Users’ manuals and useful options

MiX99 course: test-day models and single step genomic prediction
COURSE DAY 2, April 11th, 2025

Matti Taskinen
Natural Resources Institute Finland (Luke)

April 11, 2025

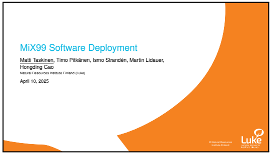
NATURAL RESOURCES INSTITUTE FINLAND

Content

- MiX99 documents
- Some useful MiX99 features

MiX99 Documentation

MiX99 distribution comes with 4 documentation files:



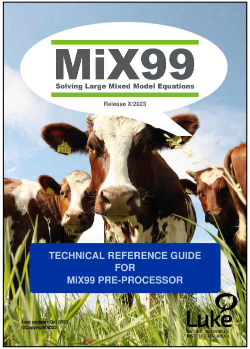
MiX99 Software Deployment

Matti Taskinen, Timo Pitkänen, Ismo Strandén, Martin Lidauer, Hongding Gao
Natural Resources Institute Finland (Luke)

April 10, 2025

© Natural Resources Institute Finland

MiX99 Sofware Deployment
`MiX99_deployment.pdf`

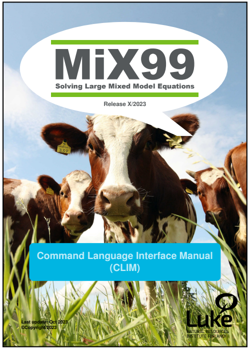


MiX99
Solving Large Mixed Model Equations
Release X/2023

TECHNICAL REFERENCE GUIDE
FOR
MiX99 PRE-PROCESSOR

Last update: Oct 2023
©Copyright 2023

Technical reference guide for MiX99 pre-processor
`refguide_preproc.pdf`

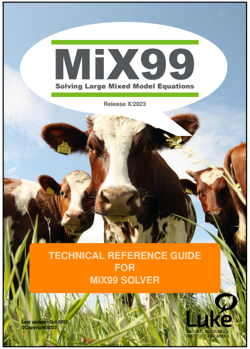


MiX99
Solving Large Mixed Model Equations
Release X/2023

Command Language Interface Manual
(CLIM)

Last update: Oct 2023
©Copyright 2023

Command Language Interface Manual
`refguide_clim.pdf`



MiX99
Solving Large Mixed Model Equations
Release X/2023

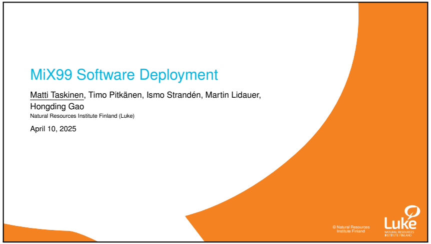
TECHNICAL REFERENCE GUIDE
FOR
MiX99 SOLVER

Last update: Oct 2023
©Copyright 2023

Technical reference guide for MiX99 solver
`refguide_solver.pdf`

MiX99 Software Deployment

- What is included in MiX99 **distribution**
- How to **install** MiX99
- How to **start using** MiX99
- Introduction to **documentation**



MiX99 Software Deployment

Matti Taskinen, Timo Pitkänen, Ismo Strandén, Martin Lidauer,
Hongding Gao

Natural Resources Institute Finland (Luke)

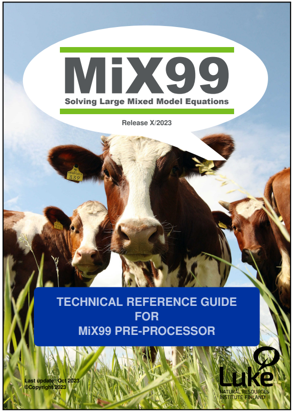
April 10, 2025

© Natural Resources Institute Finland

MiX99 Sofware Deployment
`MiX99_deployment.pdf`

Technical reference guide for MiX99 pre-processor

- How to use MiX99 **preprocessor**.
- What are **input files**.
- Describes MiX99 **instruction syntax**.
- Simple **examples**.



MiX99

Solving Large Mixed Model Equations

Release X/2023

TECHNICAL REFERENCE GUIDE
FOR
MiX99 PRE-PROCESSOR

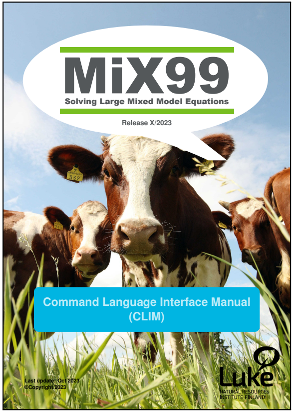
Last update: Oct 2023
©Copyright 2023

Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Technical reference guide for MiX99 pre-processor
`refguide_preproc.pdf`

Command Language Interface Manual (CLIM)

- Alternative, **more user-friendly** instruction syntax for preprocessor.
- **Theoretical background** information.
- Additional information on:
  - Models.
  - Files.
  - Solver.
- More **examples**.



MiX99
Solving Large Mixed Model Equations
Release X/2023

Command Language Interface Manual (CLIM)

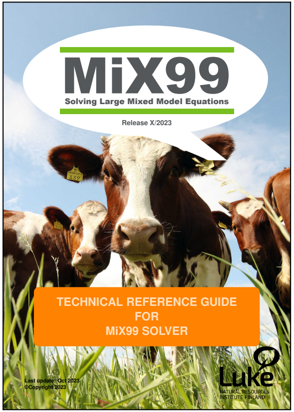
Last update: Oct 2023
©Copyright 2023

Luke
NATURAL RESOURCES INSTITUTE FINLAND

Command Language Interface Manual
`refguide_clim.pdf`

Technical reference guide for MiX99 solver

- How to use MiX99 **solver**.
- What are **output files**.
- Special topics:
    - Reliabilities.
    - Daughter Yield Deviations.
    - Non-linear models.
    - Estimation of variance components.
    - Accounting for heterogeneous variance.



MiX99

Solving Large Mixed Model Equations

Release X/2023

TECHNICAL REFERENCE GUIDE
FOR
MiX99 SOLVER

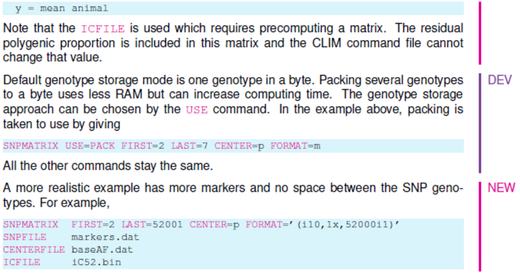
Last update: Oct 2023
©Copyright 2023

Luke
NATURAL RESOURCES INSTITUTE FINLAND

Technical reference guide for MiX99 solver
`refguide_solver.pdf`

Additional features in MiX99 reference guides

- **New** and development **features**:
  - Indicated with colored vertical margin bars.



```
  y = mean animal
```

Note that the `ICFILE` is used which requires precomputing a matrix. The residual polygenic proportion is included in this matrix and the CLIM command file cannot change that value.

Default genotype storage mode is one genotype in a byte. Packing several genotypes to a byte uses less RAM but can increase computing time. The genotype storage approach can be chosen by the `USE` command. In the example above, packing is taken to use by giving

```
SNPMATRIX USE=PACK FIRST=2 LAST=7 CENTER=p FORMAT=m
```

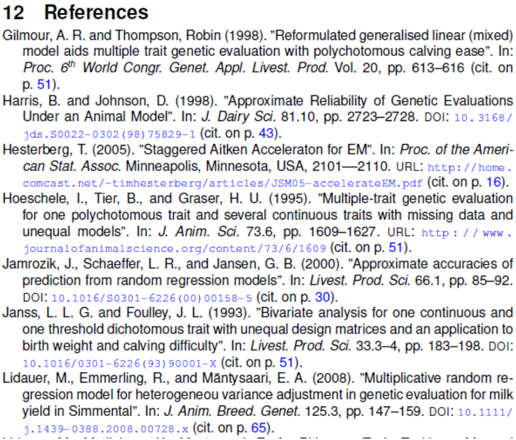
All the other commands stay the same.

A more realistic example has more markers and no space between the SNP genotypes. For example,

```
SNPMATRIX  FIRST=2 LAST=52001 CENTER=p FORMAT='(i10,1x,52000i1)'
SNPFILE    markers.dat
CENTERFILE baseAF.dat
ICFILE     iC52.bin
```

Additional features in MiX99 reference guides

- **New** and development **features**:
  - Indicated with colored vertical margin bars.
- Citations and list of **references**.



12 References

Gilmour, A. R. and Thompson, Robin (1998). ”Reformulated generalised linear (mixed) model aids multiple trait genetic evaluation with polychotomous calving ease”. In: *Proc. 6th World Congr. Genet. Appl. Livest. Prod.* Vol. 20, pp. 613–616 (cit. on p. 51).

Harris, B. and Johnson, D. (1998). ”Approximate Reliability of Genetic Evaluations Under an Animal Model”. In: *J. Dairy Sci.* 81.10, pp. 2723–2728. DOI: `10.3168/jds.S0022-0302(98)75829-1` (cit. on p. 43).

Hesterberg, T. (2005). ”Staggered Aitken Acceleraton for EM”. In: *Proc. of the American Stat. Assoc.* Minneapolis, Minnesota, USA, 2101—2110. URL: `http://home.comcast.net/~timhesterberg/articles/JSM05-accelerateEM.pdf` (cit. on p. 16).

Hoeschele, I., Tier, B., and Graser, H. U. (1995). ”Multiple-trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models”. In: *J. Anim. Sci.* 73.6, pp. 1609–1627. URL: `http://www.journalofanimalscience.org/content/73/6/1609` (cit. on p. 51).

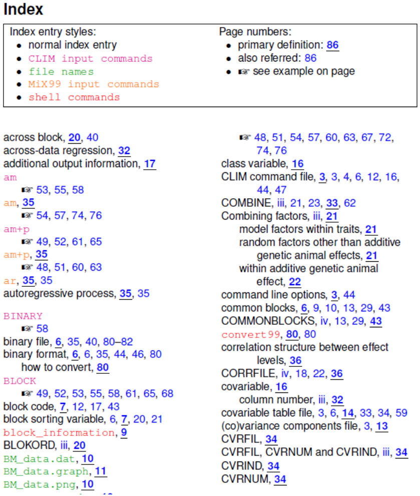
Jamrozik, J., Schaeffer, L. R., and Jansen, G. B. (2000). ”Approximate accuracies of prediction from random regression models”. In: *Livest. Prod. Sci.* 66.1, pp. 85–92. DOI: `10.1016/S0301-6226(00)00158-5` (cit. on p. 30).

Janss, L. L. G. and Foulley, J. L. (1993). ”Bivariate analysis for one continuous and one threshold dichotomous trait with unequal design matrices and an application to birth weight and calving difficulty”. In: *Livest. Prod. Sci.* 33.3–4, pp. 183–198. DOI: `10.1016/0301-6226(93)90001-X` (cit. on p. 51).

Lidauer, M., Emmerling, R., and Mäntysaari, E. A. (2008). ”Multiplicative random regression model for heterogeneou variance adjustment in genetic evaluation for milk yield in Simmental”. In: *J. Anim. Breed. Genet.* 125.3, pp. 147–159. DOI: `10.1111/j.1439-0388.2008.00728.x` (cit. on p. 65).

Additional features in MiX99 reference guides

- **New** and development **features**:
  - Indicated with colored vertical margin bars.
- Citations and list of **references**.
- **Index**:
  - Alphabetical list of covered terms.
  - Locations of primary definitions, secondary references, and examples.
  - Color and typeface coded categories.



Index

Index entry styles:
- normal index entry
- `CLIM input commands`
- `file names`
- `MiX99 input commands`
- `shell commands`

Page numbers:
- primary definition: **86**
- also referred: 86
- ☞ see example on page

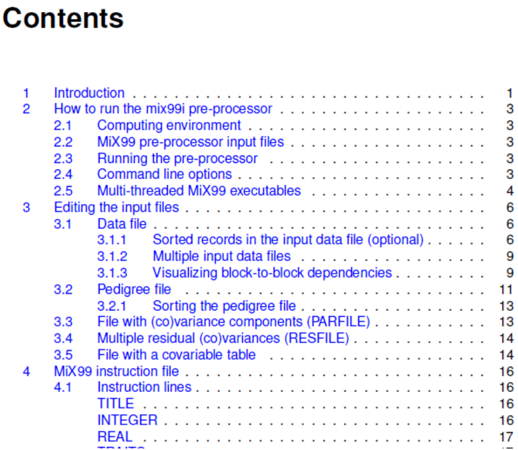
across block, **20**, 40
across-data regression, **32**
additional output information, **17**
`am`
  ☞ 53, 55, 58
`am`, **35**
  ☞ 54, 57, 74, 76
`am+p`
  ☞ 49, 52, 61, 65
`am+p`, **35**
  ☞ 48, 51, 60, 63
`ar`, **35**, 35
autoregressive process, **35**, 35

`BINARY`
  ☞ 58
binary file, **6**, 35, 40, 80–82
binary format, **6**, 6, 35, 44, 46, 80
  how to convert, **80**
`BLOCK`
  ☞ 49, 52, 53, 55, 58, 61, 65, 68
block code, **7**, 12, 17, 43
block sorting variable, 6, **7**, 20, 21
`block_information`, **9**
BLOKORD, iii, **20**
`BM_data.dat`, **10**
`BM_data.graph`, **11**
`BM_data.png`, **10**

  ☞ 48, 51, 54, 57, 60, 63, 67, 72, 74, 76
class variable, **16**
CLIM command file, **3**, 3, 4, 6, 12, 16, 44, 47
COMBINE, iii, 21, 23, **33**, 62
Combining factors, iii, **21**
  model factors within traits, **21**
  random factors other than additive genetic animal effects, **21**
  within additive genetic animal effect, **22**
command line options, **3**, 44
common blocks, **6**, 9, 10, 13, 29, 43
COMMONBLOCKS, iv, 13, 29, **43**
`convert99`, **80**, 80
correlation structure between effect levels, **36**
CORRFILE, iv, 18, 22, **36**
covariable, **16**
  column number, iii, **32**
covariable table file, 3, 6, **14**, 33, 34, 59
(co)variance components file, 3, **13**
CVRFIL, **34**
CVRFIL, CVRNUM and CVRIND, iii, **34**
CVRIND, **34**
CVRNUM, **34**

Hyperlinks in MiX99 reference guides

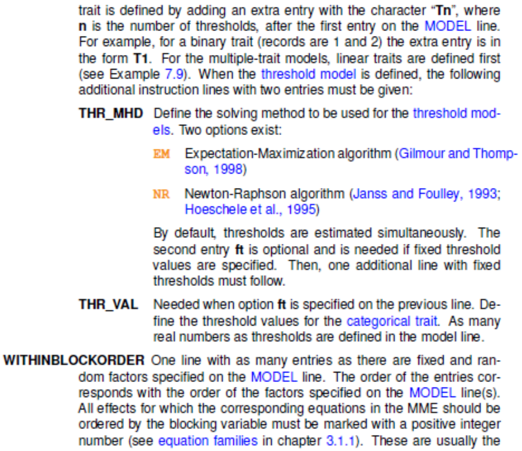
- List of contents: to chapters



Contents

Hyperlinks in MiX99 reference guides

- List of contents: to chapters
- Keywords, commands, files: to primary definitions
- Citations: to list of references



trait is defined by adding an extra entry with the character "**Tn**", where **n** is the number of thresholds, after the first entry on the MODEL line. For example, for a binary trait (records are 1 and 2) the extra entry is in the form **T1**. For the multiple-trait models, linear traits are defined first (see Example 7.9). When the threshold model is defined, the following additional instruction lines with two entries must be given:

THR_MHD Define the solving method to be used for the threshold models. Two options exist:

- **EM** Expectation-Maximization algorithm (Gilmour and Thompson, 1998)
- **NR** Newton-Raphson algorithm (Janss and Foulley, 1993; Hoeschele et al., 1995)

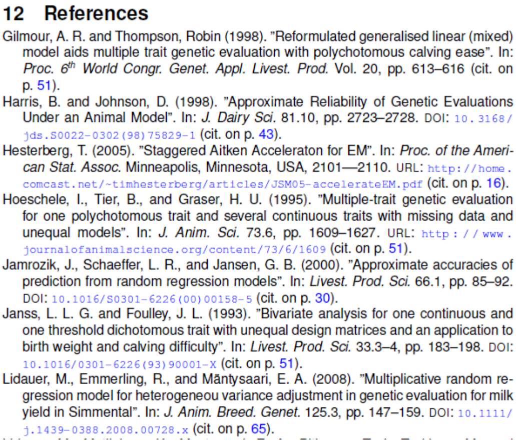
By default, thresholds are estimated simultaneously. The second entry **ft** is optional and is needed if fixed threshold values are specified. Then, one additional line with fixed thresholds must follow.

THR_VAL Needed when option **ft** is specified on the previous line. Define the threshold values for the categorical trait. As many real numbers as thresholds are defined in the model line.

WITHINBLOCKORDER One line with as many entries as there are fixed and random factors specified on the MODEL line. The order of the entries corresponds with the order of the factors specified on the MODEL line(s). All effects for which the corresponding equations in the MME should be ordered by the blocking variable must be marked with a positive integer number (see equation families in chapter 3.1.1). These are usually the

Hyperlinks in MiX99 reference guides

- List of contents: to chapters
- Keywords, commands, files: to primary definitions
- Citations: to list of references
- List of references:
  - page numbers: to citation locations
  - ULRs and DOIs: to Internet



12 References

Gilmour, A. R. and Thompson, Robin (1998). "Reformulated generalised linear (mixed) model aids multiple trait genetic evaluation with polychotomous calving ease". In: *Proc. 6^{th} World Congr. Genet. Appl. Livest. Prod.* Vol. 20, pp. 613–616 (cit. on p. 51).

Harris, B. and Johnson, D. (1998). "Approximate Reliability of Genetic Evaluations Under an Animal Model". In: *J. Dairy Sci.* 81.10, pp. 2723–2728. DOI: 10.3168/jds.S0022-0302(98)75829-1 (cit. on p. 43).

Hesterberg, T. (2005). "Staggered Aitken Acceleraton for EM". In: *Proc. of the American Stat. Assoc.* Minneapolis, Minnesota, USA, 2101—2110. URL: http://home.comcast.net/~timhesterberg/articles/JSM05-accelerateEM.pdf (cit. on p. 16).

Hoeschele, I., Tier, B., and Graser, H. U. (1995). "Multiple-trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models". In: *J. Anim. Sci.* 73.6, pp. 1609–1627. URL: http://www.journalofanimalscience.org/content/73/6/1609 (cit. on p. 51).

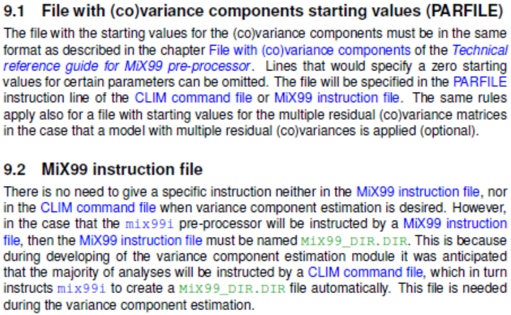
Jamrozik, J., Schaeffer, L. R., and Jansen, G. B. (2000). "Approximate accuracies of prediction from random regression models". In: *Livest. Prod. Sci.* 66.1, pp. 85–92. DOI: 10.1016/S0301-6226(00)00158-5 (cit. on p. 30).

Janss, L. L. G. and Foulley, J. L. (1993). "Bivariate analysis for one continuous and one threshold dichotomous trait with unequal design matrices and an application to birth weight and calving difficulty". In: *Livest. Prod. Sci.* 33.3–4, pp. 183–198. DOI: 10.1016/0301-6226(93)90001-X (cit. on p. 51).

Lidauer, M., Emmerling, R., and Mäntysaari, E. A. (2008). "Multiplicative random regression model for heterogeneou variance adjustment in genetic evaluation for milk yield in Simmental". In: *J. Anim. Breed. Genet.* 125.3, pp. 147–159. DOI: 10.1111/j.1439-0388.2008.00728.x (cit. on p. 65).

Hyperlinks in MiX99 reference guides

- List of contents: to chapters
- Keywords, commands, files: to primary definitions
- Citations: to list of references
- List of references:
  - page numbers: to citation locations
  - ULRs and DOIs: to Internet
- Cross-document: to another MiX99 reference guide



9.1 File with (co)variance components starting values (PARFILE)

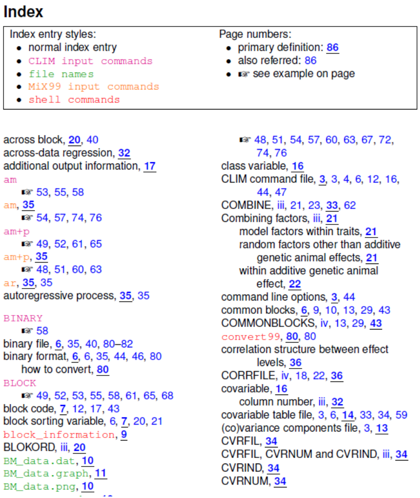
The file with the starting values for the (co)variance components must be in the same format as described in the chapter File with (co)variance components of the *Technical reference guide for MiX99 pre-processor*. Lines that would specify a zero starting values for certain parameters can be omitted. The file will be specified in the PARFILE instruction line of the CLIM command file or MiX99 instruction file. The same rules apply also for a file with starting values for the multiple residual (co)variance matrices in the case that a model with multiple residual (co)variances is applied (optional).

9.2 MiX99 instruction file

There is no need to give a specific instruction neither in the MiX99 instruction file, nor in the CLIM command file when variance component estimation is desired. However, in the case that the `mix99i` pre-processor will be instructed by a MiX99 instruction file, then the MiX99 instruction file must be named `MiX99_DIR.DIR`. This is because during developing of the variance component estimation module it was anticipated that the majority of analyses will be instructed by a CLIM command file, which in turn instructs `mix99i` to create a `MiX99_DIR.DIR` file automatically. This file is needed during the variance component estimation.

Hyperlinks in MiX99 reference guides

- List of contents: to chapters
- Keywords, commands, files: to primary definitions
- Citations: to list of references
- List of references:
  - page numbers: to citation locations
  - ULRs and DOIs: to Internet
- Cross-document: to another MiX99 reference guide
- Index page numbers: to definitions, references, and examples.



Index

Index entry styles:
- normal index entry
- `CLIM input commands`
- `file names`
- `MiX99 input commands`
- `shell commands`

Page numbers:
- primary definition: **86**
- also referred: 86
- ☞ see example on page

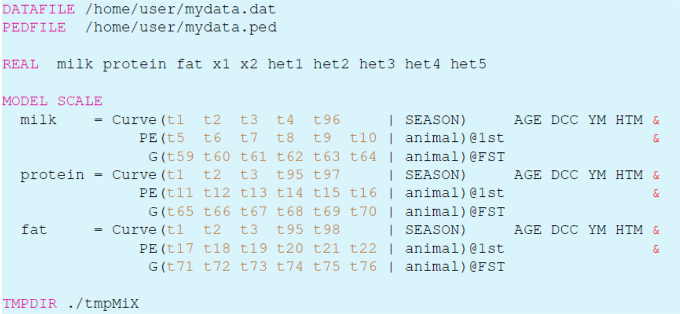
across block, **20**, 40
across-data regression, **32**
additional output information, **17**
`am`
  ☞ 53, 55, 58
`am`, **35**
  ☞ 54, 57, 74, 76
`am+p`
  ☞ 49, 52, 61, 65
`am+p`, **35**
  ☞ 48, 51, 60, 63
`ar`, **35**, 35
autoregressive process, **35**, 35

`BINARY`
  ☞ 58
binary file, **6**, 35, 40, 80–82
binary format, **6**, 6, 35, 44, 46, 80
  how to convert, **80**
`BLOCK`
  ☞ 49, 52, 53, 55, 58, 61, 65, 68
block code, **7**, 12, 17, 43
block sorting variable, 6, **7**, 20, 21
`block_information`, **9**
BLOKORD, iii, **20**
`BM_data.dat`, **10**
`BM_data.graph`, **11**
`BM_data.png`, **10**

  ☞ 48, 51, 54, 57, 60, 63, 67, 72, 74, 76
class variable, **16**
CLIM command file, **3**, 3, 4, 6, 12, 16, 44, 47
COMBINE, iii, 21, 23, **33**, 62
Combining factors, iii, **21**
  model factors within traits, **21**
  random factors other than additive genetic animal effects, **21**
  within additive genetic animal effect, **22**
command line options, **3**, 44
common blocks, **6**, 9, 10, 13, 29, 43
COMMONBLOCKS, iv, 13, 29, **43**
`convert99`, **80**, 80
correlation structure between effect levels, **36**
CORRFILE, iv, 18, 22, **36**
covariable, **16**
  column number, iii, **32**
covariable table file, 3, 6, **14**, 33, 34, 59
(co)variance components file, 3, **13**
CVRFIL, **34**
CVRFIL, CVRNUM and CVRIND, iii, **34**
CVRIND, **34**
CVRNUM, **34**

Usefull MiX99 features: CLIM macros

- Defining **macros** with replacement text.
- **Range expansion** for numbered identifiers.

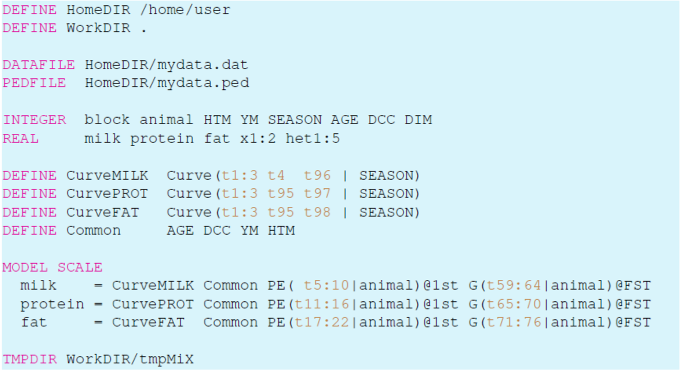


```
DATAFILE /home/user/mydata.dat
PEDFILE  /home/user/mydata.ped

REAL  milk protein fat x1 x2 het1 het2 het3 het4 het5

MODEL SCALE
  milk    = Curve(t1  t2  t3  t4  t96     | SEASON)     AGE DCC YM HTM &
               PE(t5  t6  t7  t8  t9  t10 | animal)@1st                &
                G(t59 t60 t61 t62 t63 t64 | animal)@FST
  protein = Curve(t1  t2  t3  t95 t97     | SEASON)     AGE DCC YM HTM &
               PE(t11 t12 t13 t14 t15 t16 | animal)@1st                &
                G(t65 t66 t67 t68 t69 t70 | animal)@FST
  fat     = Curve(t1  t2  t3  t95 t98     | SEASON)     AGE DCC YM HTM &
               PE(t17 t18 t19 t20 t21 t22 | animal)@1st                &
                G(t71 t72 t73 t74 t75 t76 | animal)@FST

TMPDIR ./tmpMiX
```

Usefull MiX99 features: CLIM macros

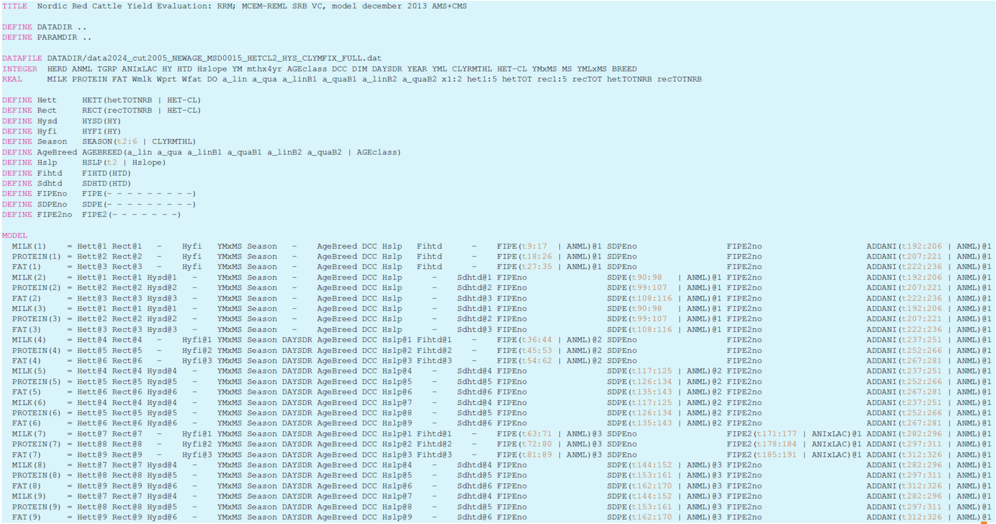
- Defining **macros** with replacement text.

- **Range expansion** for numbered identifiers.



```
DEFINE HomeDIR /home/user
DEFINE WorkDIR .

DATAFILE HomeDIR/mydata.dat
PEDFILE  HomeDIR/mydata.ped

INTEGER  block animal HTM YM SEASON AGE DCC DIM
REAL     milk protein fat x1:2 het1:5

DEFINE CurveMILK  Curve(t1:3 t4  t96 | SEASON)
DEFINE CurvePROT  Curve(t1:3 t95 t97 | SEASON)
DEFINE CurveFAT   Curve(t1:3 t95 t98 | SEASON)
DEFINE Common     AGE DCC YM HTM

MODEL SCALE
  milk    = CurveMILK Common PE( t5:10|animal)@1st G(t59:64|animal)@FST
  protein = CurvePROT Common PE(t11:16|animal)@1st G(t65:70|animal)@FST
  fat     = CurveFAT  Common PE(t17:22|animal)@1st G(t71:76|animal)@FST

TMPDIR WorkDIR/tmpMiX
```

CLIM macros: Timo’s RDC example



```
TITLE  Nordic Red Cattle Yield Evaluation: RRM; MCEM-REML SRB VC, model december 2013 AMS+CMS

DEFINE DATADIR ..
DEFINE PARAMDIR ..

DATAFILE DATADIR/data2024_cut2005_NEWAGE_MSD0015_HETCL2_HYS_CLYMFIX_FULL.dat
INTEGER  HERD ANML TGRP ANIxLAC HY HTD Hslope YM mthx4yr AGEclass DCC DIM DAYSDR YEAR YML CLYRMTHL HET-CL YMxMS MS YMLxMS BREED
REAL     MILK PROTEIN FAT Wmlk Wprt Wfat DO a_lin a_qua a_linB1 a_quaB1 a_linB2 a_quaB2 x1:2 het1:5 hetTOT rec1:5 recTOT hetTOTNRB recTOTNRB

DEFINE Hett     HETT(hetTOTNRB | HET-CL)
DEFINE Rect     RECT(recTOTNRB | HET-CL)
DEFINE Hysd     HYSD(HY)
DEFINE Hyfi     HYFI(HY)
DEFINE Season   SEASON(t2:6 | CLYRMTHL)
DEFINE AgeBreed AGEBREED(a_lin a_qua a_linB1 a_quaB1 a_linB2 a_quaB2 | AGEclass)
DEFINE Hslp     HSLP(t2 | Hslope)
DEFINE Fihtd    FIHTD(HTD)
DEFINE Sdhtd    SDHTD(HTD)
DEFINE FIPEno   FIPE(- - - - - - - - -)
DEFINE SDPEno   SDPE(- - - - - - - - -)
DEFINE FIPE2no  FIPE2(- - - - - - -)

MODEL
  MILK(1)    = Hett@1 Rect@1   -    Hyfi   YMxMS Season   -    AgeBreed DCC Hslp   Fihtd      -    FIPE(t9:17  | ANML)@1 SDPEno                  FIPE2no                      ADDANI(t192:206 | ANML)@1
  PROTEIN(1) = Hett@2 Rect@2   -    Hyfi   YMxMS Season   -    AgeBreed DCC Hslp   Fihtd      -    FIPE(t18:26 | ANML)@1 SDPEno                  FIPE2no                      ADDANI(t207:221 | ANML)@1
  FAT(1)     = Hett@3 Rect@3   -    Hyfi   YMxMS Season   -    AgeBreed DCC Hslp   Fihtd      -    FIPE(t27:35 | ANML)@1 SDPEno                  FIPE2no                      ADDANI(t222:236 | ANML)@1
  MILK(2)    = Hett@1 Rect@1 Hysd@1   -    YMxMS Season   -    AgeBreed DCC Hslp      -    Sdhtd@1 FIPEno                SDPE(t90:98   | ANML)@1 FIPE2no                      ADDANI(t192:206 | ANML)@1
  PROTEIN(2) = Hett@2 Rect@2 Hysd@2   -    YMxMS Season   -    AgeBreed DCC Hslp      -    Sdhtd@2 FIPEno                SDPE(t99:107  | ANML)@1 FIPE2no                      ADDANI(t207:221 | ANML)@1
  FAT(2)     = Hett@3 Rect@3 Hysd@3   -    YMxMS Season   -    AgeBreed DCC Hslp      -    Sdhtd@3 FIPEno                SDPE(t108:116 | ANML)@1 FIPE2no                      ADDANI(t222:236 | ANML)@1
  MILK(3)    = Hett@1 Rect@1 Hysd@1   -    YMxMS Season   -    AgeBreed DCC Hslp      -    Sdhtd@1 FIPEno                SDPE(t90:98   | ANML)@1 FIPE2no                      ADDANI(t192:206 | ANML)@1
  PROTEIN(3) = Hett@2 Rect@2 Hysd@2   -    YMxMS Season   -    AgeBreed DCC Hslp      -    Sdhtd@2 FIPEno                SDPE(t99:107  | ANML)@1 FIPE2no                      ADDANI(t207:221 | ANML)@1
  FAT(3)     = Hett@3 Rect@3 Hysd@3   -    YMxMS Season   -    AgeBreed DCC Hslp      -    Sdhtd@3 FIPEno                SDPE(t108:116 | ANML)@1 FIPE2no                      ADDANI(t222:236 | ANML)@1
  MILK(4)    = Hett@4 Rect@4   -    Hyfi@1 YMxMS Season DAYSDR AgeBreed DCC Hslp@1 Fihtd@1    -    FIPE(t36:44 | ANML)@2 SDPEno                  FIPE2no                      ADDANI(t237:251 | ANML)@1
  PROTEIN(4) = Hett@5 Rect@5   -    Hyfi@2 YMxMS Season DAYSDR AgeBreed DCC Hslp@2 Fihtd@2    -    FIPE(t45:53 | ANML)@2 SDPEno                  FIPE2no                      ADDANI(t252:266 | ANML)@1
  FAT(4)     = Hett@6 Rect@6   -    Hyfi@3 YMxMS Season DAYSDR AgeBreed DCC Hslp@3 Fihtd@3    -    FIPE(t54:62 | ANML)@2 SDPEno                  FIPE2no                      ADDANI(t267:281 | ANML)@1
  MILK(5)    = Hett@4 Rect@4 Hysd@4   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@4    -    Sdhtd@4 FIPEno                SDPE(t117:125 | ANML)@2 FIPE2no                      ADDANI(t237:251 | ANML)@1
  PROTEIN(5) = Hett@5 Rect@5 Hysd@5   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@5    -    Sdhtd@5 FIPEno                SDPE(t126:134 | ANML)@2 FIPE2no                      ADDANI(t252:266 | ANML)@1
  FAT(5)     = Hett@6 Rect@6 Hysd@6   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@6    -    Sdhtd@6 FIPEno                SDPE(t135:143 | ANML)@2 FIPE2no                      ADDANI(t267:281 | ANML)@1
  MILK(6)    = Hett@4 Rect@4 Hysd@4   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@7    -    Sdhtd@4 FIPEno                SDPE(t117:125 | ANML)@2 FIPE2no                      ADDANI(t237:251 | ANML)@1
  PROTEIN(6) = Hett@5 Rect@5 Hysd@5   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@8    -    Sdhtd@5 FIPEno                SDPE(t126:134 | ANML)@2 FIPE2no                      ADDANI(t252:266 | ANML)@1
  FAT(6)     = Hett@6 Rect@6 Hysd@6   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@9    -    Sdhtd@6 FIPEno                SDPE(t135:143 | ANML)@2 FIPE2no                      ADDANI(t267:281 | ANML)@1
  MILK(7)    = Hett@7 Rect@7   -    Hyfi@1 YMxMS Season DAYSDR AgeBreed DCC Hslp@1 Fihtd@1    -    FIPE(t63:71 | ANML)@3 SDPEno                  FIPE2(t171:177 | ANIxLAC)@1 ADDANI(t282:296 | ANML)@1
  PROTEIN(7) = Hett@8 Rect@8   -    Hyfi@2 YMxMS Season DAYSDR AgeBreed DCC Hslp@2 Fihtd@2    -    FIPE(t72:80 | ANML)@3 SDPEno                  FIPE2(t178:184 | ANIxLAC)@1 ADDANI(t297:311 | ANML)@1
  FAT(7)     = Hett@9 Rect@9   -    Hyfi@3 YMxMS Season DAYSDR AgeBreed DCC Hslp@3 Fihtd@3    -    FIPE(t81:89 | ANML)@3 SDPEno                  FIPE2(t185:191 | ANIxLAC)@1 ADDANI(t312:326 | ANML)@1
  MILK(8)    = Hett@7 Rect@7 Hysd@4   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@4    -    Sdhtd@4 FIPEno                SDPE(t144:152 | ANML)@3 FIPE2no                      ADDANI(t282:296 | ANML)@1
  PROTEIN(8) = Hett@8 Rect@8 Hysd@5   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@5    -    Sdhtd@5 FIPEno                SDPE(t153:161 | ANML)@3 FIPE2no                      ADDANI(t297:311 | ANML)@1
  FAT(8)     = Hett@9 Rect@9 Hysd@6   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@6    -    Sdhtd@6 FIPEno                SDPE(t162:170 | ANML)@3 FIPE2no                      ADDANI(t312:326 | ANML)@1
  MILK(9)    = Hett@7 Rect@7 Hysd@4   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@7    -    Sdhtd@4 FIPEno                SDPE(t144:152 | ANML)@3 FIPE2no                      ADDANI(t282:296 | ANML)@1
  PROTEIN(9) = Hett@8 Rect@8 Hysd@5   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@8    -    Sdhtd@5 FIPEno                SDPE(t153:161 | ANML)@3 FIPE2no                      ADDANI(t297:311 | ANML)@1
  FAT(9)     = Hett@9 Rect@9 Hysd@6   -    YMxMS Season DAYSDR AgeBreed DCC Hslp@9    -    Sdhtd@6 FIPEno                SDPE(t162:170 | ANML)@3 FIPE2no                      ADDANI(t312:326 | ANML)@1
```

Modifying MiX99 solver operation at run-time

- External `STOP` file:
  - Solver stops iteration if file exists.

```
touch STOP
```

Modifying MiX99 solver operation at run-time

- External `STOP` file: `touch STOP`
  - Solver stops iteration if file exists.

- External `PEEK` file: `[[-]PITER [PSTEP]]` `echo "-100 10" > PEEK`
  - Stores intermediate solutions.
  - Optionally at certain iteration and interval.

Modifying MiX99 solver operation at run-time

- External `STOP` file:
  - Solver stops iteration if file exists.

```
touch STOP
```

- External `PEEK` file: `[[-]PITER [PSTEP]]`
  - Stores intermediate solutions.
  - Optionally at certain iteration and interval.

```
echo "-100 10" > PEEK
```

- External `ITER` file:
  - Modifies iteration parameters.
  - One or two lines: PCG and VCE or HV:

```
echo "6000 1.0e-6 R F" > ITER
```

```
# STOP: maxiter, tolerance,  criterion (A/R/M/D), [enforce (F)]:
        6000     1.0e-6      M                     F
# STOPE: REMLrounds, nSamples, stopCritVCE
         200         10        1.0e-10
```

Modifying MiX99 solver operation at run-time

- External `STOP` file:
  - Solver stops iteration if file exists.

```
touch STOP
```

- External `PEEK` file: `[[-]PITER [PSTEP]]`
  - Stores intermediate solutions.
  - Optionally at certain iteration and interval.

```
echo "-100 10" > PEEK
```

- External `ITER` file:
  - Modifies iteration parameters.
  - One or two lines: PCG and VCE or HV:

```
# STOP: maxiter, tolerance,  criterion (A/R/M/D), [enforce (F)]:
        6000     1.0e-6      M                     F
# STOPE: REMLrounds, nSamples, stopCritVCE
         200         10        1.0e-10
```

```
echo "6000 1.0e-6 R F" > ITER
```

```
touch ITER.LOCK
cat <<EOD > ITER
# PCG: maxiter tol criterion F
6000 1.0e-6 M F
# REML: maxrounds samples tol
200 10 1.0e-10
EOD
rm -f ITER.LOCK
```

MiX99 solver memory options for SNP-BLUP

When solving **pure SNP-BLUP**:

- `REGMATRIX FORMAT` options "`m`" (markers), "`s`" (squeezed), and "`pb`" (PLINK), store marker values (0,1,2) **byte-packed** to internal files:

  ```
  REGMATRIX RANDOM SNP ID=1 FIRST=2 LAST=46913 FORMAT=markers
  ```

  ⇒ 20 times smaller internal files.

MiX99 solver memory options for SNP-BLUP

When solving **pure SNP-BLUP**:

- `REGMATRIX FORMAT` options "`m`" (markers), "`s`" (squeezed), and "`pb`" (PLINK), store marker values (0,1,2) **byte-packed** to internal files:

  ```
  REGMATRIX RANDOM SNP ID=1 FIRST=2 LAST=46913 FORMAT=markers
  ```

  ⇒ 20 times smaller internal files.

- MiX99 "serial" solver `mix99s` has various memory options:
  - Marker values fully in **double precision** real memory (`RDX`).
  - Fully byte-packed in memory (40 times smaller), **unpacked in blocks**.
  - **Reads in blocks** from internal file (again and again).
  - Block size can be controlled (`RDS,RDM,RDL,RDB b`) and memory usage limited (`RDU m`).

MiX99 solver memory options for SNP-BLUP

When solving **pure SNP-BLUP**:

- `REGMATRIX FORMAT` options "`m`" (markers), "`s`" (squeezed), and "`pb`" (PLINK), store marker values (0,1,2) **byte-packed** to internal files:

  ```
  REGMATRIX RANDOM SNP ID=1 FIRST=2 LAST=46913 FORMAT=markers
  ```

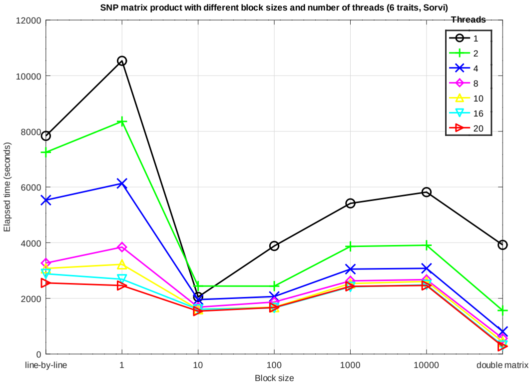
  ⇒ 20 times smaller internal files.

- MiX99 "serial" solver `mix99s` has various memory options:
  - ► Marker values fully in **double precision** real memory (`RDX`).
  - ► Fully byte-packed in memory (40 times smaller), **unpacked in blocks**.
  - ► **Reads in blocks** from internal file (again and again).
  - ► Block size can be controlled (`RDS`,`RDM`,`RDL`,`RDB b`) and memory usage limited (`RDU m`).

- What is fastest depends on available memory and **bandwidth**.

Performance testing SNP-BLUP in MiX99

- **1M genotypes, 50k SNPs, 6 traits.**
- 20 PCG iterations of MiX99 solver.
- Different memory options:
  - ▶ Line-by-line (reading from file).
  - ▶ Unpacked in blocks from memory.
  - ▶ Full double precision matrix in memory (373 GB).
- **Two sweet spots:**
  - ▶ Unpacked in small blocks.
  - ▶ Full double precision matrix in memory. Faster in this server.



Luke

NATURAL RESOURCES
INSTITUTE FINLAND