



UNIVERSITY OF
GEORGIA

College of Agricultural &
Environmental Sciences

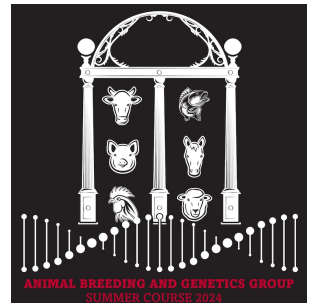
*Animal Breeding and
Genetics Group*

Introduction to RENUMF90

Daniela Lourenco

Ignacio Aguilar

BLUPF90 TEAM – 05/2024

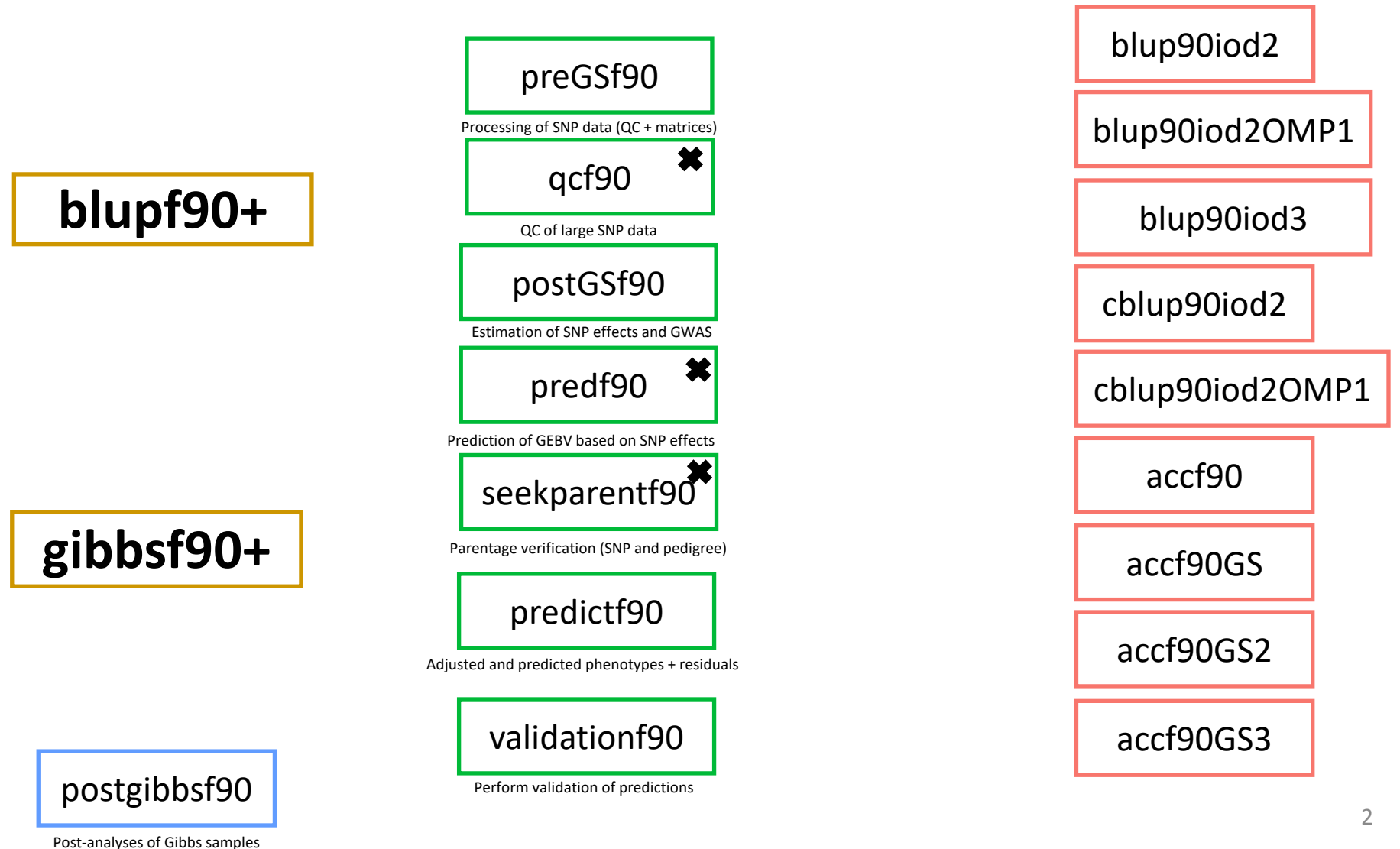


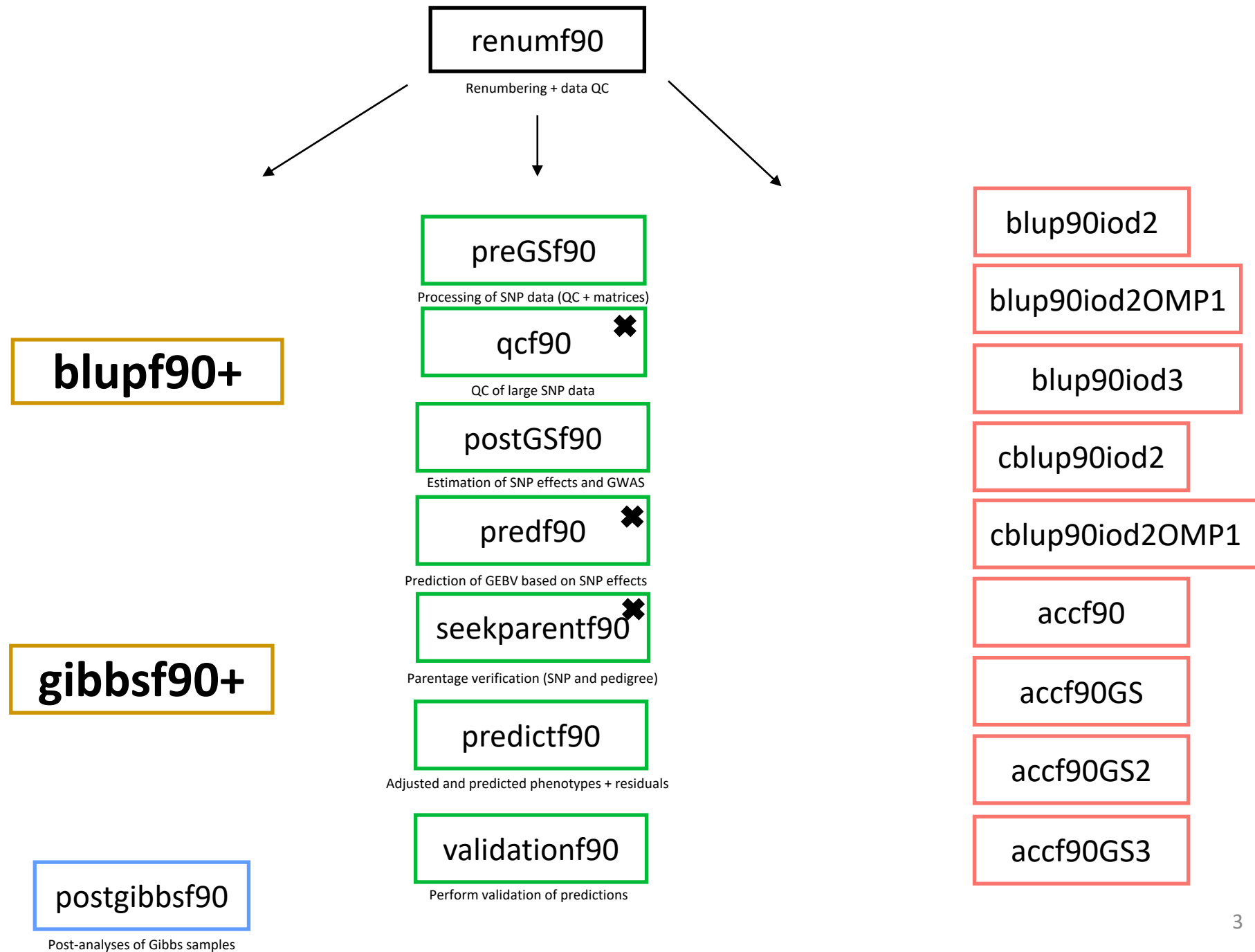
1) Controlled by the same parameter file

2) Data File for blupf90 family:

a) Only numbers – Integer or real

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





RENUMF90

**The renumbering software for the
BLUPF90 suite**

RENUMF90

- **Renumbers data and pedigree**
- **Creates a parameter file for BLUPF90 family**
 - **Parameter file can be modified by the users for new models**
- **Traces back pedigree for individuals in the data**
- **Performs comprehensive pedigree checks**
- **Provides data statistics**
- **Creates an Xref file for genotyped individuals**
- **Computes inbreeding by default in $v \geq 1.157$**



RENUMF90

- **Supports**

- virtually any datasets
- multiple traits
- different models (effects) per trait
- alphanumeric and numeric fields
- unknown parent groups
- covariates for random regression models

RENUMF90 – Input files

- **Parameter file with specific keywords**
- **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

RENUMF90 – Output files

- **Creates files to be used by BLUPF90 family**
 - renf90.dat - renumbered data
 - renaddxx.ped - renumbered pedigree + statistics
 - renf90.par - new parameter file

 - renf90.inb - file with inbreeding
 - renf90.tables - cross reference file with renumbered and original effects
 - renf90.fields - description of the effects in each field of renf90.dat

RENUMF90 parameter file

MANDATORY

Keyword	possible value	description
DATAFILE	character	The name of data file to be processed
TRAITS	integer	Position for phenotype (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 – 1 for each effect in the model

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: $y = farm + sex + \beta age + e$

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

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

Fixed linear model

data1.txt

ID	farm	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

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 only)	possible value	description
FILE	character	Name of the pedigree file for animal 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 <i>If maternal effect alternate_dam</i>

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
	group_unisex	UPG based on the information in pedigree Ex. UPG by breed FILE_POS 1 2 3 0 0 4 #the 6th field indicates which column the UPG code is in the pedigree
	group_sex	Separate UPG code for unknown sire and dam FILE_POS 1 2 3 0 0 4 5 #the 6 th and 7 th fields indicate which columns the UPG codes are in the pedigree

RENUMF90 parameter file

Inbreeding option

Keyword (for RANDOM animal only)	possible value	description
INBREEDING	pedigree	Default in RENUMF90 \geq v1.157 Calculates inbreeding code and saves it in the renumbered pedigree file (Default in RENUMF90 \geq v1.157)
	File <name>	Reads inbreeding from an external file format: original_ID inbreeding (0 to 1)
	self \times	Calculates inbreeding with selfing \times is the column in the pedigree file with the number of selfing generations
	no-inbreeding	Turn inbreeding calculation off in RENUMF90 \geq v1.157

Inbreeding methods in renumf90

- Six methods
 - Speed up depends on the pedigree structure (depth)

OPTION inbreeding_method *n*  # method number

- 1: Meuwissen and Luo (1992)
- 2: Modified Meuwissen & Luo by Sargolzaei & Iwaisaki (2004)
- 3: Modified Colleau by Sargolzaei et al. (2005)
- 4: Recursive tabular
- 5: Tier (1990) with groups (not finished yet)
- 6: Hybrid parallel computing

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
	legendre	If the program will compute Legendre polynomials

Keyword	possible value	description
RR_POSITON	Integer	Specifies positions of covariables in the data if RANDOM_REGRESSION type is data
	r1 ... rq	Positions of covariables if “data”
	r1 r2	r1 is the order of the polynomial and r2 is the position of covariable if “legendre”

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 among 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 covariates
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

- If the data and pedigree files have header

```
#Parameter file for renumf90
DATAFILE
data.txt
SKIP_HEADER
1
TRAITS
3
FIELDS_PASSED TO OUTPUT
1 #Line_ID
WEIGHT(S)

RESIDUAL_VARIANCE
1.0
EFFECT
2 cross alpha
EFFECT
1 cross alpha
RANDOM
animal
FILE
ped.txt
SKIP_HEADER
1
(CO)VARIANCES
1.0
OPTION sol se
```


RENUMF90 options

Options

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

- the default size of character fields (default = 20)

```
OPTION alpha_size nn
```

where *nn* is the new size.

- the size of the record length (default = 800)

```
OPTION max_string_readline nn
```

where *nn* is the new size.

- the maximum number of fields (default = 100)

```
OPTION max_field_readline nn
```

where *nn* is the number of fields.

```
OPTION missing x
```

allows indicating that the missing value is the number *x* (e.g., 999), for instance, if 0 is a valid record. This is only to represent the missing value in the data. If there are covariables in the data, 0 is treated as a value, not missing information. Missing pedigree is always 0 and cannot be changed to another value.

```
OPTION remove_all_missing
```

removes lines in the data where phenotypes are missing. Keeping those lines may cause unexpected behavior in some programs.

RENUMF90 options

```
OPTION missing_in_weights
```

in addition, this indicates that if a weight for the trait is 0, then the value of the trait is converted to "missing" in the output file `renf90.dat`, i.e. 0 by default or another value is set if `OPTION missing` is used.

```
OPTION no_basic_statistics
```

avoids the computation of basic statistics (min, max, correlations, ...), which take a certain time for very large data file.

```
OPTION inbreeding_method m
```

allows choosing a method for inbreeding calculation. The inbreeding coefficients are used later (in the other programs) to set up the coefficients for the A-inverse. Acceptable values for m are:

- 1: Meuwissen and Luo (1992)
- 2: Modified Meuwissen & Luo by Sargolzaei & Iwaisaki (2004)
- 3: Modified Colleau by Sargolzaei et al. (2005)
- 4: recursive tabular method
- 5: method of Tier (1990)
- 6: Hybrid parallel computing, which is basically a parallel (OMP) version of Meuwissen and Luo (1992)
- 7: Recursive tabular with self-breeding generations. For populations with selfing, i.e., wheat

The default is method 1. Large speed-ups are made using method 6, but this requires using several threads (e.g., using `OMP_NUM_THREADS=4`)

RENUMF90 options

Run renumf90 with the following option:

```
OPTION animal_order genotypes
```

- Optimizing gibbsf90+ when using genomic data
 - Run gibbsf90+ with the following option:

```
OPTION separate_dense
```

RENUMF90 parameter file

Hints

- Keyword EFFECT is repeated as many times 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

RENUMF90 parameter file

Options passed to blupf90

- All lines that begin with the keyword `OPTION` are passed to the parameter file `renf90.par`
 - Unless they are specific to `renumf90`
- This allows automation of process by using scripts
- For example:
 - `OPTION sol se`
 - `OPTION use_yams`

RENUMF90 parameter file

Model: $y = \text{farm} + \text{sex} + \beta \text{ age} + e$

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

RESIDUAL_VARIANCE
1.0
EFFECT      #1st effect - farm
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 farm as random
with variance = 0.5?*

data1.txt

ID	farm	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: $y = \text{farm} + \text{sex} + \beta \text{ age} + \text{animal} + e$

```

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

RESIDUAL_VARIANCE
1.0
EFFECT      #1st effect - farm
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
0.2
    
```

What if we want to consider animal effect as random with $\sigma_u^2 = 0.2$?

ped1.txt			data1.txt				
ID	Sire	Dam	ID	farm	sex	age	phen
ID006	ID001	ID003	ID006	A	1	1.0	3.0
ID009	ID001	ID004	ID009	A	2	1.0	2.0
ID012	ID001	ID005	ID012	A	1	2.0	4.0
ID007	ID001	ID003	ID007	B	2	2.0	6.0
ID010	ID001	ID004	ID010	B	1	1.0	3.0
ID013	ID002	ID005	ID013	B	2	2.0	6.0
ID008	ID002	ID003	ID008	C	1	2.0	6.0
ID011	ID002	ID004	ID011	C	2	1.0	6.0
ID014	ID002	ID005	ID014	C	1	1.0	8.0
ID015	ID002	ID003	ID015	C	2	2.0	4.0

RENUMF90 output files

Pedigree file: `renaddxx.ped`

Data file: `renf90.dat`

Parameter file: `renf90.par`

Inbreeding file: `renf90.inb`

Renumbering table: `renf90.table`

Fields table: `renf90.fields`

RENUMF90 output files

Pedigree file: `renaddxx.ped`

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 genotyped: 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`

- As inbreeding is default:

Column 4:

$$\text{inb/upg code} = 4000 / [(1+m_s)(1-F_s) + (1+m_d)(1-F_d)]$$

m_s (m_d) is 0 if sire (dam) is known, and 1 otherwise

F_s (F_d) is the coefficient of inbreeding of sire (dam)

Ex: For an animal with both parents known and $F=0$

$$\text{inb/upg code} = 2000$$

RENUMF90 output files

Data file: `renf90.dat`

1. Phenotype

2. Effect 1

3. Effect 2

.... ...

N. Effect n

RENUMF90 output files

parameter file: renf90.par

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

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  2          2 cross
  3          12010 cross
RANDOM_RESIDUAL_VALUES
  0.60000
RANDOM_GROUP
  2
RANDOM_TYPE
  add_an_upginb
FILE
  renadd02.ped
(CO)VARIANCES
  0.40000
OPTION SNP_file genotypes.txt
OPTION map_file gen_map.txt
```

```
DATAFILE
  phenotypes.txt
TRAITS
  3
FIELDS_PASSED TO OUTPUT

WEIGHT(S)

RESIDUAL_VARIANCE
  0.60
EFFECT
  2 cross alpha #sex
EFFECT
  1 cross alpha
RANDOM
  animal
FILE
  pedigree.txt
FILE_POS
  1 2 3 0 0
SNP_FILE
  genotypes.txt
PED_DEPTH
  4
(CO)VARIANCES
  0.40
OPTION map_file gen_map.txt
```

RENUMF90 output files

Inbreeding file: `renf90.inb`

origID	Inbreeding	newID
A71342462	0.059204	6927175
A17194772	0.032106	29
A13476873	0.002958	6550405
A1ZEP4813	0.000000	61
A14347077	0.019187	6550336
A64547711	0.026603	12
A71922414	0.000000	6942899
A17274771	0.019961	42
A53301967	0.000000	6550416
A4ZGF7566	0.000000	167
A3ZZS6645	0.000000	25
A07818367	0.000000	7117564
A17354770	0.050361	55
A53401908	0.000000	31
A13556872	0.063467	6550439
A14507075	0.071151	6550347

RENUMF90 – default inbreeding in A^{-1}

Computed using Henderson-Quaas' algorithm **with inbreeding**

before now

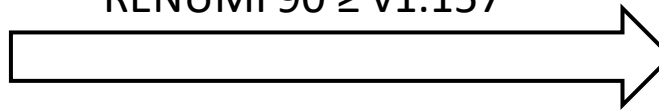
Computed using VanRaden's formula, which considers inbreeding

Computed using Colleau's algorithm, which considers inbreeding

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - \mathbf{X} A_{22}^{-1} \end{bmatrix}$$

RENUMF90 \geq v1.157



```
DATAFILE
phenotypes.txt
TRAITS
3
FIELDS_PASSED TO OUTPUT
WEIGHT(S)
RESIDUAL_VARIANCE
0.60
EFFECT
2 cross alpha
EFFECT
1 cross alpha
RANDOM
animal
FILE
pedigree.txt
FILE_POS
1 2 3 0 0
SNP_FILE
genotypes.txt
PED_DEPTH
0
INBREEDING
pedigree
(CO)VARIANCES
0.40
```

```
DATAFILE
phenotypes.txt
TRAITS
3
FIELDS_PASSED TO OUTPUT
WEIGHT(S)
RESIDUAL_VARIANCE
0.60
EFFECT
2 cross alpha
EFFECT
1 cross alpha
RANDOM
animal
FILE
pedigree.txt
FILE_POS
1 2 3 0 0
SNP_FILE
genotypes.txt
PED_DEPTH
0
(CO)VARIANCES
0.40
```

RENUMF90 output files

Renumbering table: `renf90.tables`

```
Effect group 1 of column 1 with 2 levels, effect # 1
  Value      #      consecutive number
1 5034 1
2 4966 2
```

RENUMF90 output files

Fields table: `renf90.fields`

field	variable	origfield	group	column	random	effect	file
1	trait	3	0	0	*	cov	*
2	renumbered	2	1	1	*	cross	*
3	renumbered	1	2	1	animal	cross	renadd02.ped

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) Fixed effects in renf90.dat are different from original

renf90.tables

5) I want to have original IDs in renf90.dat

FIELDS_PASSED TO OUTPUT

6) I used renumf90 but want to have solutions with original IDs when using blupf90+

renumf90 FAQ

- Storing solutions with original ID if renumf90 was used to renumber the data

```
OPTION origID
```

May not work with
some programs
and options

- Only *solutions.original* is created

- Storing reliabilities with original ID

```
OPTION store_accuracy X orig
```

Number of animal effect

$$Rel = 1 - \frac{PEV}{\sigma_u^2(1 + f)}$$

- Saves *acc_bf90* with renumbered and original ID

- Storing solutions and rel with original ID if renumf90 was used to renumber the data

- Combine the two options

renumf90 FAQ

6) 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

renumf90 quick trick

- `renumf90 --help`
- `renumf90 --show-template`

```
[dani@dodo2 day13]$ renumf90 --help
RENUMF90 version 1.158 with zlib

renumf90 parameter-file [--options ...]

--version          show version number
--show-template    show template parameter file
[dani@dodo2 day13]$ renumf90 --show-template
# parameter file for renumf90
DATAFILE

TRAITS

FIELDS_PASSED TO OUTPUT

WEIGHT(S)

RESIDUAL_VARIANCE

EFFECT

#RANDOM
#
#OPTIONAL
#
#FILE
#
#FILE_POS
#
#SNP_FILE
#
#PED_DEPTH
#
#UPG_TYPE
#
#INBREEDING
#
#FIXED_REGRESSION
#
#RANDOM_REGRESSION
#
#RR_POSITION
#
#(CO)VARIANCES
#
#(CO)VARIANCES_PE
#
#(CO)VARIANCES_MPE
#
#OPTION alpha_size 20
#OPTION max_string_readline 800
#OPTION max_field_readline 100
```

General output from RENUMF90

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

renumf90

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

```
$vi renum.sh
```

```
#type the following commands inside renum.sh
```

```
    renumf90 <<AA > renum.log
```

```
    renum.par
```

```
    AA
```

```
#save and exit
```

```
$bash renum.sh & #can replace bash with sh
```