MiX99

Solving Large Mixed Model Equations

MTT

CLIM tutorial

Ismo Strandén, MTT

Command Language Interface for MiX99

- Original MiX99 interface:
  - Directive file
  - Answers to questions
  - ➔ not so easy for user

- CLIM:
  - Command driven
  - Order of commands flexible
  - Easy to read and change

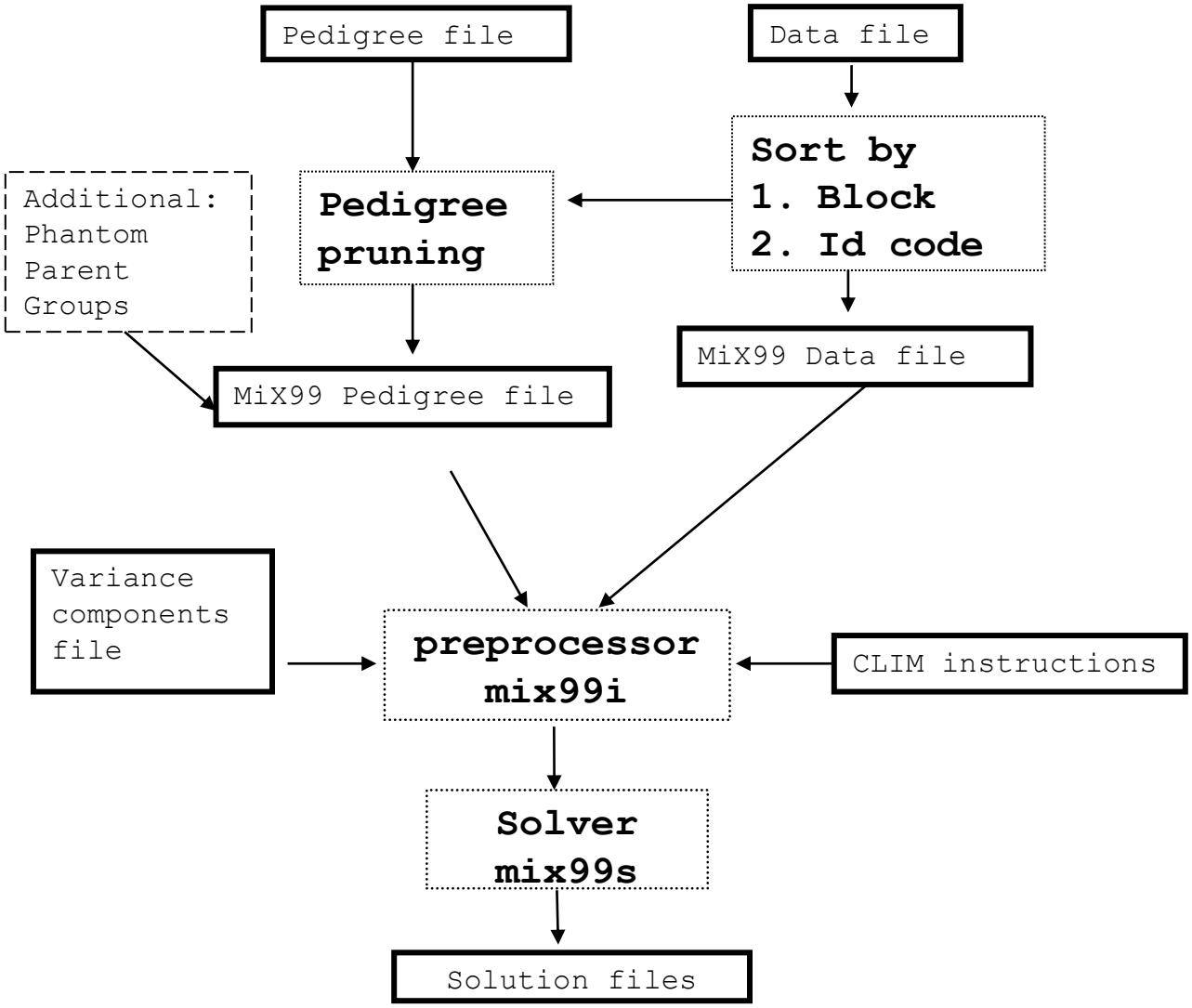
This presentation:

WARNING: contains material from the CLIM manual

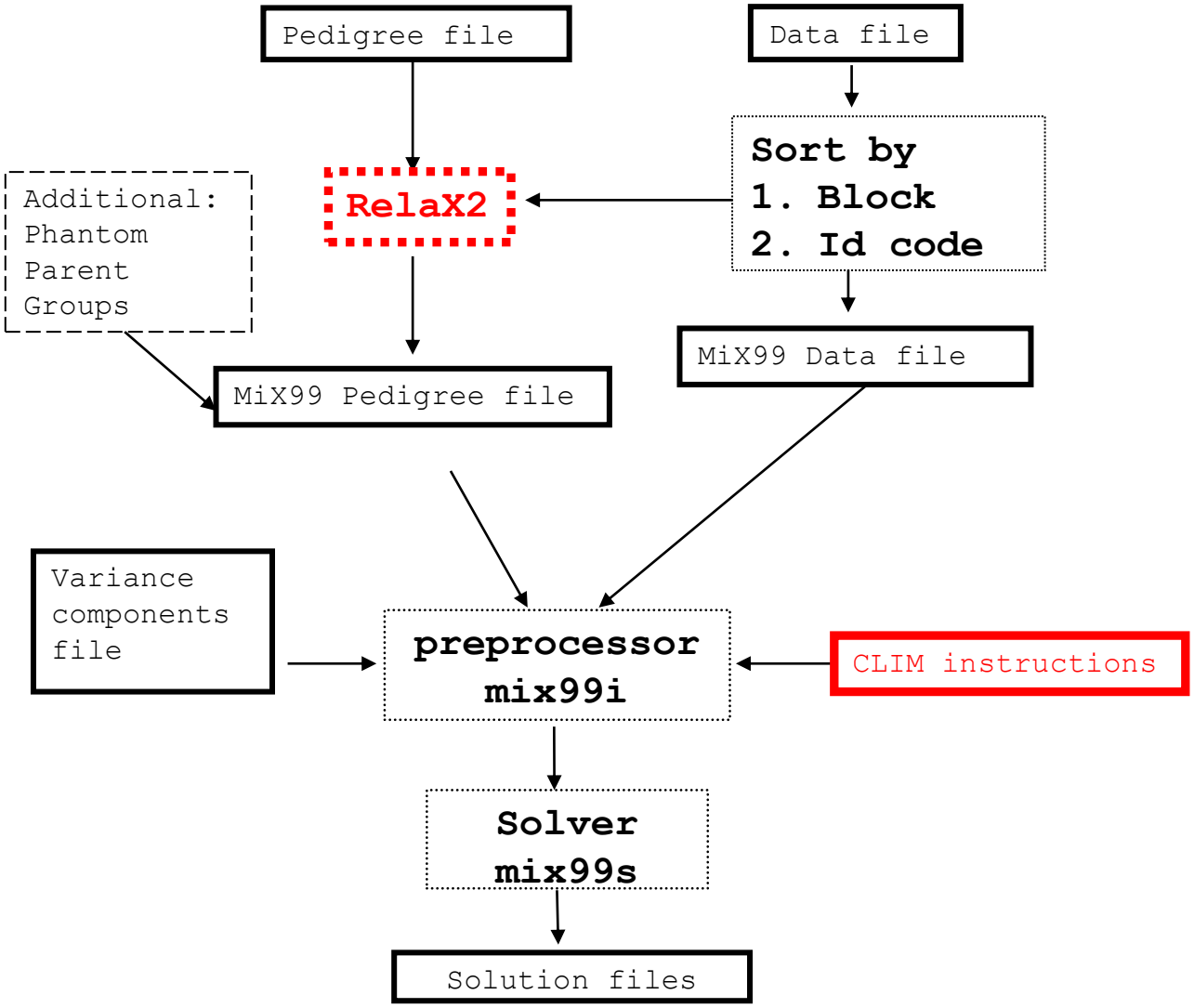
How to use MiX99 solver through CLIM

& pedigree for MiX99 through RelaX2

MiX99 workflow



MiX99 workflow



RelaX2: Pedigree analysis program

1. Check pedigree for consistency, correct some errors
2. Prune pedigree
   - Different methods available
3. New pedigree
   - Pruned, corrected, formatted, ordered …
4. Relationship information
   - Inbreeding coefficients, relationship matrix
5. Population statistics
   - Effective population size, generation interval
6. Breed proportions
7. …

Command language driven

Simple command file for RelaX2

```
# NOTE: no output to file is requested!

input pedigree          # Pedigree input
  file relaxc.ped       # This file has pedigree
  record id sire dam    # This input information

# Following are some additions that might be useful:
# output overwrite inbreeding simple.inbr # Inreeding
# output overwrite cycles     simple.cycl # Cycles
# nopopulations # No calculation of number of populations
```

Execution of RelaX2 having commands in file `simple.dir`:

```
relax2 < simple.dir
```

Format for pedigree

Integer or character string in id code

Integers:

1. Record id 1 sire 2 dam 3
2. Record id sire dam
3. Record id 1:10 sire 50:59 dam 70:79

Character strings:

4. Record id $ 1:10 sire $ 50:59 dam $ 70:79

Unknown parent can be space in 3. and 4.

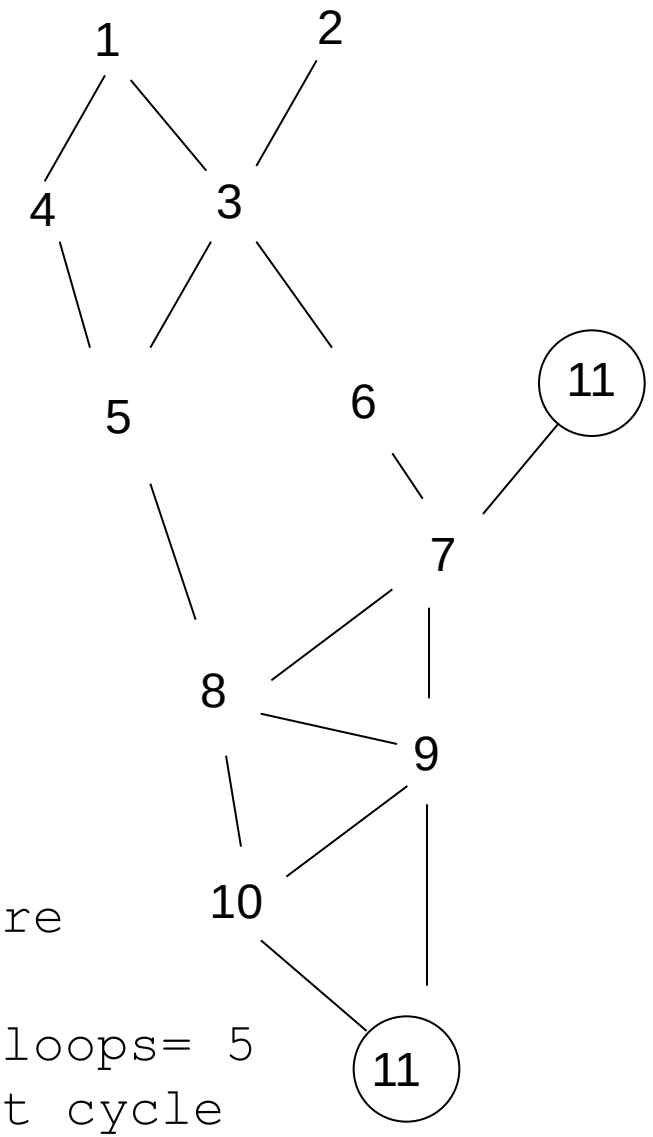
RelaX2: Pedigree

Some commands:

- `Input/Output pedigree`: animal model pedigree
- `Input/Output smpedigree`: sire model pedigree
- `Singlepopulation`
  - Chooses the largest pedigree among separate ones
- `Prune Some_Method`
  - Prunes pedigree of unnecessary animals.
- `Input select` (selecting for statistics calculations)
  - Chooses individuals according to some external information for calculation of statistics (mean generation interval, inbreding coefficients etc.)
- `Output Amatrix`: relationship matrix

RelaX2 example: pedigree loops

```
> cat relaxc.ped
 1   0   0 10 01/1998
 2   0   0 10 01/1999
 3   1   2 10 01/2000
 4   1   0 20 02/2000
 5   4   3 20 01/2001
 6   0   3 20 02/2001
 7   6 11 30 01/2002
 8   7   5 30 01/2003
 9   7   8 30 01/2004
10   9   8 30 02/2005
11   9 10 30 01/2006
```



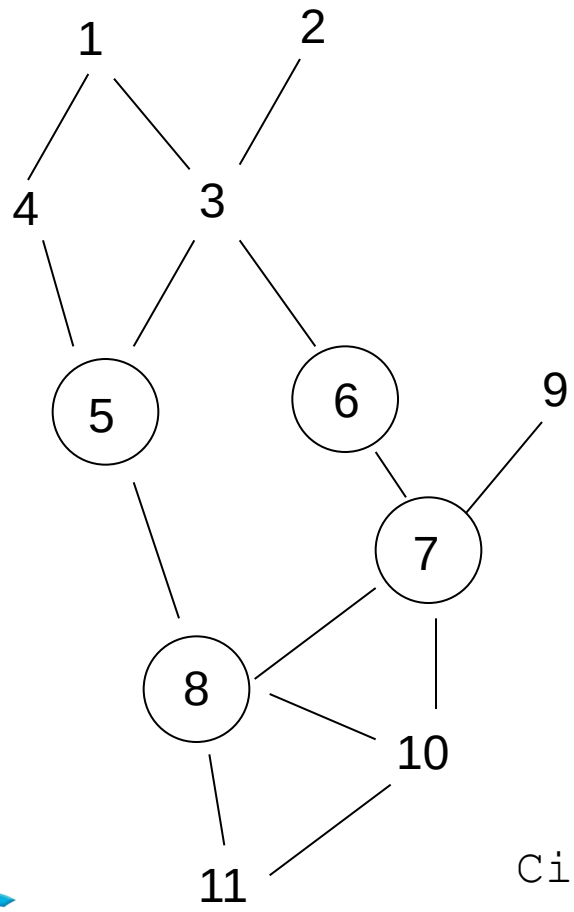
RelaX2 says:

```
There are individual(s) that are
parents for themselves.
Number of individuals in such loops= 5
Please rerun RelaX2 and request cycle
output file (e.g., output cycles cycles.dat)
```

Example on pruning for MiX99

```
input pedigree
  file some.ped     # pedigree file
  record id sire dam group
  prune prediction # animals with data and their
                   # ancestors, only informative
                   # except always sire and dam

input data         # animals in data file have observ.
  file some.dat
  record id group  # id code and group/block code

mix99 keep # MiX99 format for output pedigree;
           # group in pedigree is kept if
           # animal has no observation

output overwrite pedigree pruned.ped
```

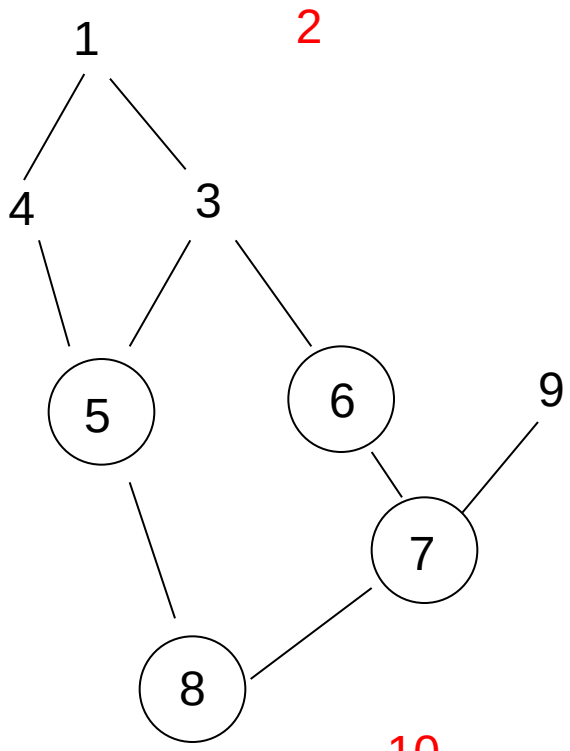
Pruning: original pedigree



1	-1	-1	10
2	-2	-2	10
3	1	2	10
4	1	-4	20
5	4	3	20
6	-6	3	20
7	6	9	30
8	7	5	30
9	-5	-9	30
10	7	8	30
11	10	8	30

Circled animals are in the data file

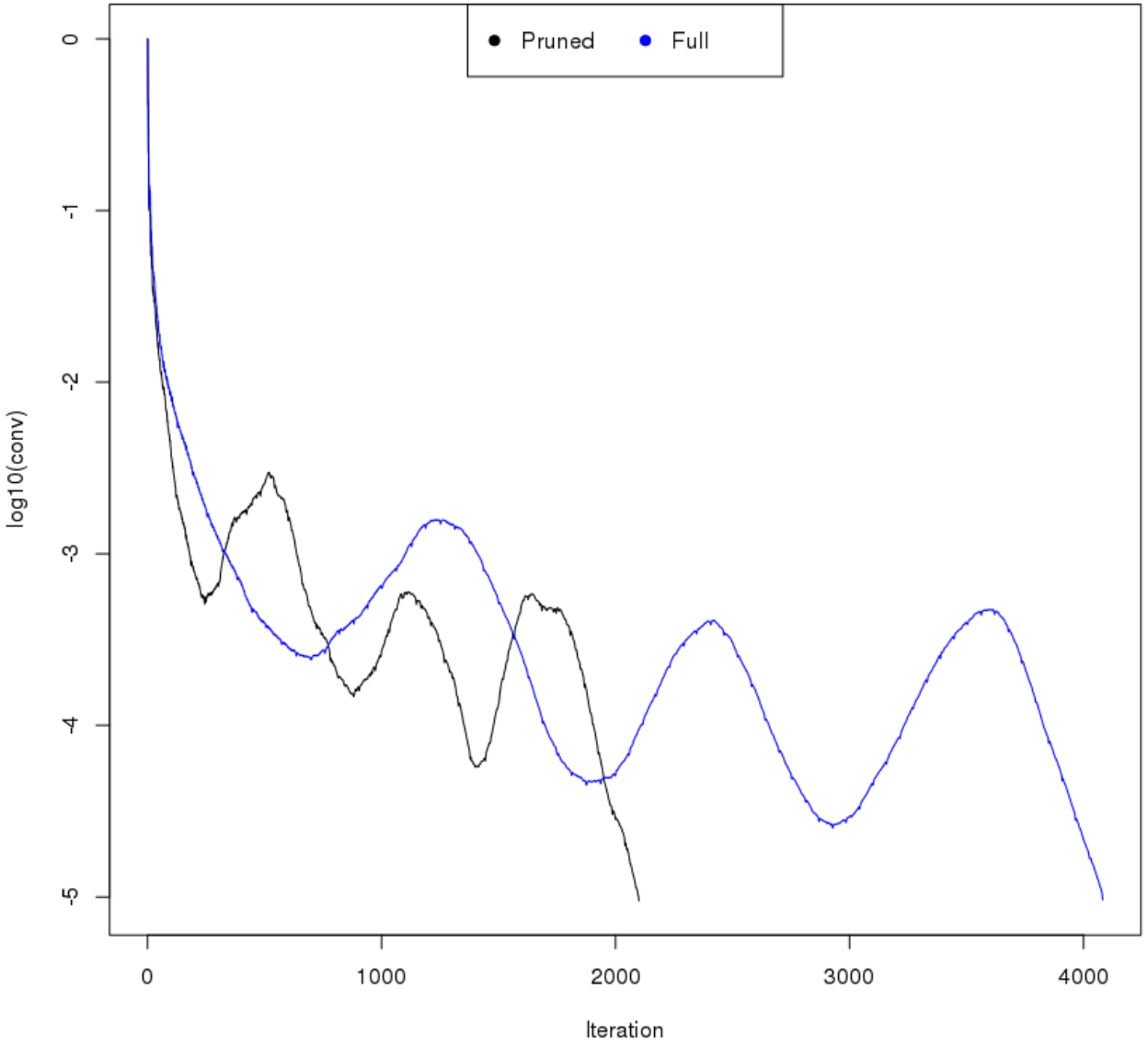
Pruned pedigree



1	-1	-1	10
2	-2	-2	10
3	1	**0**	10
4	1	-4	20
5	4	3	20
6	-6	3	20
7	6	9	30
8	7	5	30
9	-5	-9	30
10	7	8	30
11	10	8	30

NOTE: Offspring of Dam 9 has observation but offspring of Dam 2 (animal 3) has no observation

Effect of pruning to convergence



Short summary on RelaX2

1. Read pedigree and check its goodness
   - Need any corrections? Can use so called `errata` file
2. Prune pedigree, use data file
3. Output pedigree in MiX99 format

4. Other options of interest:
   - Calculate inbreeding coefficients (MiX99 cannot calculate)
   - Calculate relationship matrix (useful for single-step)
   - Calculate breed proportions

CLIM by example
Example: Single trait animal model

A simple single trait animal model has the form

$$y = Xb + Za + e$$

where

- y is $n \times 1$ vector of observations,
- b is $p \times 1$ vector of fixed effects,
- X is $n \times p$ design matrix to link observations to appropriate fixed effects,
- a is $q \times 1$ vector of random additive genetic effects,
- Z is $n \times q$ design matrix to link observations to appropriate random effects,
- e is $n \times 1$ random residual vector.

Animal model example by CLIM:

R.Mrode. ”Linear models for the Prediction of Animal Breeding Values”. Ex. 3.1

Model: $\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$

$$\begin{aligned} \mathrm{E}(\boldsymbol{a}) &= \mathbf{0} & \mathrm{Var}(\boldsymbol{a}) &= \boldsymbol{A}\sigma_a^2 \\ \mathrm{E}(\boldsymbol{e}) &= \mathbf{0} & \mathrm{Var}(\boldsymbol{e}) &= \boldsymbol{I}\sigma_e^2 \\ \mathrm{E}(\boldsymbol{y}) &= \boldsymbol{X}b & \mathrm{Cov}(\boldsymbol{a},\boldsymbol{e}) &= \mathbf{0} \end{aligned}$$

Mixed model equations:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Example case:

Fixed effects: sex

Random effects: additive genetics and residual

CLIM instruction file: AM.clm

```
herakles:~/MiX99/example/serial/animal_model> cat AM_example_3_1.clm
#-----------------------------------------------------------------
#  R. Mrode: Linear Models for the Prediction of Animal Breeding Values
#
#  A model for an animal evaluation (animal model)
#  Example 3.1
#
#  MODEL:  Preweaning gain = sex + animal
#-----------------------------------------------------------------

DATAFILE   AM_ex31.dat # Name of data file

INTEGER    animal sex  # Integer number column names
REAL       weaningW    # Real number column names

PEDFILE    AM_ex31.ped # Name of Pedigree file
PEDIGREE   animal am   # Genetics associated with animal code
                       # am=animal model

PARFILE    AM_ex31.var # Name of variance component file

MODEL
  weaningW = sex animal # The model
```

sex

weaningW

animal

```
herakles:~/clim_examples/animal_model_RMrode> cat AM_ex31.da
4   1   4.5
5   2   2.9
6   2   3.9
7   1   3.5
8   1   5.0
```

Sire and dam ID

animal ID

```
herakles:~/clim_examples/animal_model_RMrode> cat AM_ex31.ped
1   0   0
2   0   0
3   0   0
4   1   0
5   3   2
6   1   2
7   4   5
8   3   6
```

```
herakles:~/clim_examples/animal_model_RMrode> cat AM_ex31.var
1   1   1   20.0    Genetic variance
2   1   1   40.0    Residual variance
```

MiX99 solving

- MiX99 preprocessor:

  `mix99i AM_example_3_1.clm >AM_mix99i.log`

  - `Tmp-files`
  - `Modlog, Parlog, Tralog`
  - `Resid.List`
  - `OK_mix99i`

- MiX99 solver: `mix99s -s > AM_mix99s.log`

  - `Conlog`
  - `Solfix, Solani`
  - `Solvec`
  - `OK_mix99s`

Repeatability animal model

- Consider model:
  - y = herd-year + permanent environment + add.genet. + residual
  - **y** = **Xb** + **Zp** + **Za** + **e**, with

$$\begin{aligned} \mathrm{Var}(\boldsymbol{p}) &= \boldsymbol{I}\sigma_p^2 \\ \mathrm{Var}(\boldsymbol{a}) &= \boldsymbol{A}\sigma_a^2 \\ \mathrm{Var}(\boldsymbol{e}) &= \boldsymbol{I}\sigma_e^2 \end{aligned}$$

- When no permanent environment effect:
  - **y** = **Xb** + **Za** + **e**

and CLIM instructions are previous example

CLIM and repeatability model

- Problem: two effects use the same integer number column: `animal`
  - Which of the effects is additive genetic effect
  - How about number of random effect
  - Within block ordering?

- Solution: user defined component names

```
INTEGER    animal sire herd_year ones
REAL       tr12

PEDIGREE   animal am
MODEL
  tr12 = herd_year animal animal
```

Not acceptable

```
Model error: Same component name appears twice on the line.
  On line:   9
  Near token: animal
  Complete line information:
  tr12 = herd_year animal animal
```

Component names for repeatability model

$$tr_{12} = herd \times year + p + a + e$$

User defined component name: **G**

```
INTEGER    animal sire herd_year ones
REAL       tr12

PEDIGREE   G am    # G for animal genetics
RANDOM     animal # permanent environment

MODEL
  tr12 = herd_year animal G(animal)
```

```
INTEGER
REAL

PEDIGREE   animal am
MODEL
  tr12 = herd_year animal animal
```

Permanent environment is random

Repeatability model: alternatives

```
INTEGER    animal sire herd_year ones
REAL       tr12

PEDIGREE   G am    # G for animal genetics
RANDOM     animal # permanent environment

MODEL
  tr12 = herd_year animal G(animal)
```

1:

```
PEDIGREE    G am #  G for animal genetics
RANDOM      PE   # PE for permanent environment
MODEL
  tr12 = herd_year PE(animal) G(animal)
```

2:

```
PEDIGREE    animal am # animal for animal genetics
RANDOM      PE        # PE     for permanent environment
MODEL
  tr12 = herd_year PE(animal) animal
```

Component names

- Allows accessing the same integer column
- Can be used to combine effects for easier reading
- Can be used to group random effects with same variance-covariance structure:
  - Direct and maternal component
  - Random regression effects

- Note: unique name must be chosen, not used column name or reserved name

```
Improper pedigree effect name.
  On line:   6
  Near token: real
  Complete line information:
  PEDIGREE real  am
```

```
Error: random effect can only be within integer class. Cannot be a random effect
  On line:   7
  Near token: block
  Complete line information:
  RANDOM   block
```

Solutions

MiX99 solver output:

```
Solutions for First 20 Levels of Across-Block Fixed Effect:  1  herd_yea
-----------------------------------------------------------------------
Fact.Trt _____Level_____     N-Obs     Eq-No    Solution     Factor
   1   1              11         2        11     99.8333     herd_yea
```

```
herakles:~/MiX99/example/serial/clim_examples/repeatability_model> cat Solfix
 Fact. Trt      Level    N-Obs  Solution               Factor Trait
   1   1         11          2   99.833               herd_yea tr12
   1   1         12          3   123.01               herd_yea tr12
   1   1         21          2   194.83               herd_yea tr12
   1   1         22          3   129.68               herd_yea tr12
```

Fixed effects

```
   1   1              4         2        15   -0.888889     animal
```

```
herakles:~/MiX99/example/serial/clim_examples/repeatability_model> cat Solr01
          4         2 -0.88889
          6         2  0.88889
          8         2   1.1867
          9         2 -0.55846
         10         2 -0.62827
```

Permanent environment

```
First 20 Ani
-----------
Fact.Trt __Animal-ID____ N-Desc  N-Obs     Eq-No    Solution     Factor
   1   1              1      2      0         1    0.427759E-10 animal
   1   1              2      2      0         2    0.427759E-10 animal
```

```
herakles:~/MiX99/example/serial/clim_examples/repeatability_model> cat Solani
          1      2      0  0.42776E-10
          2      2      0  0.42776E-10
          3      2      0  0.33333
          4      2      2 -0.33333
          5      3      0  0.38932E-10
          6      3      2  0.66667
          7      1      0  0.97731E-01
          8      1      2  0.98778
          9      0      2 -0.85515E-01
         10      0      2  0.71553E-01
```

Animal genetics

Multiple traits

- Repeatability model:
- Two trait model:

$$tr_{12} = herd \times year + p + a + e$$

$$tr_1 = herd \times year + a_1 + e_1$$
$$tr_2 = \mu + a_2 + e_2$$

two_tr.clm

```
DATAFILE example.dat

INTEGER animal sire herd_year ones
REAL    tr1 tr2

PEDFILE  AM.ped
PEDIGREE animal am

DATASORT PEDIGREECODE=animal

PARFILE  mt.var

MODEL SCALE
  tr1 =  -    herd_year animal
  tr2 = ones  -         animal
```

Important indicators

mt.var

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

Genetic

Residual

(Co)variance

Position in covariance matrix

Random effect number

Two trait model solutions

Solutions for First 20 Levels of Across-Block Fixed Effect: 1 ones

Fact.Trt	Level	N-Obs	Eq-No	Solution	Factor
1 2	1	5	21	160.285	ones

Fixed effects

Solfix

	Fact.	Trt	Level	N-Obs	Solution	Factor	Trait
Solutions	1	2	1	5	160.28	ones	tr2
	2	1	1	2	88.759	herd_yea	tr1
Fact.Trt	2	1	2	3	137.73	herd_yea	tr1
2 1			1	2	22	88.7587	herd_yea
2 1			2	3	23	137.733	herd_yea

First 20 Animal Solutions

Animal genetics

Fact.Trt	Animal-ID	N-Desc	N-Obs	Eq-No	Solution	Factor
1 1	1	2	0	1	0.142930E-05	animal
2 2	1	2	0	2	0.609187E-06	animal
1 1	2	2	0	3	0.142930E-05	animal
2 2	2	2	0	4	0.609187E-06	animal
1 1	3	2	0	5	-2.28421	animal
2 2	3	2	0	6	-2.51119	animal
1 1	4	2	1	7	2.28422	animal
2 2	4	2	1	8	2.51119	animal
1 1	5	3	0	9	-5.89685	animal
2 2	5	3	0	10	-5.89685	animal
1 1	6	3	1	11	1.32842	animal
2 2	6	3	1	12	0.874481	animal
1 1	7	1	0	13	-4.27474	animal
2 2	7	1	0	14	-4.46353	animal
1 1	8	1	1	15	-7.68036	animal
2 2	8	1	1	16	-7.61893	animal
1 1	9	0	1	17	-6.69126	animal
2 2	9	0	1	18	-7.24481	animal
1 1	10	0	1	19	-9.95859	animal
2 2	10	0	1	20	-9.94590	animal

Solani

1	2	0	0.14293E-05	0.60919E-06
2	2	0	0.14293E-05	0.60919E-06
3	2	0	-2.2842	-2.5112
4	2	1	2.2842	2.5112
5	3	0	-5.8969	-5.8969
6	3	1	1.3284	0.87448
7	1	0	-4.2747	-4.4635
8	1	1	-7.6804	-7.6189
9	0	1	-6.6913	-7.2448
10	0	1	-9.9586	-9.9459

N-Desc= Number of progeny
N-Obs= number of records

Maternal effects example

```
DATAFILE   maternal_2.dat

INTEGER    calf sire dam herd sex
REAL       B_weight B2

DATASORT   PEDIGREECODE=calf

PEDFILE    data/maternal.ped
PEDIGREE   G am
RANDOM     PE

PARFILE    maternal_2.var      # Variance c

WITHINBLOCK G PE herd

MODEL
  B_weight = herd sex PE(dam) G(calf dam)
  B2       = herd sex PE(dam) G(calf -)
```

Numbering of genetic effects!

```
1 1 1   40   permanent environment
1 2 2   40   permanent environment

2 1 1  150   additive genetic 1
2 1 3  -40   cov (genetic, maternal)
2 3 3   90   maternal genetic 1

2 2 2  150   additive genetic 2

3 1 1  350
3 2 2  350
```

Genetics has 2 class effects
Missing maternal needs to be indicated

Combining of effects:
repeatability model by two trait model

Consider two trait model:

$$tr_1 = herd \times year_1 + a_1 + e_1$$
$$tr_2 = herd \times year_2 + a_2 + e_2$$

Repeatability model has genetic correlatin of one.

OR: the two trait model:

$$tr_1 = herd \times year + a + e_1$$
$$tr_2 = herd \times year + a + e_2$$

where herd x year is by trait.

```
DATAFILE example_mt_repeat.dat

INTEGER animal sire hy_1 hy_2 ones
REAL    tr1 tr2

PEDFILE  AM.ped
PEDIGREE G am

PARFILE  mt_repeat.var

MODEL SCALE
  tr1 = hy_1 -    G(animal)@fst
  tr2 = -    hy_2 G(animal)@fst
```

Effects combined

Random regression model

Consider model (Schaeffer & Dekkers, 1994, WCGALP):

$$milk = DIM + log(305/DIM) + HTD + f(a, DIM) + e$$

where random regression function is

$$f(\boldsymbol{a}, DIM) = a_{i,1} + DIM \cdot a_{i,2} + log(305/DIM) \cdot a_{i,3}$$

Fixed regression function

Herd test day effect

CLIM model line:

```
MODEL
  milk_yd = Lact_curve(DIM ln305DIM) HTD G(1 DIM ln305DIM| animal)
```

Random regression model data can be large

data file RRM.dat

HTD_1	$animal_2$	$block_3$	DIM_1	$ln(305/DIM)_2$	$milk_3$
1	1	1	73.0	1.4298500	26.0
2	1	1	123.0	0.9081270	23.0
3	1	1	178.0	0.5385280	21.0
1	2	1	34.0	2.1939499	29.0
2	2	1	84.0	1.2894900	18.0
3	2	1	139.0	0.7858380	8.0
4	2	1	184.0	0.5053760	1.0
1	3	2	8.0	3.6408701	37.0
2	3	2	58.0	1.6598700	25.0
3	3	2	113.0	0.9929240	19.0
4	3	2	158.0	0.6577170	15.0
5	3	2	218.0	0.3358170	11.0
6	3	2	268.0	0.1293250	7.0
2	4	3	5.0	4.1108699	44.0
3	4	3	60.0	1.6259700	29.0
4	4	3	105.0	1.0663500	22.0
5	4	3	165.0	0.6143660	14.0
6	4	3	215.0	0.3496740	8.0
4	5	3	14.0	3.0812500	35.0

Limited number of different numbers

Covariable tables with index in data

HTD$_1$	animal$_2$	block$_3$	index$_4$	milk$_1$
1	1	1	8	26.0
2	1	1	14	23.0
3	1	1	19	21.0
1	2	1	5	29.0
2	2	1	11	18.0
3	2	1	16	8.0
4	2	1	20	1.0
1	3	2	2	37.0
2	3	2	6	25.0
3	3	2	13	19.0
4	3	2	17	15.0
5	3	2	22	11.0
6	3	2	23	7.0
2	4	3	1	44.0
3	4	3	7	29.0
4	4	3	12	22.0
5	4	3	18	14.0
6	4	3	21	8.0

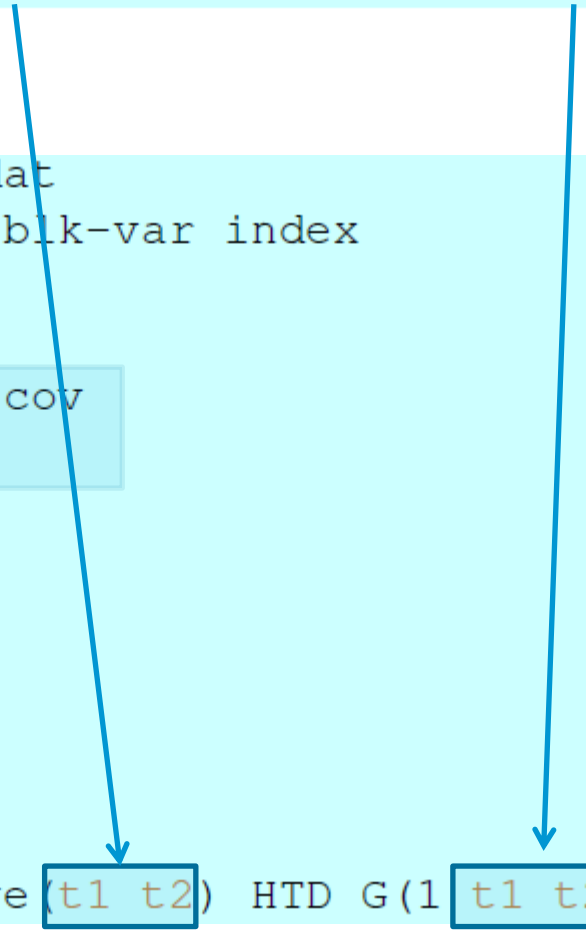
index$_1$	DIM$_1$	log(305/DIM)$_2$
1	5	4.1108699
2	8	3.6408701
3	14	3.0812500
4	31	2.2863200
5	34	2.1939499
6	58	1.6598700
7	60	1.6259700
8	73	1.4298500
9	74	1.4162500
10	81	1.3258600
11	84	1.2894900
12	105	1.0663500
13	113	0.9929240
14	123	0.9081270
15	124	0.9000300
16	139	0.7858380
17	158	0.6577170

Using covariable table files

```
MODEL
  milk_yd = Lact_curve(DIM ln305DIM) HTD G(1 DIM ln305DIM| animal)
```

With table files:

```
DATAFILE   RRM_table.dat
INTEGER    HTD animal blk-var index
REAL       milk_yd

TABLEFILE  RRM_table.cov
TABLEINDEX index

PEDFILE    RRM.ped
PEDIGREE   G am

PARFILE    RRM.var

MODEL SCALE
  milk_yd = Lact_curve(t1 t2) HTD G(1 t1 t2| animal)
```



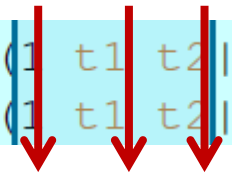
Multiple trait random regression model

```
DATAFILE   RRM_table.dat2
INTEGER    HTD animal blk-var index
REAL       milk_yd  milk_yd2

TABLEFILE   RRM_table.cov
TABLEINDEX index

PEDFILE    RRM.ped
PEDIGREE   G am

PARFILE    RRM.var2

MODEL SCALE
  milk_yd = Lact_curve(t1 t2) HTD G(1 t1 t2| animal)
  milk_yd2= Lact_curve(t1 t2) HTD G(1 t1 t2| animal)
```

Numbering in PARFILE?

Numbering of variance components within random effect

```
milk_yd = Lact_curve(t1 t2) HTD G(1 t1 t2| animal)
milk_yd2= Lact_curve(t1 t2) HTD G(1 t1 t2| animal)
```



(1,1) = 1st trait animal variance

(2,2) = 2nd trait animal variance

(3,3) = 1st trait t1 variance

(4,4) = 2nd trait t1 variance

(5,5) = 1st trait t2 variance

(6,6) = 2nd trait t2 variance

```
1  1  1    44.791
1  3  1    -0.133
1  5  1     0.351
1  3  3     0.073
1  3  5    -0.010
1  5  5     1.068

1  2  2     4.4791
1  4  2    -0.0133
1  4  2     0.0351
1  4  2     0.0073
1  4  6    -0.0010
1  6  6     0.1068

2  1  1   100.000
2  2  2    10.000
```

1st trait

2nd trait

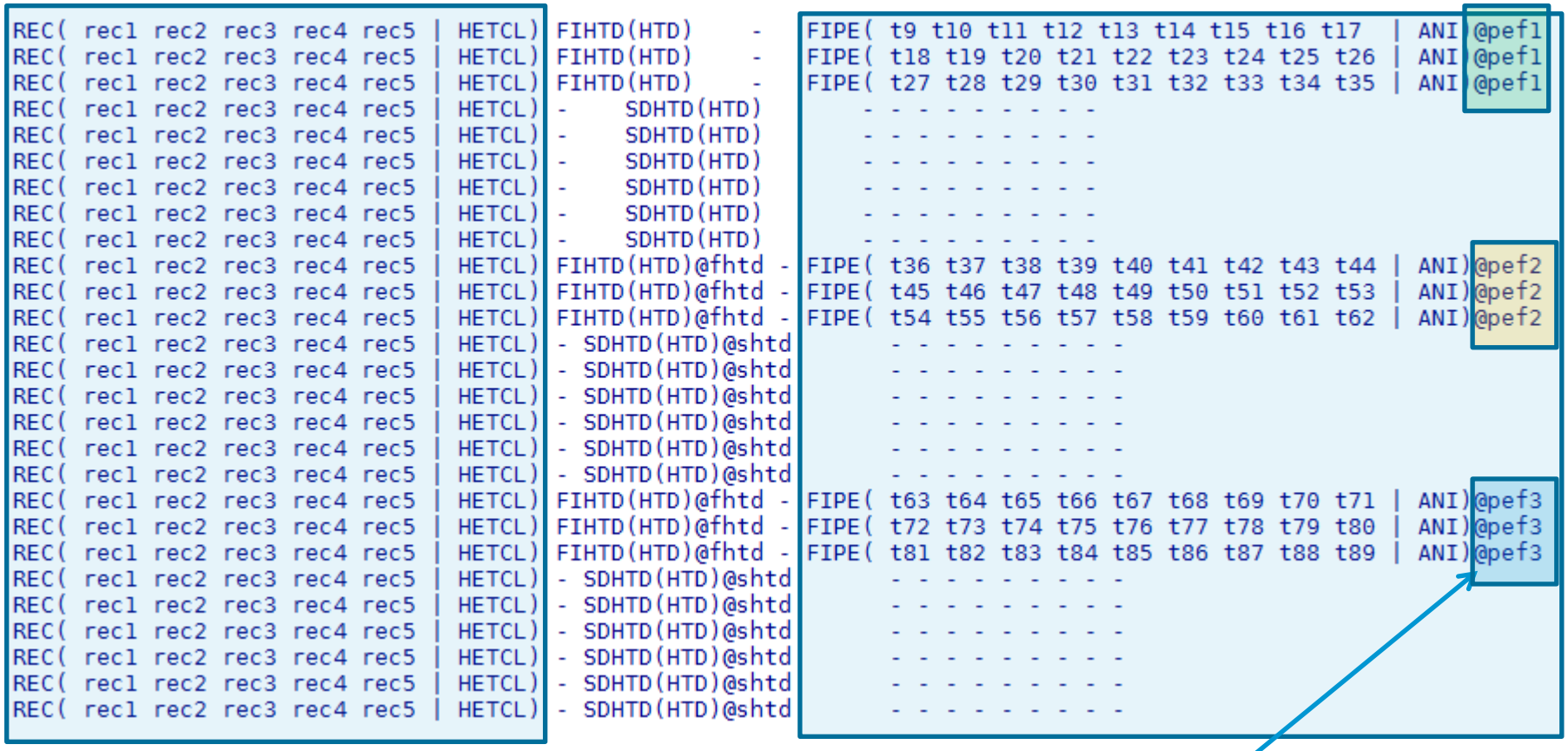
residuals

Reduced rank models

- Use:
  - Multiple traits
  - Random regression effects
  - Combining of random effects
- Large models with many effects
  - Easy to make errors in model description

Part of Nordic test-day model for production

Random regression effects with combining of effects and covariable table



REC, FIHTD, SDHTD and FIPE are all random effects

Combining indicated by symbol @

Multiple single trait analysis

- Want to analyze several models simultaneously
- Can do: multiple trait analysis with zero correlations
  - But data may be very sparse and computations very inefficient
- Instead: trait groups
  - Zero correlation between trait groups
  - Own model and variance component information
  - Trait (or model) recognized by trait group number

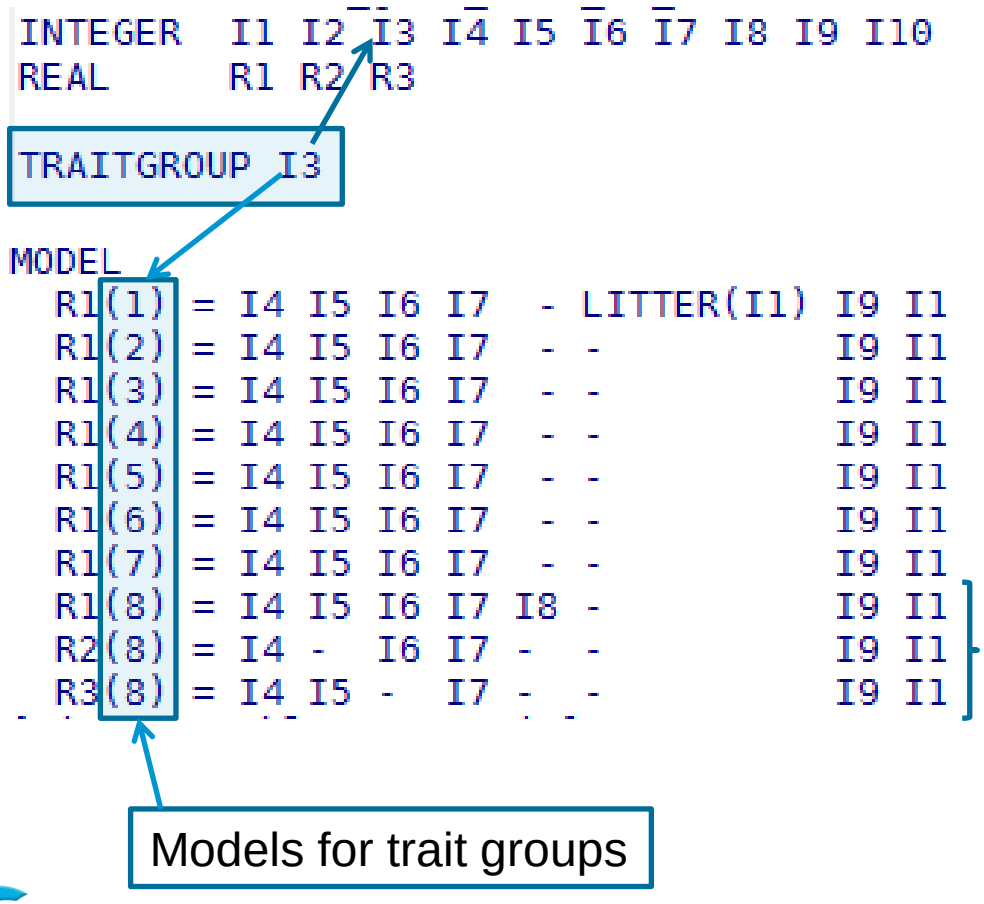
Warning: May have sometimes convergence issues!

Two single trait analyses by trait groups

```
INTEGER    animal sire herd_year ones trait
REAL       tr

TRAITGROUP trait

PEDFILE    data/AM.ped
PEDIGREE   animal am

PARFILE    mt_single.var

MODEL SCALE
   tr(1)  = -     herd_year animal
   tr(2)  = ones  -         animal
```

Trait group number in data selects model and variance components

Trait groups and models:
8 separate model analysis



Summary on CLIM

- Standard single and multiple trait models
- Many random effects
- Maternal effects
- Random regression models
- Combining of effects (Reduced rank models)

and many more … during this workshop and in the manual