RENUMF90 Cheat Sheet

The RENUMF90 parameter file

Pairs of *keyword-value* must appear in the following order. The keyword should be capital. More than one field are allowed for a multiple-trait case. Characters led by `#` will be ignored as comments.

Optional `COMBINED` keyword

`COMBINE`
`n a b ..` Fields a, b, ... combied into a single field n.

Required keywords

`DATAFILE`	Data file with observations and effects.
`filename`	
`TRAITS`	Field(s) for observations.
`f1 ..`	Multiple values if multiple-trait.
`FIELDS_PASSED TO OUTPUT`	Field(s) passed to output data.
`f1 ..`	(Empty value if not needed.)
`WEIGHT(S)`	Field for weight.
`field`	(Empty value if not needed.)
`RESIDUAL_VARIANCE`	Full residual covariance matrix.
`R`	(Scalar if single-trait)
`EFFECT`	Effect definition.
`f1 .. type form`	Repeat the keyword-value if needed.

A `EFFECT` block has the following values.

`f1 ..`	Field(s) with class code or covariate. Multiple values if multiple-trait. `0` if not needed for specific traits.
`type`	Type of effect. `cross` for cross-classified effect. `cov` for covariate.
`form`	Only for cross-classified effect. `alpha` for alphanumeric fields. `numer` for numeric fields.

Optional `NESTED` keyword

`NESTED`	Nested regression for the immediate effect.
`f1 .. form`	The same number of fields as `EFFECT`. `form` is `alpha` or `numer`.

Optional `RANDOM` and related keywords

`RANDOM`	Make the immediate effect random.
`type`	`animal` for $\mathbf{A}$; `diag` for $\mathbf{I}$.
`OPTIONAL`	Add optional random effects.
`type ..`	`pe` for permanent environmental effect; `mat` for maternal genetic effect; `mpe` for maternal PE.
`FILE`	Pedigree file.
`filename`	
`FILE_POS`	Field definition of pedigree file.
`a s d ad yob g`	`a`,`s`,`d` for animal, sire and dam ID; `ad` for alternate dam; `0` if not needed; `yob` for birth year; `0` if not needed; `g` for unknown parent groups (optional) (Default = `1 2 3 0 0`)
`SNP_FILE`	SNP marker file.
`filename`	
`PED_DEPTH`	Depth of pedigree search.
`n`	(Default = 3)
`GEN_INT`	Generation interval.
`min avg max`	minimum, average and maximum interval. Needed only if birth year is available.
`REC_SEX`	Check sex-limited traits.
`x`	`1` for male; `2` for female.
`UPG_TYPE`	Assign unknown parent groups.
`type`	`yob` = birth year; `in_pedigrees` = negative code in pedigree; `group` based = group code in extra field; `group_unisex` as above but with "unisex"
`INBREEDING`	Consider inbreeding in $\mathbf{A}^{-1}$.
`type`	`pedigree` computing with pedigree data. `file filename` reading values from file. (Default = no inbreeding considered)
`RANDOM_REGRESSION`	Define this effect as random regression.
`type`	`data` here.
`RR_POSITION`	Field(s) of covariates.
`f1 ..`	
`(CO)VARIANCES`	Full covariance matrix.
`G`	(Default: `1` on diag. and `0.1` on off-diag.)
`(CO)VARIANCES_PE`	Full covariance matrix for PE.
`G`	(Default: `1` on diag. and `0.1` on off-diag.)
`(CO)VARIANCES_PE`	Full covariance matrix for maternal PE.
`G`	(Default: `1` on diag. and `0.1` on off-diag.)

For 2-trait maternal model, the genetic covariance matrix `(CO)VARIANCES` should be 4×4 as follows.

		Direct		Maternal	
		Trait 1	Trait 2	Trait 1	Trait 2
Direct	Trait 1				
	Trait 2				
Maternal	Trait 1				
	Trait 2				

For 2-trait random regressions, the genetic covariance matrix `(CO)VARIANCES` should be 4×4 as follows.

		RR 1		RR 2	
		Trait 1	Trait 2	Trait 1	Trait 2
RR 1	Trait 1				
	Trait 2				
RR 2	Trait 1				
	Trait 2				

Additional `OPTION` lines

You can write additional options at the end of the parameter file. An option has the keyword `OPTION` and its values on the same line. Any numbers of option lines are allowed. Only the following options will be taken with RENUMF90; the other options will be just passed to the output file.

`alpha_size n`	The width of alphanumeric field. (Default=20).
`max_string_readline n`	The number of characters in a line. (Default=800).
`max_string_readline n`	The number of fields. (Default=99).

Guideline for file preparation

- Text file with tidy data like a table (each row for reacord and each field for factor).
- White spaces as the only separators; no tabs are allowed.
- Alphanumeric characters and symbols for group code and animal ID.
- Common numerics (integer, floating point, and exponential expressions) for observations, covariates, and weights.
- The default missing code is `0`. You can change it using `OPTION missing n` with an integer n; This works only for data file (not for pedigree file in which a single `0` is the missing code).

Output files

The RENUMF90 program generates "renumbered" files that can be used with BLUPF90 and related programs.

`renf90.par`	New parameter file for BLUPF90 programs.
`renf90.dat`	New data file.
`renadd??.ped`	New pedigree file. `??` replaced with numbers.
`****_XrefID`	Cross-reference ID file (optional). `****` replaced with the SNP file name.
`renf90.inb`	Inbreeding coefficients (optional).
`renf90.tables`	Code replacement table.

The new pedigree file has 10 fields with additional information.

1. New animal ID (renumbered from 1)
2. New parent 1 (sire) or unknown parent group ID
3. New parent 2 (dam) or unknown parent group ID
4. Without inbreeding, 3 minus number of known parents; With inbreeding a 4-digit code (see below).
5. Known or estimated year of birth (`0` if not provided)
6. The number of known parents (parents might be eliminated if not contributing; if animal has genotype 10+number of know parents
7. The number of records
8. The number of progeny (before elimination due to other effects) as parent 1
9. The number of progeny (before elimination due to other effects) as parent 2
10. Original animal id

The 4th field has the following four-digit code (*upg/inb code*) when inbreeding is included:

$$\frac{4000}{(1+m_s)(1-F_s)+(1+m_d)(1-F_d)}$$

where m_s (m_d) is 0 if sire (dam) is known or 1 if the parent is unknown, and F_s (F_d) is the inbreeding coefficient of sire (dam).

Based on the BLUPF90 manual
Yutaka Masuda (masuday@uga.edu)

BLUPF90 Cheat Sheet

The BLUPF90 parameter file

The parameter file can be commonly used with BLUPF90, AIREMLF90, GIBBSF90, and the other family programs. Pairs of *keyword-value* must appear in the following order. The keyword should be capital. Characters led by `#` will be ignored as comments.

Required keywords

`DATAFILE`	Data file with observations and effects.
`filename`	
`NUMBER_OF_TRAITS`	
`n`	a single integer.
`NUMBER_OF_EFFECTS`	
`n`	a single integer.
`OBSERVATION(S)`	Field(s) for observations.
`f1 ..`	Multiple fields if multiple-trait model.
`WEIGHT(S)`	Field for weight.
`field`	(Empty value if not needed.)
`EFFECTS:`	Effect definition (see below).
`f1 .. type c1 ..`	Repeat the description if needed.
`RANDOM_RESIDUAL`	Full residual covariance matrix.
`R`	(Scalar if single-trait)

The `EFFECTS:` block is followed by model-description lines; each row describes 1 effect so that the number of rows is the same to the number of effects specified at `NUMBER_OF_EFFECTS`.

`f1 ..`	Field(s) with class code or covariate. Multiple values if multiple-trait. `0` if not needed for specific traits.
`type`	Type of effect. `cross` for cross-classified effect. `cov` for covariate.
`c1 ..`	The list is optional. Field(s) with class code for nested regression. Multiple values if multiple-trait. `0` if not needed for specific traits.

The position of effect in this block is called "effect number". It will be used to specify a random effect in the next section.

Optional `RANDOM` and related keywords

`RANDOM` is followed by 3 other keywords and it makes a section to define a random effect. You can repeat the `RANDOM` section if you have several random effects. This section can define a correlated random effect involving multiple effects such as a direct-maternal genetic effect and random-regressions.

`RANDOM_GROUP`	Define a random effect group.
`e1 ..`	List of effect numbers defined above. The effect numbers must be consecutive.
`RANDOM_TYPE`	Type of random effect.
`type`	See below
`FILE`	Pedigree (or similar) file.
`filename`	Empty line if not needed.
`(CO)VARIANCES`	Full covariance matrix.
`G`	

The `type` defines the covariance structure and pedigree relationships.

- `diagonal`: Identity (**I**).
- `add_sire`: Numerator relationship matrix for sire and MGS.
- `add_animal`: Numerator relationship matrix without inbreeding.
- `add_an_upg`: As above with unknown parent groups.
- `add_an_upginb`: As above but with inbreeding.
- `par_domin`: Parental dominance.
- `user_file`: User-supplied inverse matrix.
- `user_file_inv`: User-supplied non-inverse matrix (inverted by programs).

The `filename` will be needed for all `type` except `diagonal`. In a pedigree file, an animal's ID should be a positive integer; an unknown-parent group is an integer greater than the largest ID of real animals; a missing parent is `0`. The file format is shown below.

- For `add_sire`: 1) animal, 2) sire, and 3) MGS.
- For `add_animal`: 1) animal, 2) sire, and 3) dam.
- For `add_an_upg`: 1) animal, 2) sire or UPG, 3) dam or UPG, and 4) 3 minus the number of known parents.
- For `add_an_upginb`: 1) animal, 2) sire or UPG, 3) dam or UPG, and 4) four-digit upg/inb code; see RENUMF90.
- For `par_domin`: Generated with `rendomn`; See the manual.
- For `user_file` and `user_file_inv`: 1) row, 2) column, and 3) value; Half-stored.

In a covariance matrix, the trait is nested within effect. See the following case for 2 traits and 2 correlated effects.

		Eff 1		Eff 2	
		Tr 1	Tr 2	Tr 1	Tr 2
Eff 1	Tr 1				
	Tr 2				
Eff 2	Tr 1				
	Tr 2				

Options

You can write additional options at the end of the parameter file. An option has the keyword `OPTION` and its values on the same line. Any numbers of option lines are allowed. Unsupported options will be simply ignored (but preGSf90 give you an error).

Genomic options

For genomic analyses, see a separate cheat sheet.

Common options for BLUPF90/AIREMLF90

`missing n`	Treat an integer `n` as a missing observation; default = `0`.
`conv_crit c`	Convergence criterion for iteratiions; default = 10^{-12}.
`maxrounds n`	Maximum iterations; default = 5000.
`sol se`	Calculate SE of each solution as the inverse of LHS.
`use_yams`	Faster computations with the YAMS package; should be combined with `solv_method FSPAK` for BLUPF90.

BLUPF90

`solv_method m`	Solving method: `m` = `FSPAK` for direct inversion and `pcg` for PCG (default)

AIREMLF90

`EM-REML n`	EM iterations for the first `n` rounds.
`hetres_pos f1 ..`	Fields for covariates in a function of heterogeneous residual variance; should be multiple of the number of traits.
`hetres_pol f1 ..`	Initial regression coeffeicients for the heterogeneous-residual-variance function.

`se_covar_function label function`
Calculate SE for a function of variance components by sampling; shown with arbitrary `label`; covariance `G_i_j_k_l` for random effects i and j, and traits k and l; residual covariance `R_k_l` for trait k and l.

Common options for GIBBSxF90/THRGIBBS1F90

`fixed_var mean e1 ..`	Compute poeterior mean/SD of location parameters of specified effects without updating covariances.
`fixed_var all e1 ..`	As above but store all samples.
`solution mean e1 ..`	Similar to `fixed_var mean` but updating variance components.
`solution all e1 ..`	Similar to `fixed_var all` but updating variance components.
`cont n`	Continue sampling from the previous run in the round `n`.
`seed m n`	Seeds of random number generators.

THRGIBBS1F90

`cat t1 ..`	The number of categories in each trait; `0` for contineoues traits.
`censored t1 ..`	Censoring in each trait; `1` if censored and `0` if not.
`threshold v1 ..`	Set fixed thresholds; default = 0.
`residual 1`	Set residual variance to 1.

Based on the BLUPF90 manual
Yutaka Masuda (masuday@uga.edu)

Genomic Options Cheat Sheet

Flowchart

All application programs can check the genomic data and calculate a genomic relationship matrix ($\mathbf{G}$), a subset of a pedigree matrix ($\mathbf{A}_{22}$), and those inverse matrices, followed by the statistical computations (e.g. solving equations in BLUPF90). PREGSF90 performs the genomic set-up only. The genomic data will be processed as follows.

1. Check the cross-reference ID (XrefID) file.
2. Read the pedigree and store it in memory.
3. Calculate $\mathbf{A}_{22}$.
4. Read and store the SNP markers in memory.
5. Check the quality of markers and animals and remove some of them if unqualified (*quality control*).
6. Compute $\mathbf{Z}$ as the adjusted marker genotypes with allele frequency.
7. Calculate $\mathbf{G} = \mathbf{ZZ}'/k$ with a coefficient k.
8. Update $\mathbf{G}$ as $\mathbf{G} \leftarrow \alpha\mathbf{G} + \beta\mathbf{A}_{22} + \gamma\mathbf{I} + \delta\mathbf{11}'$ (*blending*).
9. Update $\mathbf{G}$ to scale it to $\mathbf{A}_{22}$ (*tuning*).
10. Calculate statistics on $\mathbf{G}$ and $\mathbf{A}_{22}$.
11. Calculate $\omega\mathbf{A}_{22}^{-1}$ by updating $\mathbf{A}_{22}$.
12. Calculate $\tau\mathbf{G}^{-1}$ by updating $\mathbf{G}$.
13. Calculate the difference $\Delta = \tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$.
14. Save Δ in a binary file (`GimA22i`).

Files

Input files

- SNP file: 2 fields per row: an animal ID and its genotypes. The ID must have a fixed width with the tailing spaces. The genotypes can be integers (coded as `0`, `1`, `2`, and `5` as missing) or gene content (real numbers with fixed width). No spaces are allowed between markers.
- XrefID file: The file is usually generated with RENUMF90. This file has a table relating the genotyped animals with the renumbered pedigree.
- Pedigree file: It is the same as used in the standard animal-model analysis.
- Map file (optional): The file has at least 3 fields per row: 1) the marker number, 2) the chromosome number, 3) the physical location, and optional 4) the description of this markers.

Default output files

- `freqdata.count`: Minor allel frequency calculated from the original SNP file.
- `freqdata.count.after.clean`: Minor allel frequency calculated from the data after the quality control.
- `Gen_call_rate`: Call rate for genotyped animals.
- `Gen_conflicts`: Report of parentage checks.
- `sum2pq`: $2\sum_i p_i q_i$; k as above in $\mathbf{G}$.
- `GimA22i`: A binary fole for $\Delta = \tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$.

The required options for genomics

`SNP_file snpfile xrefid`
Invoke genomic module using the SNP file `snpfile`; By default, a cross-reference-ID (XrefID) file is assumed to be `snpfile` + `_XrefID`. You can optionally supply the XrefID file as the second argument. This option accompanies many other options (shown below).

Genomic options

The following lists are not complete. See the official manual for additional options.

User-supplied files

`chrinfo file`	Supply a map file.
`FreqFile file`	Supply the pre-calculated allele frequency; the same format as `freq.count`

Quality control

`no_quality_control`	Turn off the quality-control; some checks will be still performed but any unqualified data will not be removed.
`saveCleanSNPs`	Save "clean" SNP data, in which unqualified markers and animals have been removed, to files.
`minfreq x`	Remove a marker if the minor allele frequency is $< x$. (default = 0.05)
`callrate x`	Remove a marker if the call rate is $< x$. (default = 0.90)
`callrateAnim x`	Remove an animal if the call rate is $< x$. (default = 0.90)
`monomorphic x`	Remove a monomorphic marker if `x` is 1. (default = 1)
`hwe x`	Perfrom the Hardy-Weinberg test with the criterion `x` (default = not performed).
`high_correlation x y`	Check a high-correlated pair of markers if the difference in the allele frequency between the markers is larger than `x`; show warings if the correlation is higher than `y`. Default = `x=0.025` and `y=0.995`.
`verify_parentage x`	Parentage checks; `0` for skipping all checks; `1` for just checks; `2` for checking animals and removing conflicted markers and animals (default = 2).
`outparent_progeny`	Create a precise report of parentage checks.
`excludeCHR n1..`	Exclude markers on specific chromosomes from the final output; the map file in seeded.
`sex_chr n`	Exclude markers on the sex chromosomes temporarily from parantage and Hardy-Weinberg checks; the map file in seeded.

Blending and tuning

`AlphaBeta a b`	Specify α as `a` and β as `b` in blending (default: $\alpha = 0.95$ and $\beta = 0.05$).
`GammaDelta g d`	Specify γ as `g` and δ as `d` in blending (default: $\gamma = 0$ and $\delta = 0$).
`TauOmega t o`	Specify τ as `t` and ω as `o` for the inverse matrices (default: $\tau = 1$ and $\omega = 1$).

Saving matrices

`saveAscii`	All files will be saved as the text file; Without this option, the files will be saved in a binary format.
`saveG`	Save the final $\mathbf{G}$ in the file `G`.
`saveG all`	Save the all intermediate $\mathbf{G}$'s in several files.
`saveA22`	Save the final $\mathbf{A}_{22}$ in the file `A22`.
`saveGInverse`	Save the final $\tau\mathbf{G}^{-1}$ in the file `Gi`.
`saveA22Inverse`	Save the final $\omega\mathbf{A}_{22}^{-1}$ in the file `A22i`.
`saveGOrig`	Save the final $\omega\mathbf{G}$ with the original animal ID; always saved in the ASCII format regardless of `saveAscii`; the pedigree file generated with RENUMF90.
`saveA22Orig`	As above but for the final $\mathbf{A}_{22}$.
`saveHinv`	Save $\mathbf{H}^{-1}$ in a text file. Only accepted by PREGSF90.
`saveHinvOrig`	Save $\mathbf{H}^{-1}$ in a text file with the original animal ID. Only accepted by PREGSF90.

Reading matrices

With one of the following options, the program will skip all required operations to form the relationship matrix related to the specified option. The file used here should be a binary format (not ASCII).

`readGimA22i <file>`	Read $\Delta = \tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$ (default file = `GimA22i`).
`readG <file>`	Read $\mathbf{G}$ (default = `G`).
`readA22 <file>`	Read $\mathbf{A}_{22}$ (default = `A22`).
`readGInverse <file>`	Read $\mathbf{G}^{-1}$ (default = `Gi`).
`readA22Inverse <file>`	Read $\mathbf{A}_{22}^{-1}$ (default = `A22i`).

The last 2 options assume the file contains $\mathbf{G}^{-1}$ or $\mathbf{A}_{22}^{-1}$, NOT $\tau\mathbf{G}^{-1}$ or $\omega\mathbf{A}_{22}^{-1}$. If you read these matrices and also specifies `TauOmega`, the program will apply τ and ω to the matrices just read from the files. This is problematic when the matrices have been already scaled with τ and ω before being saved.

Skip creating matrices

`createGimA22i 0`	Omit $\Delta = \tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$.
`createG 0`	Omit $\mathbf{G}$.
`createA22 0`	Omit $\mathbf{A}_{22}$.
`createGInverse 0`	Omit $\mathbf{G}^{-1}$.
`createA22Inverse 0`	Omit $\mathbf{A}_{22}^{-1}$.

Based on the BLUPF90 manual
Yutaka Masuda (masuday@uga.edu)