UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Introduction to BLUPF90 software suite

Daniela Lourenco
UGA USA

Ignacio Aguilar
INIA Uruguay

UGA TEAM – 08/2019

BLUPF90 software suite

- Collection of software for computations
  - Focus on applications in Breeding and Genetics
- Fortran 90/95
- No GUI (graphical user interface) programs !!!
- Since 1998 by Ignacy Misztal
- First idea: to solve the MME

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\dfrac{\boldsymbol{\sigma_e^2}}{\boldsymbol{\sigma_a^2}} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'}\mathrm{y} \\ \mathbf{Z'}\mathrm{y} \end{bmatrix}$$

- First software: blupf90
- Second idea: variance components estimation
- Developers?

BLUPF90 software developers

Ignacy Misztal

Shogo Tsuruta

Andres Legarra

Ignacio Aguilar

Yutaka Masuda

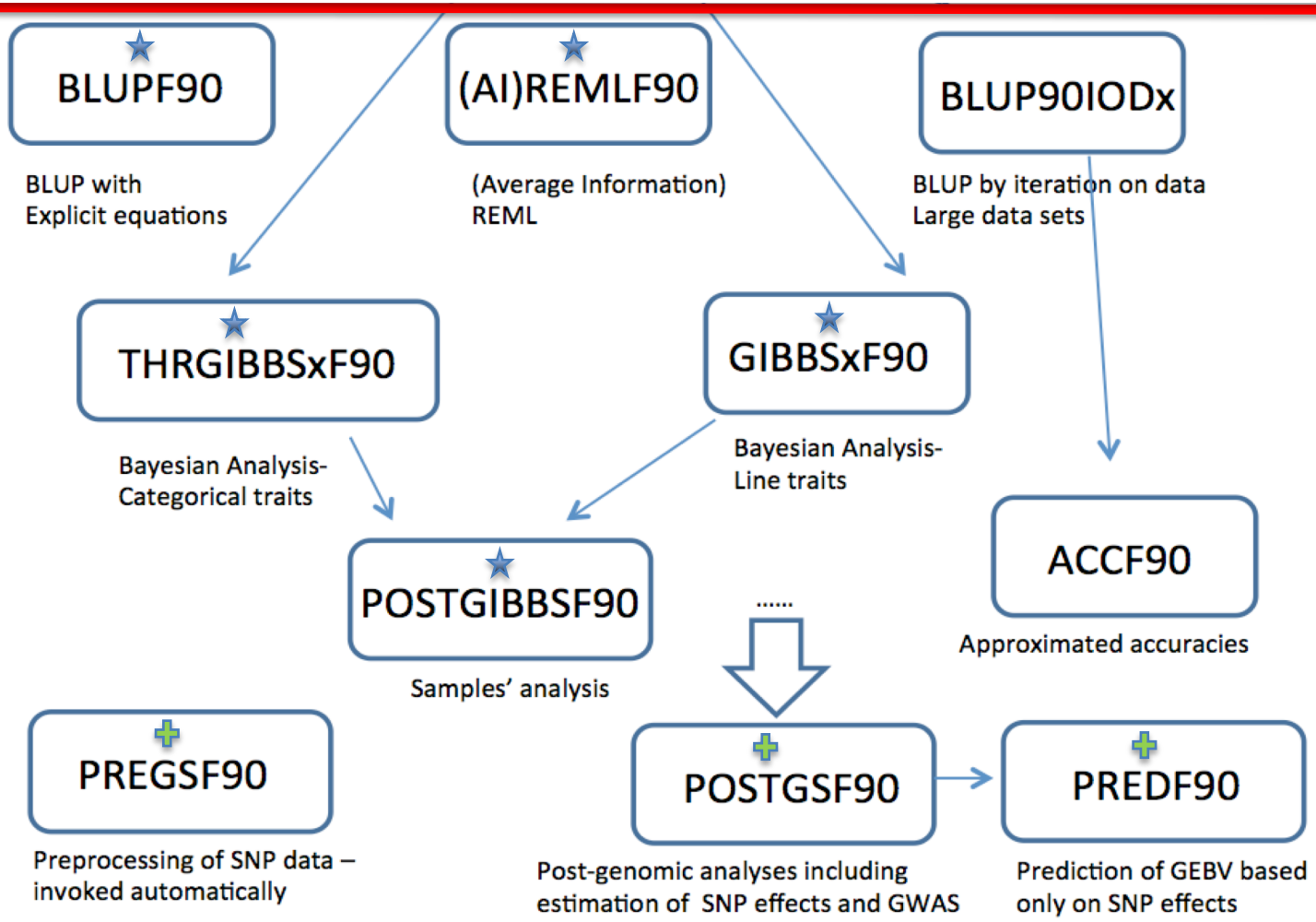
- + Several contributors
- Research turns into code

- Which programs?

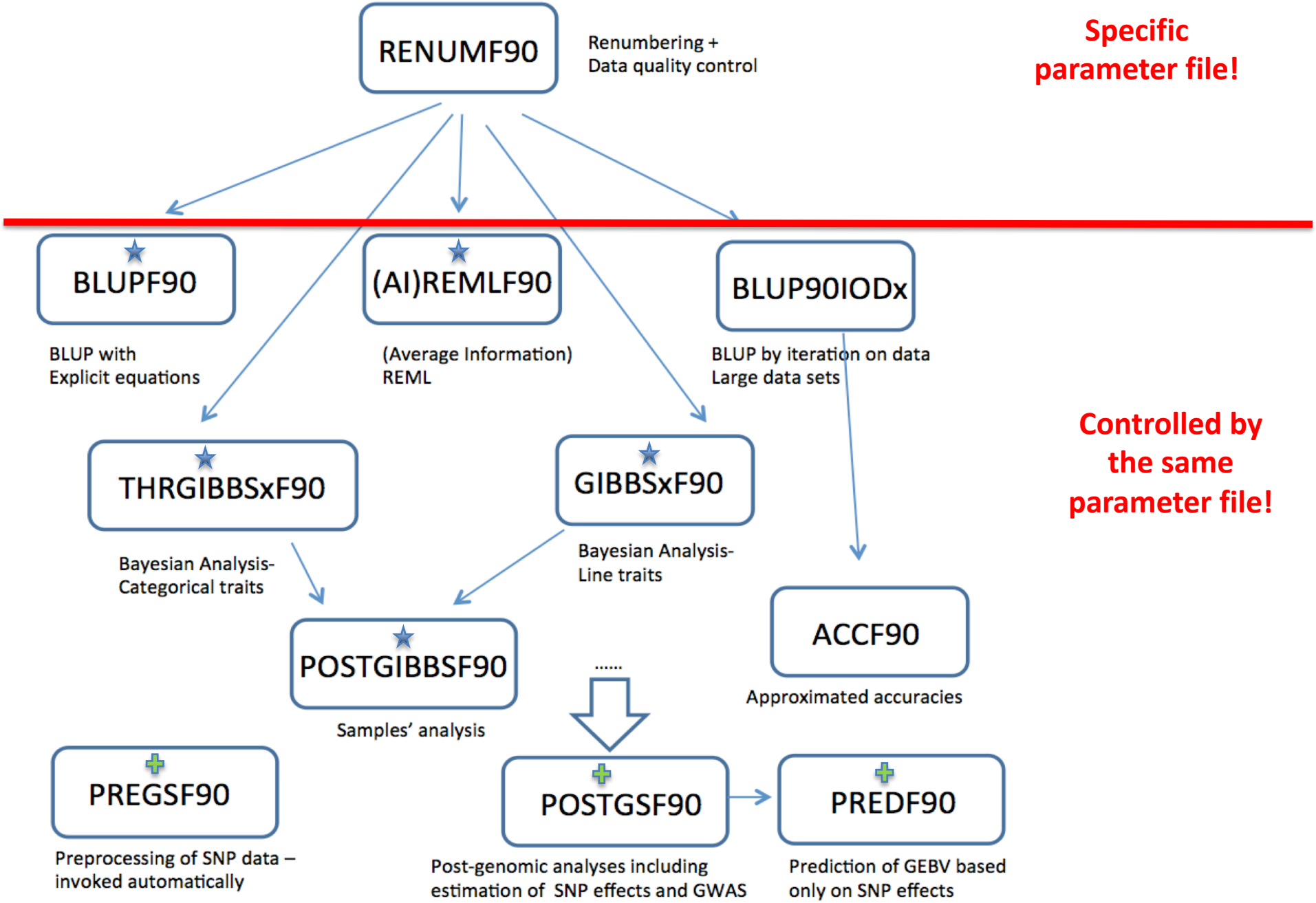
Data File for blupf90 family:

a) Only numbers – Integer or real

a) All effects need to be renumbered from 1 to N



Controlled by the same parameter file!


RENUMF90
Renumbering +
Data quality control
Specific parameter file!
BLUPF90
BLUP with
Explicit equations
(AI)REMLF90
(Average Information)
REML
BLUP90IODx
BLUP by iteration on data
Large data sets
THRGIBBSxF90
Bayesian Analysis-
Categorical traits
GIBBSxF90
Bayesian Analysis-
Line traits
Controlled by the same parameter file!
POSTGIBBSF90
Samples' analysis
.....
ACCF90
Approximated accuracies
PREGSF90
Preprocessing of SNP data –
invoked automatically
POSTGSF90
Post-genomic analyses including
estimation of SNP effects and GWAS
PREDF90
Prediction of GEBV based
only on SNP effects

RENUMF90

A **renumbering program** for the BLUPF90 family of programs

RENUMF90

- **Renumbers data and pedigree**
- **Traces back pedigree for individuals in data file**
- **Performs comprehensive pedigree checking**
- **Provides data statistics**

RENUMF90

- **Supports**
  - virtually any dataset
  - multiple traits
  - different effects per trait
  - alphanumeric and numeric fields
  - unknown parent groups
  - covariates for random regression models

RENUMF90 – Input files

- **Data file and pedigree file as flat files**
  - Columns separated by at least one SPACE
  - No TABS !!!! (current version checks for it)
  - Input files cannot contain character #
  - Missing sire/dams must have code 0
  - code 00 is treated as a known animal

- **Parameter file for renumf90**
  - Has its own parameter file!!!!
  - Different from parameter file for other programs!!!!

RENUMF90 – Output files

- **Create files to be used by BLUPF90 family**
  - renf90.dat - data recoded
  - renaddxx.ped - renumbered pedigree + statistics
  - renf90.tables - cross reference file with renumbered and original effects
  - renf90.par - parameter file

RENUMF90 parameter file

- Based on keywords in capital letter followed by a line(s) with the corresponding data item
- Keywords need to be typed exactly
- Keywords need to be in sequential order!!!
- Lines starting with # are treated as comments and are ignored

RENUMF90 parameter file

MANDATORY

Keyword	possible value	description
DATAFILE	character	The name of data file to be processed
TRAITS	integer	Position for observation (trait) in the data file
FIELDS_PASSED TO OUTPUT	integer	Position for the columns in the original data that will be passed to the renumbered data without changes Keep empty if not needed
WEIGHT(S)	integer	The position(s) for weight in the data file Keep empty if not needed
RESIDUAL_VARIANCE	real value(s)	Residual (co)variance
EFFECT	(next slide …)	Description of an effect Repeatable

RENUMF90 parameter file

Effects

Keyword	Possible value	effect type	form
EFFECT	integer (column where the effect is)	cross	alpha
			numer
		cov	

Keyword (only for covariables)	Possible value	form
NESTED	integer (column where the effect is)	alpha
		numer

RENUMF90 parameter file

Model: $y = herd + sex + \beta\ age + e$

```
DATAFILE
data1.txt
TRAITS
5
FIELDS_PASSED TO OUTPUT
2
WEIGHT(S)

RESIDUAL_VARIANCE
1.0
EFFECT        #1st effect - herd
2 cross alpha
EFFECT        #2nd effect - sex
3 cross numer
EFFECT        #3rd effect - age
4 cov
```

Fixed model

data1.txt

ID	herd	sex	age	phen
ID006	A	1	1.0	3.0
ID009	A	2	1.0	2.0
ID012	A	1	2.0	4.0
ID007	B	2	2.0	6.0
ID010	B	1	1.0	3.0
ID013	B	2	2.0	6.0
ID008	C	1	2.0	6.0
ID011	C	2	1.0	6.0
ID014	C	1	1.0	8.0
ID015	C	2	2.0	4.0

RENUMF90 parameter file

Random Effects

Keyword after EFFECT	possible value	description
RANDOM	diagonal	Non-correlated
	animal	Correlation structure among animals
	sire	Not implemented!!!

Keyword	possible value	description
OPTIONAL	pe	Permanent environmental
	mat	Maternal
	mpe	Permanent environmental maternal (only if mat is used)

RENUMF90 parameter file

Random effects file section

Keyword after RANDOM (animal or sire only)	possible value	description
FILE	character	Name of the pedigree file for animal or sire models only

Keyword after FILE (for RANDOM animal only)	possible value	description
FILE_POS	integer	Specifies positions in the pedigree for ani sire dam alternate_dam yob Default: 1 2 3 0 0 If maternal effect alternate_dam

Keyword (for RANDOM animal only)	possible value	description
SNP_FILE	character	Optional: If genomic info is to be used Name of the SNP file Format: ID 011122211155152222

RENUMF90 parameter file

Pedigree options

Keyword **(for RANDOM animal only)**	**possible value**	**description**
PED_DEPTH	Integer	Optional Specifies the depth of pedigree search Default = 3 All pedigree = 0

Keyword **(for RANDOM animal only)**	**possible value**	**description**
GEN_INT	Integer min avg max	Optional Specifies min, avg, max generation interval; if yob is present Average used to predict yob of parents

Keyword **(for RANDOM animal only)**	**possible value**	**description**
REC_SEX	Integer	Optional Specifies which parent has records Checks if records are found in specific sex

RENUMF90 parameter file

Unknown Parent Group options

Keyword **(for RANDOM animal only)**	**possible value**	**description**
UPG_TYPE		Optional
	yob 1990 1992 …	UPG assigned based on yob
	in_pedigrees	Missing parent receives -x x is the UPG number

RENUMF90 parameter file

Inbreeding option

Keyword **(for RANDOM animal only)**	**possible value**	**description**
INBREEDING		Optional
	pedigree	Calculates inbreeding and saves in the renumbered pedigree file
	file	Reads inbreeding from an external file format: original_ID inbreeding (0 to 1)

RENUMF90 parameter file

Random Regression options

Keyword	possible value	description
RANDOM_REGRESSION		Specifies that random regression should be applied to the random effects
	data	If covariables are in the data

Keyword	possible value	description
RR_POSITON	Integer	Specifies positions of covariables if RANDOM_REGRESSION type is data

RENUMF90 parameter file

(CO)VARIANCES for Random effects

Keyword	possible value	description
(CO)VARIANCES	real	(co)variance for the animal effect dimension should account for number of traits and random correlated effects

32.79	-7.22	-11.07
-7.22	258.06	87.66
-11.07	87.66	194.34

RENUMF90 parameter file

(CO)VARIANCES structure

- 3 trait (T1-T3) and 2 correlated effects (E1-E2)

		E1			E2		
		T1	T2	T3	T1	T2	T3
E1	T1						
	T2						
	T3						
E2	T1						
	T2						
	T3						

RENUMF90 parameter file

(CO)VARIANCES for Random effects

Keyword	possible value	description
(CO)VARIANCES	real	(co)variance for non-correlated random effects

Keyword	possible value	description
(CO)VARIANCES_PE	real	(co)variance for the PE effect if present

Keyword	possible value	description
(CO)VARIANCES_MPE	real	(co)variance for the MPE effect if present

RENUMF90 parameter file

Creating interaction between effects

Keyword	possible value	description
COMBINE	integer	Should come before DATAFILE COMBINE 7 2 3 4 Columns 2, 3, 4 are combined into 7 They can be numer or alpha

RENUMF90 parameter file

Keyword	optional	possible values
COMBINE	optional	definition of new field as a combination of existing fields
DATAFILE	mandatory	name of raw data file
TRAITS	mandatory	positions of observations in the raw data file
FIELDS_PASSED	mandatory	positions of items in the raw data file to be passed to renf90.dat
WEIGHT(S)	mandatory	positions of weights in the raw data file
RESIDUAL_VARIANCE	mandatory	residual covariance matrix
EFFECT	mandatory	effect description
NESTED	optional	positions of nested covatiates
RANDOM	optional	declaration of random effect
OPTIONAL	optional	declaration of MAT, PE, MPE
FILE	optional	name of raw pedigree file
FILE_POS	optional	positions of animal ID, sire ID and dam ID
SNP_FILE	optional	name of SNP marker file
PED_DEPTH	optional	the maximum generation back from animals with phenotype and/or genotype
GEN_INT	optional	generation interval to set unknown parent groups (UPG)
REC_SEX	optional	check if records are found in specific sex
UPG_TYPE	optional	UPG specification
INBREEDING	optional	create pedigree file with inbreeding code
RANDOM_REGRESSION	optional	put covariates for random regressions
RR_POSITION	optional	positions of covariates for random regressions
(CO)VARIANCES	optional	covariance components
(CO)VARIANCES_PE	optional	covariance components for animal PE effects
(CO)VARIANCES_MPE	optional	covariance components for maternal PE effects
OPTION	optional	option parameters

RENUMF90 parameter file

Options passed to blupf90

- All lines that begin with keyword OPTION are passed to parameter file `renf90.par`

- This allows automatization of process by using scripts

- For example:
  - `OPTION sol se`

RENUMF90 parameter file

Extra options

The following options can added at the end of the parameter file to redefine parameters used to read the input file:

– the default size of character fields (maximum number of characters in a column - 20)

```
OPTION alpha_size nn
```

where *nn* is the new size.

– the size of th record length (maximum number of characters in a line - 800)

```
OPTION max_string_readline nn
```

where *nn* is the new size.

– the maximun number of fields (maximum number of columns in a line - 100)

```
OPTION max_field_readline nn
```

where *nn* is the number of fields.

RENUMF90 parameter file

Hints

- Keyword EFFECT is repeated as many time as effects in the model
- If (CO)VARIANCES for any effect are missing, default matrix with 1.0 in diagonal and 0.1 on off-diagonal will be used
  - WARNING: for EM-REML convergence rate is improved if starting values are too large rather than too small !!!

RENUMF90 parameter file

Model: $y = herd + sex + \beta\ age + e$

```
DATAFILE
data1.txt
TRAITS
5
FIELDS_PASSED TO OUTPUT
2
WEIGHT(S)

RESIDUAL_VARIANCE
1.0
EFFECT          #1st effect - herd
2 cross alpha
RANDOM
diagonal
(CO)VARIANCES
0.5
EFFECT          #2nd effect - sex
3 cross numer
EFFECT          #3rd effect - age
4 cov
```

What if we want to consider herd as random?

data1.txt

ID	herd	sex	age	phen
ID006	A	1	1.0	3.0
ID009	A	2	1.0	2.0
ID012	A	1	2.0	4.0
ID007	B	2	2.0	6.0
ID010	B	1	1.0	3.0
ID013	B	2	2.0	6.0
ID008	C	1	2.0	6.0
ID011	C	2	1.0	6.0
ID014	C	1	1.0	8.0
ID015	C	2	2.0	4.0

RENUMF90 parameter file

Model: $y = herd + sex + \beta\ age + animal + e$

```
DATAFILE
data1.txt
TRAITS
5
FIELDS_PASSED TO OUTPUT
2
WEIGHT(S)

RESIDUAL_VARIANCE
1.0
EFFECT        #1st effect - herd
2 cross alpha
EFFECT        #2nd effect - sex
3 cross numer
EFFECT        #3rd effect - age
4 cov
EFFECT        #4th effect - animal
1 cross alpha
RANDOM
animal
FILE
ped1.txt
FILE_POS
1 2 3 0 0
(CO)VARIANCES
0.5
```

What if we want to consider animal effect as random?

ped1.txt

ID	Sire	Dam
ID006	ID001	ID003
ID009	ID001	ID004
ID012	ID001	ID005
ID007	ID001	ID003
ID010	ID001	ID004
ID013	ID002	ID005
ID008	ID002	ID003
ID011	ID002	ID004
ID014	ID002	ID005
ID015	ID002	ID003

data1.txt

ID	herd	sex	age	phen
ID006	A	1	1.0	3.0
ID009	A	2	1.0	2.0
ID012	A	1	2.0	4.0
ID007	B	2	2.0	6.0
ID010	B	1	1.0	3.0
ID013	B	2	2.0	6.0
ID008	C	1	2.0	6.0
ID011	C	2	1.0	6.0
ID014	C	1	1.0	8.0
ID015	C	2	2.0	4.0

RENUMF90 output files

Pedigree file: `renaddxx.ped`

Data file: `renf90.dat`

Parameter file: `renf90.par`

Renumbering table: `renf90.table`

RENUMF90 output files
Pedigree file: `renaddxx.ped`

- Structure:
  1. Animal ID (from 1)
  2. Parent 1 ID or UPG number for parent 1
  3. Parent 2 ID or UPG number for parent 2
  4. 3 minus number of known parents
  5. Known or estimated year of birth (0 if not provided)
  6. Number of known parents, if animal has genotype: 10+number of known parents
  7. Number of records
  8. Number of progeny as parent 1
  9. Number of progeny as parent 2
  10. Original animal ID

RENUMF90 output files
Pedigree file: `renaddxx.ped`

- If option for inbreeding is used:

  Column 4:

  inb/upg code = 4000 / [(1+ms)(1-Fs) + (1+md)(1-Fd)]

  ms (md) is 0 if sire (dam) is known, and 1 otherwise
  Fs (Fd) is the coefficient of inbreeding of sire (dam)

  Ex: For an animal with both parents known and F=0
  inb/upg code = 2000

RENUMF90 output files renumbering tables: `renf90.tables`

- For each cross-classified effects
  - Original ID, count, consecutive number

- Useful
  - To translate solutions from BLUPF90 program into original alphanumeric values
  - Check counts of records by level

Example of RENUMF90 parameter file

```
DATAFILE
data.test
TRAITS
3 4
FIELDS_PASSED TO OUTPUT
2 1 # passing alphanumeric
WEIGHT(S)

RESIDUAL_VARIANCE
5 2
2 4
EFFECT
1 1 cross alpha
EFFECT
2 2 cross alpha
RANDOM
animal
OPTIONAL
mat mpe pe
FILE
test.ped
(CO)VARIANCES
10 3 2 1
3 11 4 5
2 4 12 6
1 5 6 13.01
(CO)VARIANCES_PE
5.3 2.1
2.1 4.85
(CO)VARIANCES_MPE
1.03 .27
.27 .85
EFFECT
5 0 cov
NESTED
 1 0 alpha
EFFECT
6 6 cross alpha
RANDOM
diagonal
```

data file – data.test

```
1 aa 34.5 11 12 zz
3 bb 21.333 22 23 xx
8 cc 23.666 33 34 yy
1 dd 29 44 45 xx
3 aa  30 55 56 yy
5 bb 1234567.890 66 67 zz
```

pedigree file – test.ped

```
qq 0 0
aa 0 0
bb qq aa
cc qq 0
dd 0 aa
```

RENUMF90 Printout

```
RENUMF90 version 1.93
name of parameter file? testpar1
datafile:data.test
traits:           3           4
fields passed:           2           1
R
  5.000       2.000
  2.000       4.000

Processing effect  1 of type cross
item_kind=alpha

Processing effect  2 of type cross
item_kind=alpha
Optional maternal effect
Optional maternal permanent environment
Optional permanent environment
pedigree file name  "test.ped"
positions of animal, sire, dam, alternate dam and yob    1    2    3    0    0
Reading (CO)VARIANCES:           4 x           4
Reading (CO)VARIANCES_PE:           2 x           2
Reading (CO)VARIANCES_MPE:           2 x           2

Processing effect  3 of type cov
item_kind=alpha

Processing effect  4 of type cross
item_kind=alpha

Maximum size of character fields: 20

Maximum size of record (max_string_readline): 800

Maximum number of fields innput file (max_field_readline): 100
```

RENUMF90 Printout

```
hash tables for effects set up
read            6  records
table with            4  elements sorted
added count
Effect group            1  of column            1  with            4  levels
table expanded from        10000  to        10000  records
added count
Effect group            2  of column            1  with            4  levels
table with            4  elements sorted
added count
Effect group            3  of column            1  with            4  levels
table expanded from        10000  to        10000  records
table with            3  elements sorted
added count
Effect group            4  of column            1  with            3  levels
table expanded from        10000  to        10000  records
wrote statistics in file "renf90.tables"

Basic statistics for input data  (missing value code is 0)
Pos  Min         Max         Mean        SD                 N
  3    21.333     0.12346E+07 0.20578E+06 0.50400E+06       6
  4    11.000      66.000      38.500      20.579           6
  5    12.000      67.000      39.500      20.579           6

Correlation matrix
       3     4     5
 3   1.00  0.65  0.65
 4   0.65  1.00  1.00
 5   0.65  1.00  1.00

Counts of nonzero values (order as above)
         6         6         6
         6         6         6
         6         6         6
```

RENUMF90 Printout

```
random effect   2
type:animal
opened output pedigree file "renadd02.ped"
read            5  pedigree records
loaded            3  parent(s) in round            1

Pedigree checks

Number of animals with records:           4
Number of parents without records:           1
Number of phantom dams:           2
Total number of animals:           7

random effect   4
type:diag

Wrote parameter file "renf90.par"
Wrote renumbered data "renf90.dat"
```

RENUMF90 output files

data file – renf90.dat

```
34.5 11 1 3 5 12 1 3 aa 1
21.333 22 2 1 3 23 2 1 bb 3
23.666 33 4 4 7 34 4 2 cc 8
29 44 1 2 3 45 1 1 dd 1
30 55 2 3 5 56 2 2 aa 3
1234567.890 66 3 1 3 67 3 3 bb 5
```

Pedigree file – renadd02.ped

```
1 6 3 1 0 2 2 0 0 bb
6 0 0 1 0 0 0 2 0 qq
2 0 3 1 0 1 1 0 0 dd
7 0 0 1 0 0 0 0 1 D@@0000002
5 0 0 1 0 0 0 0 1 D@@0000001
3 0 5 1 0 1 2 0 2 aa
4 6 7 1 0 2 1 0 0 cc
```

renumbering tables - renf90.tables

```
 Effect group 1 of column 1 with 4 levels
 Value    #    consecutive number
1 2 1
3 2 2
5 1 3
8 1 4
 Effect group 3 of column 1 with 4 levels
 Value    #    consecutive number
1 2 1
3 2 2
5 1 3
8 1 4
 Effect group 4 of column 1 with 3 levels
 Value    #    consecutive number
xx 2 1
yy 2 2
zz 2 3
```

parameter file - testpar1

```
# Parameter file for program renf90; it is translated to parameter
# file for BLUPF90 family programs.
DATAFILE
data.test
TRAITS
3 4
FIELDS_PASSED TO OUTPUT
2 1 # passing alphanumeric
WEIGHT(S)

RESIDUAL_VARIANCE
5 2
2 4
EFFECT
1 1 cross alpha
EFFECT
2 2 cross alpha
RANDOM
animal
OPTIONAL
mat mpe pe
FILE
test.ped
(CO)VARIANCES
10 3 2 1
3 11 4 5
2 4 12 6
1 5 6 13.01
(CO)VARIANCES_PE
5.3 2.1
2.1 4.85
(CO)VARIANCES_MPE
1.03 .27
.27 .85
EFFECT
5 0 cov
NESTED
 1 0 alpha
EFFECT
6 6 cross alpha
RANDOM
diagonal
```

new parameter file - renf90.par

```
# BLUPF90 parameter file created by RENF90
DATAFILE
 renf90.dat
NUMBER_OF_TRAITS
          2
NUMBER_OF_EFFECTS
           7
OBSERVATION(S)
    1    2
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  3  3         4 cross
  4  4         7 cross
  5  5         7 cross
  5  5         7 cross
  4  4         7 cross
  6  0         4 cov   7  0
  8  8         3 cross
RANDOM_RESIDUAL VALUES
   5.000       2.000
   2.000       4.000
 RANDOM_GROUP
     2     3
 RANDOM_TYPE
 add_animal
 FILE
renadd02.ped
(CO)VARIANCES
   10.00       3.000       2.000       1.000
   3.000       11.00       4.000       5.000
   2.000       4.000       12.00       6.000
   1.000       5.000       6.000       13.01
 RANDOM_GROUP
     4
 RANDOM_TYPE
 diagonal
 FILE

(CO)VARIANCES
   1.030      0.2700
  0.2700      0.8500
 RANDOM_GROUP
     5
 RANDOM_TYPE
 diagonal
 FILE

(CO)VARIANCES
   5.300       2.100
   2.100       4.850
 RANDOM_GROUP
     7
 RANDOM_TYPE
 diagonal
 FILE

(CO)VARIANCES
   1.000      0.1000
  0.1000       1.000
```

renumf90 FAQ

1) renumf90 cannot find the data file — Check for typos
2) How to include quadratic covariable? — Column in data file
3) Error when trying to use covariable — 2 cov numer
4) I want to have original IDs in renf90.dat — FIELDS_PASSED TO OUTPUT
5) Fixed effects in renf90.dat are different from original — renf90.tables

renumf90 FAQ

8) When and how to run renumf90?

a) Objective to compare models

Run renumf90 ONCE with the most complete model

Remove effects from renf90.par

b) Objective to compare non-genomic vs genomic model

Run renumf90 ONCE with SNP file

For non-genomic: Remove option for SNP file from renf90.par

c) Objective to mask phenotypes for some animals for validation

Run renumf90 ONCE with the complete data

Remove animals from renf90.dat

blupf90

Mixed Model Equations Solver

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\dfrac{\boldsymbol{\sigma}_e^2}{\boldsymbol{\sigma}_a^2} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'}\mathrm{y} \\ \mathbf{Z'}\mathrm{y} \end{bmatrix}$$

blupf90

- Computes generalized solutions by several methods:
  - Preconditioner Conjugate Gradient (PCG)
    - Default Iterative method (fast)
  - Successive over-relaxation (SOR)
    - an iterative method based on Gauss-Seidel
  - Direct solution using sparse Cholesky factorization
    - FSPAK or YAMS (greater memory requirements)

- The solution values change among methods but estimable function should be the same
- Prediction error variances can be obtained using sparse inverse (FSPAK or YAMS)

Parameter file for blupf90

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
 ../renf90.dat
NUMBER_OF_TRAITS
           2
NUMBER_OF_EFFECTS
           5
OBSERVATION(S)
    1    2
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  3  4     40593 cross
  5  5         2 cross
  6  0         4 cross
  7  0         8 cross
  8  8    918111 cross
RANDOM_RESIDUAL VALUES
   2.5300       1.3425
   1.3425       29.714
RANDOM_GROUP
     5
RANDOM_TYPE
 add_an_upginb
FILE
../renadd05.ped
(CO)VARIANCES
   0.7600       2.2391
   2.2391       30.609
```

Unlimited number of traits and effects

Parameter file for blupf90

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
 ../renf90.dat
NUMBER_OF_TRAITS
            2
NUMBER_OF_EFFECTS
            5
OBSERVATION(S)
    1    2
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  3   4      40593 cross
  5   5          2 cross
  6   0          4 cross
  7   0          8 cross
  8   8     918111 cross
RANDOM_RESIDUAL VALUES
   2.5300        1.3425
   1.3425        29.714
RANDOM_GROUP
     5
RANDOM_TYPE
 add_an_upginb
FILE
../renadd05.ped
(CO)VARIANCES
   0.7600        2.2391
   2.2391        30.609
```

As many columns as the number of traits

Number of levels

Type of effect

- As many rows as the NUMBER_OF_EFFECTS
- Model definition for each trait
- Different models per trait are supported
- If an effect is missing for one trait use 0

Parameter file for blupf90

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
 ../renf90.dat
NUMBER_OF_TRAITS
           2
NUMBER_OF_EFFECTS
           5
OBSERVATION(S)
    1    2
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  3  4     40593 cross
  5  5         2 cross
  6  0         4 cross
  7  0         8 cross
  8  8    918111 cross
RANDOM_RESIDUAL VALUES
   2.5300       1.3425
   1.3425       29.714
RANDOM_GROUP
     5
RANDOM_TYPE
 add_an_upginb
FILE
../renadd05.ped
(CO)VARIANCES
   0.7600       2.2391
   2.2391       30.609
```

Should be a square matrix with dimension equal to the number of traits

- Use zero (0.0) to indicate uncorrelated residual effects between traits
- e.g. For a 3-trait model

43.1	0.0	0.0
0.0	5.1	3.2
0.0	3.2	10.3

Parameter file for blupf90

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
 ../renf90.dat
NUMBER_OF_TRAITS
           2
NUMBER_OF_EFFECTS
           5
OBSERVATION(S)
    1    2
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  3  4     40593 cross
  5  5         2 cross
  6  0         4 cross
  7  0         8 cross
  8  8    918111 cross
RANDOM_RESIDUAL VALUES
   2.5300       1.3425
   1.3425       29.714
RANDOM_GROUP
     5
RANDOM_TYPE
 add_an_upginb
FILE
../renadd05.ped
(CO)VARIANCES
   0.7600       2.2391
   2.2391       30.609
```

Definition of random effects

RANDOM_GROUP
RANDOM_TYPE
FILE
(CO)VARIANCES

Definition of random effects

- RANDOM_GROUP
  - Number(s) of effect from list of effects
  - Correlated effects should be consecutive e.g. Maternal effects, Random Regression
- RANDOM_TYPE
  - diagonal, add_animal, add_sire, add_an_upg, add_an_upginb, user_file, user_file_i, or par_domin
- FILE
  - Pedigree file, parental dominance, or user file
- (CO)VARIANCES
  - Square matrix with dimension equal to the number_of_traits*number_of_correlated_effects

(CO)VARIANCES

- Assuming a 3 trait (T1-T3) and 3 correlated effects (E1-E3)

		E1			E2			E3		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
E1	T1									
	T2									
	T3									
	…..									

RANDOM_TYPE

- *Diagonal*
  - for permanent environment effects
  - assumes no correlation between levels of the effect

- *add_sire*
  - To create a relationship matrix using sire and maternal grandsire
  - Pedigre file:
    - `individual number, sire number, maternal grandsire number`

- *add_animal*
  - To create a relationship matrix using sire and dam information
  - Pedigre file:
    - `animal number, sire number, dam number`

RANDOM_TYPE

- *add_an_upg*
  - As before but using rules for unknown parent group
  - Pedigre file:
    - `animal number, sire number, dam number, parent code`
    - missing sire/dam can be replaced by upg number, usually greater than maximum number of animals
    - Parent code = 3 – # of known parents
      - 1 both parents known
      - 2 one parent known
      - 3 both parents unknown

- *add_an_upginb*
  - As before but using rules for unknown parent group and inbreeding
  - Pedigre file:
    - `animal number, sire number, dam number, inb/upg code`
    - missing sire/dam can be replaced by upg number, usually greater than maximum number of animals
    - inb/upg code = 4000 / [(1+md)(1-Fs) + (1+ms)(1-Fd)]
    - ms (md) is 0 if sire (dam) is known and 1 otherwise
    - Fs(Fs) inbreeding coefficient of the sire (dam)

RANDOM_TYPE

- *user_file*
  - An inverted matrix is read from file
  - Matrix is stored only upper- or lower-triangular
  - Matrix file:
    - `row, col, value`
- *user_file_i*
  - As before but the matrix will be inverted by the program
- *par_domin*
  - A parental dominance file created by program RENDOM

OPTIONS for blupf90

- Program behavior can be modified by adding extra options at the end of the parameter file
- `OPTION option_name x1 x2 …`
- option_name: each program has its own definition of options
- The number of optional parameters (x1, x2…) to control the behavior depends on the option

Options for blupf90

Options

```
OPTION conv_crit 1e-12
```

Set convergence criteria (deault 1e-12).

```
OPTION maxrounds 10000
```

Set maximum number of rounds (default 5000).

```
OPTION solv_method FSPAK
```

Selection solutions by FSPAK, SOR or PCG (default PCG).

```
OPTION r_factor 1.6
```

Set relaxation factor for SOR (default 1.4).

```
OPTION sol se
```

Store solutions and standard errors.

```
OPTION store_pev_pec 6
```

Store triangular matrices of standard errors and its covariances for correlated random effects such as direct-maternal effects and random-regression effects in “pev_pec_bf90”.

Options for blupf90

Missing data
Not pedigree!

```
OPTION missing -999
```

Specify missing observations (default 0) in integer.

```
OPTION residual
```

y-hat and residual will be included in “yhat_residual”.

```
OPTION blksize 3
```

Set block size for preconditioner (default 1).

```
OPTION use_yams
```

Run the program with YAMS (modified FSPAK).

```
OPTION SNP_file snp
```

Specify the SNP file name to use genotype data.

Example of parameter file for blupf90

Single trait "USDA-type" animal model

$$y_{ijkl} = hys_i + hs_{ij} + p_k + a_k + e_{ijkl}$$

where

y_{ijkl} - production yield
hys_i - fixed herd year season
hs_{ij} - random herd x sire interaction
p_k - random permanent environment
a_k - random animal

and

$var(hs_{ij}) = .05$, $var(p_k)=.1$, $var(a_k)=.5$, $var(e_{ijkl})=1$

From blupf90.pdf documentation: http://nce.ads.uga.edu/wiki/doku.php

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
renf90.dat
NUMBER_OF_TRAITS
          1
NUMBER_OF_EFFECTS
          4
OBSERVATION(S)
    1
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
 2          3 cross
 3          6 cross
 4          4 cross
 5         14 cross
RANDOM_RESIDUAL VALUES
   1.0000
 RANDOM_GROUP
     2
 RANDOM_TYPE
 diagonal
 FILE

(CO)VARIANCES
  0.10000
 RANDOM_GROUP
     3
 RANDOM_TYPE
 diagonal
 FILE

(CO)VARIANCES
  0.50000E-01
 RANDOM_GROUP
     4
 RANDOM_TYPE
 add_an_upg
 FILE
renadd04.ped
(CO)VARIANCES
  0.50000
```

Model

$$y_{ijkl} = hys_i + hs_{ij} + p_k + a_k + e_{ijkl}$$

where

y_{ijkl} - production yield
hys_i - fixed herd year season
hs_{ij} - random herd x sire interaction
p_k - random permanent environment
a_k - random animal

and

$var(hs_{ij}) = .05$, $var(p_k)=.1$, $var(a_k)=.5$, $var(e_{ijkl})=1$

renf90.dat

Format: phen/hys/pe/hs/ani

```
1 1 1 1 3
1 1 2 1 6
2 2 3 2 2
3 2 4 3 5
4 3 5 4 1
3 3 6 3 4
```

renadd04.ped

Format: ani/sire/dam/code/……

```
1 12 11 2 0 1 1 1 0 5
2 6 8 1 0 2 1 0 0 3
7 1 11 1 0 2 0 0 3 8
9 10 14 2 0 1 0 0 2 10
3 12 7 2 0 1 1 2 0 1
4 3 9 1 0 2 1 0 0 6
11 13 14 3 0 0 0 0 2 11
5 10 9 1 0 2 1 0 0 4
8 13 7 2 0 1 0 0 1 9
6 3 7 1 0 2 1 1 0 2
10 13 14 3 0 0 0 2 0 7
```

Output from blupf90

```
 name of parameter file?
renf90.par
     BLUPF90 ver. 1.66


 Parameter file:              renf90.par
 Data file:                   renf90.dat
 Number of Traits              1
 Number of Effects             4
 Position of Observations       1
 Position of Weight (1)         0
 Value of Missing Trait/Observation           0


EFFECTS
 #  type                 position
(2)        levels   [positions for nested]
    1  cross-classified       2          3
    2  cross-classified       3          6
    3  cross-classified       4          4
    4  cross-classified       5         14
```

Parameter file

data file

Number of levels for each effect

```
 read            6  records in   6.1703999E-02  s,
nonzeroes
  read           11  additive pedigrees
 finished peds in   6.1760999E-02  s,
round =     1  convergence =  0.1435
round =     2  convergence =  0.3000E-01
round =     3  convergence =  0.1714E-02
round =     4  convergence =  0.2914E-03
round =     5  convergence =  0.1207E-03
round =     6  convergence =  0.1278E-03
round =     7  convergence =  0.1085E-03
round =     8  convergence =  0.1405E-03
round =     9  convergence =  0.1904E-03
round =    10  convergence =  0.1119E-03
round =    11  convergence =  0.1562E-04
round =    12  convergence =  0.6141E-05
round =    13  convergence =  0.4609E-05
round =    14  convergence =  0.1750E-04
round =    15  convergence =  0.8708E-04
round =    16  convergence =  0.2800E-03
round =    17  convergence =  0.1286E-04
round =    18  convergence =  0.2309E-06
round =    19  convergence =  0.2566E-08
round =    20  convergence =  0.1131E-09
round =    21  convergence =  0.2413E-12
   21 iterations,   convergence criterion= 0.2413E-12
 solutions stored in file: "solutions"
```

Records read from data file

Records read from pedigree file

Solutions file

File *"solutions"*

$$y_{ijkl} = hys_i + hs_{ij} + p_k + a_k + e_{ijkl}$$

where

y_{ijkl} - production yield

hys_i - fixed herd year season

hs_{ij} - random herd x sire interaction

p_k - random permanent environment

a_k - random animal

Parameter File

```
DATAFILE
 renf90.dat
NUMBER_OF_TRAITS
           1
NUMBER_OF_EFFECTS
           4
OBSERVATION(S)
    1
WEIGHT(S)


EFFECTS: POSITIONS_IN_DATAFILE
 2          3 cross
 3          6 cross
 4          4 cross
 5         14 cross
```

trait/effect	level	solution
1	1	0.49585171
1	2	2.52240030
1	3	2.91017217
1	1	-0.00965409
1	2	0.00965406
1	3	-0.01930794
1	4	0.01930787
1	5	0.03861622
1	6	-0.03861599
1	1	-0.00000001
1	2	-0.00965384
1	3	-0.00965406
1	4	0.01930810
1	1	0.64574095
1	2	-0.30035705
1	3	0.61034316
1	4	0.52426082
1	5	0.27486415
1	6	0.39795337
1	7	0.18556405
1	8	-0.90212661
1	9	0.63126003
1	10	-0.17807241
1	11	-0.17807202
1	12	1.25233833
1	13	-1.91741245
1	14	1.51299821

Common problem in blupf90

- Wrong data file and pedigree name
  - Program may not stop if file name does not exist
  - Check outputs for data file name and number of records and pedigree read

```
round =  4995  convergence =        NaN
round =  4996  convergence =        NaN
round =  4997  convergence =        NaN
round =  4998  convergence =        NaN
round =  4999  convergence =        NaN
round =  5000  convergence =        NaN
 5001 iterations,   convergence criterion=       NaN
 solutions stored in file: "solutions"
```

blupf90 FAQ

1) Why solutions are not with original ID?

ask Ignacy Misztal

2) How to match solutions with original ID?

Bash scripting

For trait 1 when animal is effect number 2

```
$awk '{ if ($1==1 && $2==2) print $3,$4}' solutions | sort +0 -1 > sol.temp
$awk '{print $1,$10}' renadd02.ped | sort +0 -1 > ids.temp
$join -1 +1 -2 +1 ids.temp sol.temp > renumbered_original_EBV
```

remlf90 and airemlf90

Variance components estimation

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\dfrac{\boldsymbol{\sigma}_e^2}{\boldsymbol{\sigma}_a^2} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'}\mathrm{y} \\ \mathbf{Z'}\mathrm{y} \end{bmatrix}$$

remlf90 and airemlf90

- REML = restricted/residual maximum likelihood
  - Patterson and Thompson (1971)

- Most used method to estimate variance components in breeding and genetics

- BLUPF90 family has 2 REML programs
  - remlf90: expectation-maximization (EM) algorithm
  - airemlf90: average information (AI) algorithm

- remlf90 and airemlf90 use the same parameter file as blupf90

remlf90

$$\hat{\sigma}_a^2 = \frac{\hat{\mathbf{u}}'\mathbf{A}^{-1}\hat{\mathbf{u}} + \mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{C}^{uu}\right)}{N_a}$$

$$\hat{\sigma}_e^2 = \frac{\mathbf{y}'(\mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\mathbf{u}})}{N - \mathrm{rank}(\mathbf{X})}$$

Inverse of LHS for animal effect

animals (rank of A)

- The equations contain BLUE and BLUP but those values are calculated with known variance components

- This method requires iterations:

1. set initial variance components
2. compute $\hat{\beta}$ and $\hat{u}$ via mixed model equations
3. update and variance components with above equations
4. go to 1 or stop if the parameters do not change any more

Instructions from Masuda (2018)

remlf90

- Simpler equations
- Easier to understand
- More complicated equations in multiple-trait models
- Very slow convergence (looks stable but may not converge)
- Computationally demanding especially for **C^{uu}**

$$\begin{bmatrix}\hat{\beta}\\ \hat{u}\end{bmatrix} = \begin{bmatrix}\mathbf{X'X} & \mathbf{X'Z}\\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\dfrac{\sigma_e^2}{\sigma_a^2}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X'}y\\ \mathbf{Z'}y\end{bmatrix}$$

airemlf90

Vector of variance components

$$\theta_{n+1} = \theta_n - \mathbf{H}^{-1}(\theta_n)\mathbf{d}(\theta_n)$$

Gradient (score vector)

Hessian Matrix

Average-information algorithm uses this matrix as Hessian,

$$\mathbf{H}(\theta) = \mathcal{I}_A(\theta) = \begin{bmatrix} -\frac{1}{2}\mathbf{y}'\mathbf{PZAZ}'\mathbf{PZAZ}'\mathbf{Py} & -\frac{1}{2}\mathbf{y}'\mathbf{PZAZ}'\mathbf{PPy} \\ -\frac{1}{2}\mathbf{y}'\mathbf{PPZAZ}'\mathbf{Py} & -\frac{1}{2}\mathbf{y}'\mathbf{PPPy} \end{bmatrix}$$

Gradient

expensive

$$-2\mathbf{d}(\theta) = \begin{bmatrix} \mathrm{tr}(\mathbf{PZAZ}') - \mathbf{y}'\mathbf{PZAZ}'\mathbf{Py} \\ \mathrm{tr}(\mathbf{P}) - \mathbf{y}'\mathbf{PPy} \end{bmatrix} = \begin{bmatrix} \frac{N_a}{\sigma_a^2} - \frac{\mathrm{tr}(\mathbf{A}^{-1}\mathbf{C}^{uu})}{(\sigma_a^2)^2} - \frac{\hat{\mathbf{u}}'\mathbf{A}^{-1}\hat{\mathbf{u}}}{(\sigma_a^2)^2} \\ \frac{N-\mathrm{rank}(\mathbf{X})}{\sigma_e^2} - \frac{1}{\sigma_e^2}\left[N_a - \frac{\mathrm{tr}(\mathbf{A}^{-1}\mathbf{C}^{uu})}{\sigma_a^2}\right] - \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{(\sigma_e^2)^2} \end{bmatrix}$$

Instructions from Masuda (2018)

airemlf90

- Much faster than EM-REML
- Provide estimation of standard errors
- BUT
  - For complex models and poor starting values
    - Slow convergence
    - Parameters estimates out of the parameter space
  - In some cases initial rounds with EM-REML help

Options for remlf90 and airemlf90

```
OPTION conv_crit 1d-12
```

Convergence criterion (default 1d-10).

```
OPTION maxrounds 10000
```

Maximum rounds (default 5000).

```
OPTION sol se
```

Store solutions and se.

```
OPTION residual
```

y-hat and residuals will be included in "yhat_residual".

```
OPTION missing -999
```

Specify missing observations (default 0) in integer.

```
OPTION use_yams
```

Run the program with YAMS (modified FSPAK). The computing time can be dramatically improved.

```
OPTION constant_var 5 1 2
```

5: effect number
1: first trait number
2: second trait number
implying the covariance between traits 1 and 2 for effect 5.

Options for airemlf90

```
OPTION EM-REML 10
```

Run EM-REML (REMLF90) for first 10 rounds to get initial variances within the parameter space (default 0).

```
OPTION tol 1d-12
```

Tolerance (or precision) (default 1d-14) for positive definite matrix and g-inverse subroutines.
Convergence may be much faster by changing this value.

```
OPTION store_pev_pec 6
```

Store triangular matrices of standard errors and its covariances for correlated random effects such as direct-maternal effects and random-regression effects in "pev_pec_bf90".

Heterogeneous residual variances for a single trait

```
OPTION hetres_pos 10 11
```

Specify the column positions of (two) covariables in the data file.

```
OPTION hetres_pol 4.0 0.1 0.1
```

Initial values of coefficients for heterogeneous residual variances using *ln*(a0, a1, a2, …) to make these values.
To transform back to the original scale, use exp(a0+a1*X1+a2*X2)

log-residual function (Foulley and Quaas, 1995)

Options for airemlf90

```
OPTION se_covar_function <label> <function>
```

`<label>`
A name for a particular function (e.g., `P1` for phenotypic variance of trait 1, `H2_1` for heritability for trait 1, `rg12` for genetic correlation between traits 1 and 2, ...).

`<function>`
A formula to calculate a function of (co)variances to estimate SD. All terms of the function should be written with no spaces.

Each term of the function corresponds to (co)variance elements and could include any random effects (G) and residual (R) (co)variances.

`G_eff1_eff2_trt1_trt2`

`R_trt1_trt1`

Examples:

```
OPTION se_covar_function P G_2_2_1_1+G_2_3_1_1+G_3_3_1_1+G_4_4_1_1+R_1_1
```

```
OPTION se_covar_function H2d G_2_2_1_1/(G_2_2_1_1+G_2_3_1_1+G_3_3_1_1+G_4_4_1_1+R_1_1)
```

```
OPTION se_covar_function rg12 G_2_2_1_2/(G_2_2_1_1*G_2_2_2_2)**0.5
```

Does reml always converge?

- When the expected variance is very small or the covariance matrix is close to non-positive definite, try the following starting values:
  - much smaller = 0.00001
  - much bigger = 1000

- If AIREMLF90 does not converge but REMLF90 converges with the same data set and the same model:
  - run REMLF90 again but with a small starting value to check the estimate because it could be artifact

  - use an option to use EM-REML inside AI-REML:

    `OPTION EM-REML xx`

    where xx is the number of rounds of EM

gibbsf90

Bayes Theorem

Likelihood function

indicates how likely the observations are from a distribution (with particular parameters)

$$p(\theta|y) = p(y|\theta)\, p(\theta)$$

prior probability of unknown θ

posterior probability of unknown θ with known y

- Basic idea of Gibbs sampling:
- Gibbs sampling is a numerical method to draw samples from a posterior distribution (not always explicitly available)
- Draw samples = generate random numbers following a distribution
- The results are random numbers (not theoretical formulas)
- The posterior distribution will be drawn based on the numerical values (like a histogram)

Instructions from Masuda (2018)

gibbsf90

Ingredients for Gibbs sampling

1) Theoretical derivation: conditional posterior distribution for each unknown parameter

2) Software: a random number generator for a particular distribution

```
# Basic Gibbs sampling for mu (normal) and sigma2 (inverted chi-square)
y <- c(14,16,18)
N <- length(y)
n.samples <- 100
mu <- rep(0,n.samples)
sigma2 <- rep(0,n.samples)

# initial value
mu[1] <- 0
sigma2[1] <- 10

# sampling
for(i in 2:n.samples){
    mu[i] <- rnorm(1, mean=mean(y), sd=sqrt(sigma2[i-1]/N)) # using the most recent sigma2
    df <- N-2
    S <- sum((y-mu[i])^2)
    sigma2[i] <- rinvchisq(1, df=df, scale=S) # using the most recent mu
}
```

Instructions from Masuda (2018)

gibbsXf90

- gibbs1f90: faster for multiple trait models
- gibbs2f90: better for correlated random effects
- gibbs3f90: for heterogeneous residual variance

- Name of parameter file?

  gibbs1.par
- Number of samples and length of burn-in?

  samples=10,000 to 100,000; burn-in=0
- Give n to store every n-th sample?

  10

```
gibbs1f90 gibbs1.par --rounds 10000 --burnin 0 --thin 10
```

gibbsXf90

- Procedure
  - Run gibbsXf90 to estimate variance components
  - Run postgibbsf90 to process the samples and verify convergence
  - Run gibbsXf90 with new variance components to estimate breeding values (2k to 10k rounds)

gibbsXf90

```
OPTION fixed_var all 1 2 3
```

All solutions and posterior means and SD for effects for effects1, 2, and 3 are stored in “all_solutions” and in “final_solutions” every round using fixed variances. Without numbers, all solutions for all effects are stored.

```
OPTION fixed_var mean 1 2 3
```

Posterior means and SD for effects1, 2, and 3 in “final_solutions”.

```
OPTION solution all 1 2 3
```

Caution: this option will create a huge output solution file when you run many rounds and/or use a large model. All solutions and posterior means and SD for effects1, 2, and 3 are stored in “all_solutions” and in “final_solutions” every round. Without numbers, all solutions for all effects are stored.

```
OPTION solution mean 1 2 3
```

Caution: this option will create a huge output solution file when you run many rounds and/or use a large model. Posterior means and SD for effects1, 2, and 3 in “final_solutions”.

```
OPTION cont 10000
```

10000 is the number of samples run previously when restarting the program from the last run.

gibbsXf90

```
OPTION prior 5 2 -1 5
```

The (co)variance priors are specified in the parameter file.
Degree of belief for all random effects should be specified using the following structure:
OPTION prior eff1 db1 eff2 db2 … effn dbn -1 dbres
effx correspond to the effect number and dbx to the degree of belief for this random effect, -1 corresponds to the degree of belief of the residual variance.
In this example 2 is the degree of belief for the 5th effect, and 5 is the degree of belief for the residual.

```
OPTION seed 123 321
```

Two seeds for a random number generator can be specified.

```
OPTION SNP_file snp
```

Specify the SNP file name to use genotype data.

```
OPTION se_covar_function <label> <function>
```

gibbs3f90

```
OPTION hetres_int col nlev
```

```
OPTION hetres_int 5 10
```

The position “5” to identify the interval in the data file and the number of intervals “10” for heterogeneous residual variances.

gibbs3f90

Data (datasire)

```
1 - HYS
2 - sire
3 - y1
4 - heterogeneous clas
5 - y2
```

cat datasire

```
6 13 317.55 1 644.26
3 10 280.44 1 563.05
....................
37 1 270.52 5 543.63
53 10 286.43 5 579.84
```

Parameter file (ex5)

```
DATAFILE
datasire
NUMBER_OF_TRAITS
NUMBER_OF_EFFECTS
OBSERVATION(S)
WEIGHT(S)
EFFECTS: POSITIONS_IN_DATAFILE
1 1 100 cross
2 2 50 cross
RANDOM_RESIDUAL VALUES
500 100
100 1000
RANDOM_GROUP
RANDOM_TYPE
diagonal
FILE
(CO)VARIANCES
75 10
10 150
OPTION hetres_int 4 5
```

```
round   98
 209.      416.
 416.      828.
Residual variance, interval   1
df_r  1997 ee/n  99.4738134864675
 101.      202.
 202.      412.
Residual variance, interval   2
df_r  1997 ee/n  146.518188769043
 148.      296.
 296.      602.
Residual variance, interval   3
df_r  1997 ee/n  198.183671561078
 198.      397.
 397.      806.
Residual variance, interval   4
df_r  1997 ee/n  232.307903786663
 228.      455.
 455.      917.
Residual variance, interval   5
df_r  1997 ee/n  301.189371418363
 311.      622.
 622.     0.126E+04
```

thrgibbsXf90

- thrgibbs1f90: Gibbs sampler for mixed threshold-linear models

  Thresholds and variances can be estimated or assumed
- thrgibbs3f90: for heterogeneous residual variance

- Name of parameter file?

  gibbs1.par
- Number of samples and length of burn-in?

  samples=10,000 to 100,000; burn-in=0
- Give n to store every n-th sample?

  10

```
thrgibbs1f90 gibbs1.par --rounds 10000 --burnin 0 --thin 10
```

thrgibbsXf90

Options

```
OPTION cat 0 0 2 5
```

"0" indicate that the first and second traits are linear. "2" and "5" indicate that the third and fourth traits are categorical with 2 (binary) and 5 categories.

```
OPTION save_halfway_samples 5000
```

The program saves every "5000" samples to restart or recover the job right after the last saved samples. It is useful when the program accidentally stopped.

To restart, add `OPTION cont 1` to your parameter file and run `thrgibbs1f90` again

```
OPTION thresholds 0.0 1.0 2.0
```

Set the fixed thresholds. No need to set 0 for binary traits.

```
OPTION residual 1
```

The residual variance can be set to 1 but not necessary for categorical traits more than 2 categories. For binary traits, the residual variance is automatically set to 1, so no need to use this option.

postgibbsf90

- Basic idea of post Gibbs analysis:
- Summarize and visualize the samples drawn by gibbsXf90
- Confirm if the chain converged
- Find the most probable value = posterior mode as a “point estimate”
- Find the reliability of the estimates = the highest posterior density as a “confidence interval”

Instructions from Masuda (2018)

postgibbsf90

- Name of parameter file?
  gibbs1.par
- Burn-in?
  0
- Give n to store every n-th sample? (1 means read all samples)
  10

- input files
  gibbs_samples, fort.99
- output files
  "postgibbs_samples"
    all Gibbs samples for additional post analyses
  "postmean"
    posterior means
  "postsd"
    posterior standard deviations
  "postout"

postgibbsf90

at least > 10 is recommended
> 30 may be better

number of independent cycles of Gibbs samples

******** Monte Carlo Error by Time Series ********

Pos.	eff1	eff2	trt1	trt2	MCE	Mean	HPD Interval (95%)		Effective sample size	Median	Mode	Independent chain size
1	4	4	1	1	1.362E-02	0.9889	0.7788	1.215	70.4	0.9844	0.9861	18
2	4	4	1	2	1.288E-02	1.006	0.777	1.219	84.1	1.006	0.952	18
3	4	4	2	2	1.847E-02	1.66	1.347	1.987	80.3	1.652	1.579	25
4	0	0	1	1	9.530E-03	24.47	24.07	24.84	425.6	24.47	24.53	2
5	0	0	1	2	8.253E-03	11.84	11.54	12.18	395.8	11.83	11.82	2
6	0	0	2	2	1.233E-02	30.1	29.65	30.58	387.8	30.09	29.97	5

******** P

Lower and upper bounds of Mean ± 1.96PSD

io

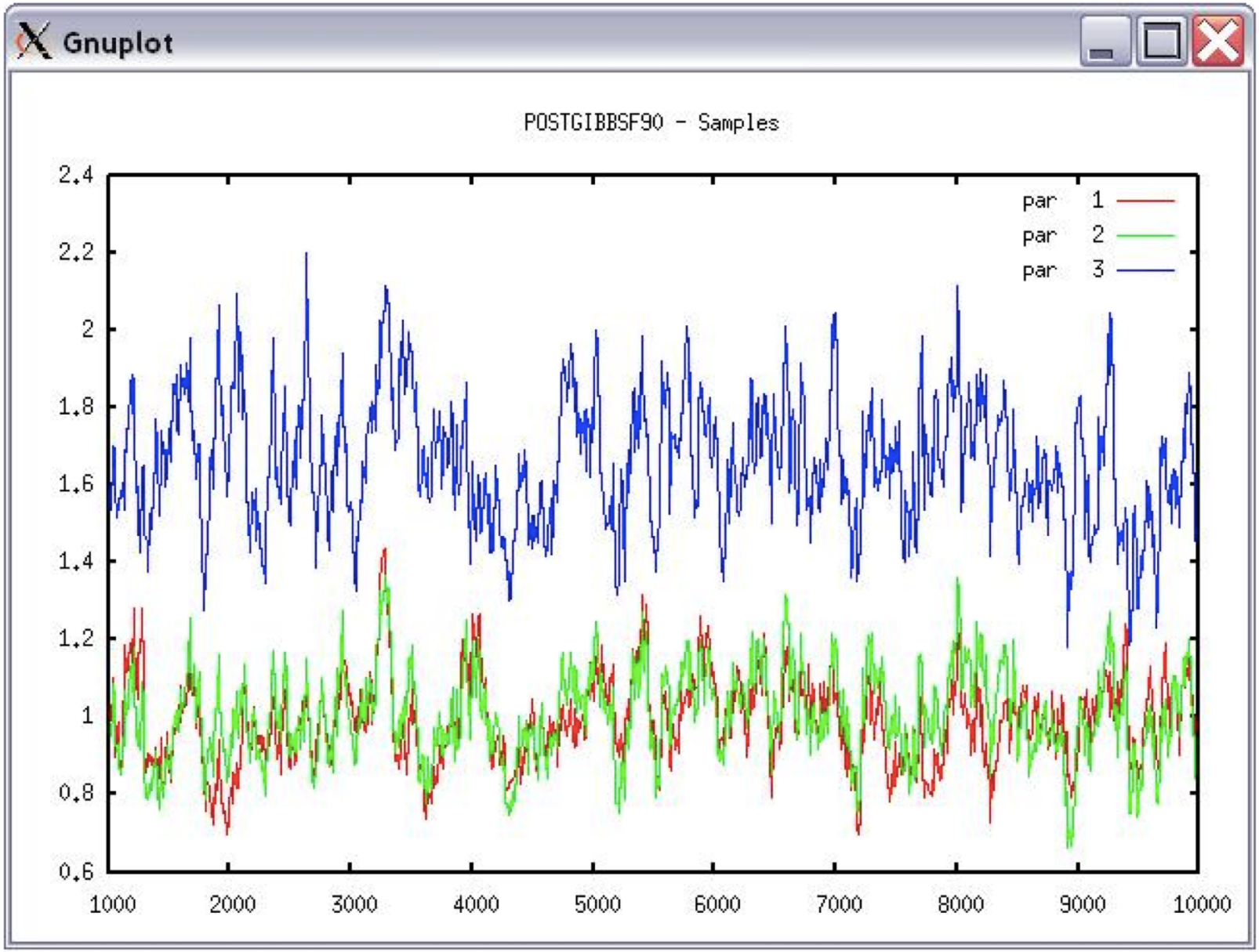
ratio between first half and second half of the samples ; should be < 1.0

Pos.	eff1	eff2	trt1	trt2	PSD	Mean	PSD Interval (95%)		Geweke diagnostic	Autocorrelations lag: 1	10	50	Independent # batches
1	4	4	1	1	0.1144	0.9889	0.7648	1.213	-0.02	0.853	0.188	0.049	50
2	4	4	1	2	0.1182	1.006	0.7742	1.237	-0.11	0.828	0.111	-0.066	50
3	4	4	2	2	0.1656	1.66	1.335	1.984	0.06	0.828	0.108	-0.021	36
4	0	0	1	1	0.1967	24.47	24.09	24.86	-0.01	0.034	0.029	-0.062	450
5	0	0	1	2	0.1643	11.84	11.51	12.16	0.03	0.032	-0.006	-0.016	450
6	0	0	2	2	0.2429	30.1	29.62	30.57	-0.02	0.07	-0.014	0.037	180

postgibbsf90

```
Choose a graph for samples (= 1) or histogram (= 2); or exit (= 0)
1

positions

1 2 3 # choose from the position numbers 1 through 6

If the graph is stable (not increasing or decreasing), the convergence is met.
All samples before that point should be discarded as burn-in.

print = 1; other graphs = 2; or stop = 0
2
```

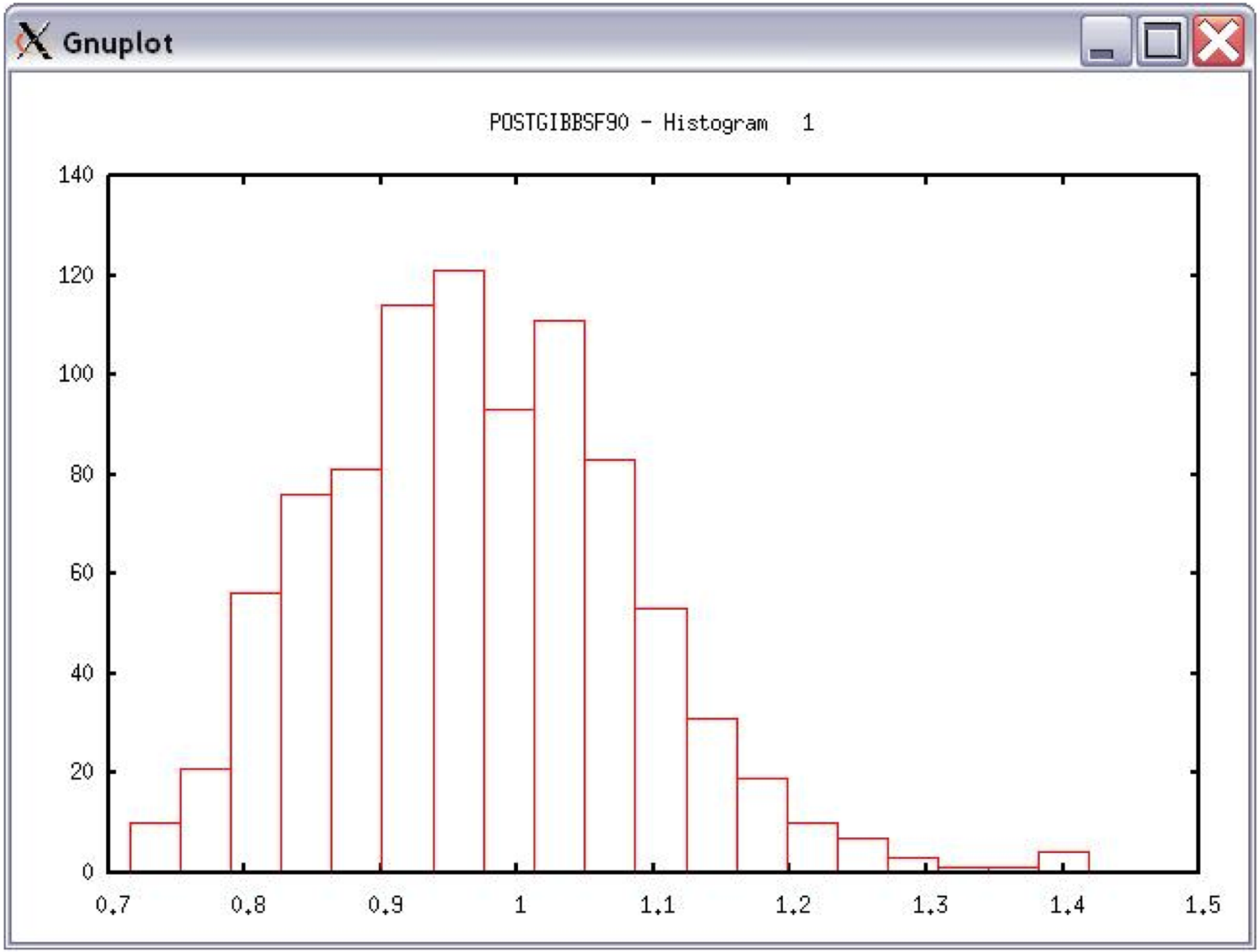
postgibbsf90



postgibbsf90

```
Choose a graph for samples (= 1) or histogram (= 2); or exit (= 0)
2

Type position and # bins
1 20
```

postgibbsf90



Common problems for blupf90 family

- Wrong position or formats for observation and effects
- Misspelling of Keywords
  - Program may stop
- (Co)variance matrices not symmetric, not positive definite
  - Program may not stop
- Large numbers (e.g. 305-day milk yield 10,000 kg)
  - Scale down i.e. 10,000 /1,000 = 10

General output form blupf90 family

- Output printed on the screen is not saved to any file!
- Should use redirection or pipes to store output

renumf90

```
echo renum.par | renumf90 | tee renum.log
```

blupf90

```
echo renf90.par | blupf90 | tee blup.log
```

airemlf90

```
echo renf90.par | remlf90 | tee airemlf.log
```

Run in background + Save output

```
$vi ai.sh
#type the following commands inside ai.sh
        blupf90 <<AA > blup.log
        renf90.par
        AA
#save and exit
$bash ai.sh &    #can replace bash by sh
```

```
$vi gibbs.sh
#type the following commands inside ai.sh
        gibbs2f90 <<AA > gibbs.log
        renf90.par
        1000 0
        10
        AA
#save and exit
$bash gibbs.sh &  #can replace bash by sh
```