute Finland
© Naturrekursinstitutet