

GSEVM v.4-INRA

(Genetical Structured Environmental Variance Model)

User Guide

M, Garcia[‡], N. Ibáñez-Escriche[†] and D. Sorensen[‡]

[‡]Station d'amélioration génétique des animaux, INRA, 31326 Castanet-Tolosan cedex, France.

[†]Genètica i Millora Animal, IRTA, 25198 Lleida, Spain.

[‡]Danish Institute of Agricultural Sciences, Dept. of Animal Breeding and Genetics, PB50, 8830, Tjele.
milagros.garcia@toulouse.inra.fr; Noelia.Ibanez@irta.cat

Last modification: January 4, 2012

1 Introduction

The GSEVM is a program for analysing data with heterogeneous variances. This model links the phenotype of a given animal with two sets of systematic (fixed effects in frequentist approach) and random effects which contribute to the mean and to a function of the environmental variance. It has the following structure:

$$Y = X\beta + Za + f(\tilde{X}\tilde{\beta}, Z\tilde{a})\epsilon$$

where $X\beta$ and a represent the systematic and random factors which have an effect on the mean and $\tilde{X}\tilde{\beta}$ and \tilde{a} those which act on the variability of the trait; f being the function for modelling the environmental variance and $\epsilon \sim N(0, I)$. The genetic effects a and \tilde{a} are assumed to be Gaussian:

$$\begin{pmatrix} a \\ \tilde{a} \end{pmatrix} | G \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, G \otimes A\right)$$

where, A is the additive genetic relationship matrix, and

$$G = \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_{\tilde{a}} \\ \rho\sigma_a\sigma_{\tilde{a}} & \sigma_{\tilde{a}}^2 \end{bmatrix}$$

with σ_a^2 and $\sigma_{\tilde{a}}^2$ the additive genetic variance for a and \tilde{a} respectively, and ρ the coefficient of genetic correlation in the joint distribution of a and \tilde{a} (see appendix C for further details).

The core of this software was written by Noelia Ibáñez, it uses an MCMC algorithm and considered an animal model for continuous traits. It has been modified by Milagros Garcia in order to deal with three different functions for modelling the environmental variance :

$$\text{Exp} : f_1(\tilde{X}\tilde{\beta}, Z\tilde{a}) = \exp\left(\frac{\tilde{X}\tilde{\beta} + Z\tilde{a}}{2}\right), \quad \text{as in San Cristobal et al (1998).}$$

$$\text{Sqr} : f_2(\tilde{X}\tilde{\beta}, Z\tilde{a}) = \sqrt{\tilde{X}\tilde{\beta} + Z\tilde{a}}, \quad \text{as in Mulder et al (2007).}$$

$$\text{Lin} : f_3(\tilde{X}\tilde{\beta}, Z\tilde{a}) = \tilde{X}\tilde{\beta} + Z\tilde{a}, \quad \text{as in Garcia et al (2009).}$$

Main specifications of the present version are:

- Univariate analysis
- Continuous trait
- Animal model with repeated records
- Only numerical fields for data and pedigree files.

The program is in active development so errors are possible and some features may not work or work incorrectly. This beta version admits input files similar to those of ASREML.

2 Software limitations

This program admits up to eight systematic effects (β plus $\tilde{\beta}$). However, not more than one covariable (that may acts on the mean and on the variance or on both) and one no genetic random effect (that may acts on the mean and on the variance) can be fitted in the model.

3 How to run the program

To run the program you need to put in a text file called 'Parameter' the path and name of the parameters file you want to use (ex: C:\home\parameters\my_parameters.file.txt). Then type the command GSEVM in your terminal.

To obtain graphical output, it is necessary to get R software installed on the machine. Two R-files are needed, "principal.r" and "statistics.r", they can be launched at any time after the burn-in period typing "R CMD BATCH principal.r". Graphical outputs are in a pdf file described in the parameter file section, to see this file type command "acroread" followed by the name of the graphical file.

4 Backup copy process

After the burning period, all computed values are saved each 2 hours in all files created by the software. After any stop (accidental computer break or normal end) of the program, if these files still exist and were not modified, it is possible to restart using the values stored at the last saving iterations (see Reload).

5 How to cite GSEVM

Use this software for research must be acknowledge in publications citing the relevant paper:

N. Ibáñez-Escriche, M. Garcia and D. Sorensen. 2009. GSEVM v2 (Genetical Structured Environmental Variance Model) User Guide.

6 Parameters and Input files

All data files used for the program are text files containing real and integer values only in free format. This program is driven by a parameter file, which is called 'Parameter' file (see section 3), the format of which is shown below. Some lines of this parameter file can be empty, the word Optional appears to indicate them. Key-words in capital and **Arial** need to be included as they appear. Fields in *italic* are those replaced by users.

6.1 SUMMARY OF THE PARAMETER FILE

TITLE

Description or title of the problem

Optional; description of the problem. If the first word in the title is "Reload", then the program restarts using the same input files and the files saved at the last saving iteration (see Details of parameter file)

SEED

seed

Optional; random seed to start the program (with a format of an integer of up to 8 digits). If more than 8 digits are given, the program take only the first eighth. If a real number is given, the program take only the integer part. If this line is empty the program uses a seed from the system.

INBREEDING FILE

f1

Optional; the inbreeding file is f1. Put an empty line if no inbreeding file is provided.

DATA FILE

f2

The data file is f2.

MODEL

mod

indicates the function for modelling the environmental variance; Exp, Sqr or Lin.

nf, np, nc

nf is the total number of systematic effects (β plus $\tilde{\beta}$. If the model includes the average mean, the value of nf is the total number of systematic effects +1). np and nc represent the existence of no genetic random effect and covariable respectively, np=1 if "no genetic random effect" exists or 0 if not; nc=1 if the covariable exists or 0 if not.

nlf₁, nlf₂, ..., nlf_{nf}, nlp, nanim

number of levels for the systematic effect 1, 2, ..., nf,
levels for the no genetic random effect if it exists (if it does not exist nlp is not specified in this line),
total number of animals in the pedigree file.

system₁, system₂, ..., system_{nf}, randomm, additivem, covariablem

model for the mean as a vector of ntf terms (0 or 1), where ntf=nf+np+nc+1, is the total number of the factors (average mean, systematic, no genetic random and genetic effect and covariable) present in the model (1 indicates that the factor is present in the model, otherwise 0).
If overl mean is fitted, one systematic effect with only one level must be added (see page 8).

systv₁, systv₂, ..., systv_{nf}, randomv, additivev, covariablev

model for the environmental variance as a vector of ntf terms (0 or 1, 1 indicates that the factor is present in the model, otherwise 0) .

PEDIGREE FILE

f3

The pedigree file is f3.

ITER LACK BURN-IN

iter lack burn

Parameters for the chain sampling. Total number of iterations, interval between samples and burn-in period.

FIX FIXR COV COVR

fixm fixv covm covv

Initial values for the mean of the mean and the variance (fixm and fixv), and for the regression coefficient for the mean and the variance (covm and covv).

VA VAR R

va vr r

Initial values for the variances of the genetic effect on the mean model (σ_a^2) and on the environmental variance model (σ_e^2), and correlation (ρ).

VP VPR

vp vpr

Initial values for the variance of the permanent effect on the mean model (σ_p^2) and on the environmental variance model (σ_p^2).

GENERATION FILE

f4

Optional; if a name file is indicated, the program computes at the end of the iterations, the means and the variances of the genetic values per generation and per line.

Remark 1 *It has to be noticed that doing that, it does not include these factors in the model.*

SUMMARY OUTPUT FILE

f5

Optional; name of the output file that will contain the summary of the results. If this line is empty, the program uses the default name file: SUMMARY in the current directory.

BREEDING VALUES FILE

f6

Optional; name of the file that will contain the estimated breeding values for the mean and for the variance model respectively. If this line is empty, the program uses the default name file: BREEDINGV in the current directory.

SYSTEMATIC EFFECTS FOR THE MEAN FILE

f7

Optional; name of the file that will contain the marginal posterior distribution of the systematic effect values for the mean model estimated by the program. If this line is empty, the program uses the default name file: SYSTEMATICS in the current directory.

SYSTEMATIC EFFECTS FOR THE VARIANCE FILE

f8

Optional; name of the file that will contain the marginal posterior distribution of the systematic effect values for the environmental variance model estimated by the program. If this line is empty, the program uses the default name file: SYSTEMATICS1 in the current directory.

VARIANCE FILE

f9

Optional; name of the file that will contain the marginal posterior distribution of the variance values for the mean and the environmental variance models respectively, estimated by the program. If this line is empty, the program uses the default name file: VARIANCES in the current directory.

GENERATION LINE OUTPUT FILE

f10

Optional; name of the file that will contain the marginal posterior distribution of the generation and line effects in case when a generation file has been given. If this line is empty, the program uses the default name file: MEGELI in the current directory.

GRAPHICAL OUTPUT FILE

f11

Optional; name of the file that will contain the graphical results. This file is a "pdf" file. If this line is empty, the program uses the default name file: GRAPHICS in the current directory.

6.2 EXAMPLE OF A PARAMETERS FILE:

TITLE

This is a description of the problem

SEED

67825564587

INBREEDING FILE

Inbreed.dat

DATA FILE

Datafile2

MODEL

Exp

1 0 0

1 1000

1 1

1 1

PEDIGREE FILE

Pedigreefile

ITER LACK BURN-IN

500 10 50

FIX FIXR COV COVR

3.5 0.5 0.00 0.0

VA VAR R

1.0 0.2 0.5

VP VPR

0.0001 0.0001

GENERATION FILE

SUMMARY OUTPUT FILE

Summaryfile

BREEDING VALUES FILE

../Test/Breedingfile

SYSTEMATIC EFFECTS FOR THE MEAN FILE

Systfile

SYSTEMATIC EFFECTS FOR THE VARIANCE FILE

../../Syst1file

VARIANCE FILE

Varfile

GENERATION LINE OUTPUT FILE

Generat

GRAPHICAL OUTPUT FILE

Graphfile

6.3 DETAILS OF THE PARAMETER FILE AND OTHER INPUT FILES

TITLE

This line is for the title of the problem and to restart the software after an accidental break or a normal end.

At each hour the software saves all values of the current iteration in a saving file and updates the output files. If the first word of the title line is "Reload", then the software reads the saved files to continue the computation with the number of iterations in the parameter file (This parameter ITER is the only value that could be changed in case of reload).

The output files will be updated and will contain the values for all iterations (initial + reload)

INBREEDING FILE

Name and path of the inbreeding file, for example: ../../Variance/file/inbreeding.txt

In this inbreeding file all animals in the pedigree file, have to be listed from the oldest to the youngest with the Wright's inbreeding coefficient (F).

If this inbreeding file does not exist or is not indicated in the parameters file, the program will not take into account the inbreeding between animals in the analysis.

Example of the inbreeding file: suppose the oldest animal is identified by the number 1 and the youngest one by the number 470, so the inbreeding file could be:

```
1 0.0
2 0.0
3 0.0
:
468 0.125
469 0.0
470 0.07
```

MODEL

In this part is indicated:

The function for modelling the environmental variance.

The total number of systematic effects, plus one if the average of the model has to be calculated, (either for the mean or for the variance), the existence or not of a no genetic random effect and the existence or not of a covariable.

The levels for the systematic and no genetic random effects.

The structure of the model.

Exp

```
2 1 1
2 3 400 470
1 0 1 1 1
1 1 0 1 0
```

first line:

Exp

here we indicate the residual variance function (in this case Exponential as in San Cristobal et al (1998)); two other functions are also available: Sqr and Lin (see page 1).

second line:

2 1 1

here we consider 2 systematic effects, a no genetic random effect and a covariable. At least one systematic effect (overall mean) must exist.

third line:

2 3 400 470

the first systematic effect contains 2 levels and the second one 3 levels; we have 400 levels for the no genetic random effect and 470 animals in the pedigree.

The program will compute only the overall mean if only one systematic effect with one level is indicated.

1 0 1 1 1

this line indicates the factors which affect the mean. For example, in this case the second systematic effect does not influence the mean. And we consider two random effects (additive genetic effect, no genetic random effect) and the existence of covariable.

1 1 0 1 0

this line indicates the factors which affect the environmental variance (here, two systematic effects are used on the variance model). In this case, the no genetic random effect and the covariable are not affecting the residual variance.

Remark 2 *Only one no genetic random effect is allowed.*

Only one covariable can be fit on the mean trait or the environmental variance.

At least it is necessary one systematic effect with one level as such the overall mean (μ).

PEDIGREE FILE

The pedigree file has the following format

animal number, sire number and dam number

```
1   471  0
2   471  0
3   471  0
   :
467  8   54
468  2   22
469  3   48
470  5   19
```

Remark 3 *When sire and/or dam are unknown their number must be replaced by either $n+1$ or 0.*

Remark 4 *The file has to be sorted so that the line giving the pedigree of an individual appears before any line where that individual appears as a parent.*

DATA FILE

Data file must contain levels of systematic and no genetic random effects, covariable and data values. Systematic and no genetic random effects have to be integer; covariable and data value can be real.

Animal have to be arranged from the oldest to the youngest and there is no text header in the file.

Data file has the following format:

animal number, $syst_1$ effect level, $syst_2$ effect level, ... no genetic random effect level, covariable value, data value.

Some example of data file are:

for instance: data file with a complete model:

```
1   1  3  1   50  101
2   1  2  2   20   8.3
2   2  3  2   30   9.5
   :
400  2  3  400  50  12.0
```

In this there are 400 animals with records, two systematic effects, one no genetic random effect and a covariable.

ITER LACK BURN-IN

This program runs a single chain in each analysis. In this part of the parameter file the parameters of the Markov chain of MCMC are defined:

- **ITER** determines the length of the MCMC
- **LACK** determines the interval between samples that will be saved from the MCMC, if **LACK** is equal to 1, all samples will be saved, if **LACK** is equal to 5, each five iterations the sample will be saved. The interval within samples, **LACK**, is used to reduce autocorrelation between computed samples.
- **BURN-IN** determines the length of the burn-in period. All posterior samples before of the burn-in period are discarded in the analysis and not saved.

FIX FIXR COV COVR

In this part the user provides the initial values used to start the program. **FIX** and **FIXR** are the average value of the systematic effects on the mean and the environmental variance model respectively. **COV** and **COVR** are the initial values for the regression coefficient on the mean and the environmental variance model respectively.

```
FIX  FIXR  COV  COVR
1.8  0.2   0.2  0.5
```

Remark 5 *FIXR usually need smaller values than FIX.*

VA VAR R

VA and **VAR** are the initial genetic variance values. **R** is the initial value for the correlation.

```
VA  VAR  R
1.0 0.2  0.5
```

VP VPR

VP and **VPR** are the initial values for the no genetic random effects on the mean and residual variance respectively.

```
Vp  VPR
0.8 0.3
```

GENERATION FILE

This file contain a first line with the number of generations and the population lines. After this first line, the following lines have three components: the animal identification, the generation number and the line number. This file requires to have only integer values. The generation file has the following format:

```
10 3
1  1 1
2  1 2
  :
469 5 3
470 6 2
```

7 Output files

While the program is running some information are written on the screen:

- name of files used
- number of the current iteration (each lack), variance components: σ_a^2 , σ_a^2 , ρ , (σ_p^2 , σ_p^2) if they exist
- DIC value

This program provides seven output files, if the names and paths of these files are not given in the parameter file the program uses default file names: SUMMARY, BREEDINGV, VARIANCES, SYSTEMATICS, SYSTEMATICSR, MEGELI and GRAPHICS and put them in the working directory. Otherwise, the program takes the names and paths given in the parameters file. For explanation purposes we will suppose the default names. Three others files are automatically generated in order to use the "Reload" mode, this files are: "gammas", "no_genetic" and "no_genetic.last", they contain the last values used to compute the last estimated genetic values, mean of no genetic effect and the last estimated values of the no genetic effect, respectively.

7.1 SUMMARY

This file provides a summary of the analysis results.

In this part, the variances of the proposal distributions are shown (kb*,kbCOV*,kp,...). The variance proposal is an important key value, because it can determine the acceptance ratio of the samples in the Metropolis Hasting. Acceptance ratios in the neighborhood of 1 imply very similar values between previous and proposed states and the chain will move very slowly. On the other hand, if the proposed displacement is too large and falls where the posterior has not support, this will lead to a high rejection rate. This software uses an adaptative algorithm in order to adjust these variances and obtain acceptance ratios of around 0.4. Here, the chain will remain in the same state for many iterations.

A more detail description of the proposal variances is in the subsection algorithm and implementation (appendix C)

The summary file is shown bellow with the following format (fields in *italic* and with xxx would be the result):

```
***** GSEVM RESULTS *****
TITLE: This is a description of the problem
*****
    This analysis (CPU time) took: xxx seconds
    Date of the analysis (yyyy/mm/dd): xxx/xx/xx
    Time of the analysis (hh/mm/ss) xx/xx/xx
*****
INITIAL SEED: xxx
FINAL SEED: xxx
*****
    Parameter file: parameters file name
    Inbreeding file: inbreeding file name and number of animals
    Data file: data file name
    Pedigree file: pedigree file name
    Generation file: generation file nname, number of generations and number of lines
*****
MCMC PARAMETERS
ITER: xxx
ITER SAUV: xxx
```

LACK: *xxx*
BURN-IN: *xxx*

MODEL

Number of animals in the pedigree file: *xxx*
Number of records in the data file: *xxx*
Number of systematic effects: *xxx*
Number of no genetic random effects: *xxx*
Number of covariables: *xxx*
Number of level for systematic effects 1: *xxx*
⋮
Number of level for no genetic random effect: *xxx*

Model as in parameter file: *xxx*

xxx xxx xxx
xxx xxx xxx xxx
xxx xxx xxx xxx xxx
xxx xxx xxx xxx xxx
xxx xxx xxx xxx xxx

DIC VALUES

Deviance: *xxx*
Effective number of parameters: *xxx*
DIC: *xxx*

VARIANCES AND CORRELATION

POSTERIOR MEAN OF THE GENETIC ADDITIVE VARIANCE FOR THE MEAN σ_a^2 [SD]:
xxx[xxx]
POSTERIOR MEAN OF THE GENETIC ADDITIVE VARIANCE FOR THE VARIANCE σ_a^2
[SD]: *xxx[xxx]*
POSTERIOR MEAN CORRELATION [SD]: *xxx[xxx]*
POSTERIOR MEAN OF THE NO GENETIC RANDOM EFFECT FOR THE MEAN σ_p^2 [SD]:
xxx[xxx]
POSTERIOR MEAN OF THE NO GENETIC RANDOM EFFECT FOR THE VARIANCE σ_p^2
[SD]: *xxx[xxx]*

ACCEPTANCE RATIOS

b*	=	xxx	partial b*	=	xxx	kb*	=	xxx
bCOV*	=	xxx	partial bCOV*	=	xxx	kbCOV*	=	xxx
p	=	xxx	partial p	=	xxx	kp	=	xxx
p*	=	xxx	partial p*	=	xxx	kp*	=	xxx
vp	=	xxx	partial vp	=	xxx	kvp	=	xxx
vp*	=	xxx	partial vp*	=	xxx	kvp*	=	xxx
a	=	xxx	partial a	=	xxx	ka	=	xxx
a*	=	xxx	partial a*	=	xxx	ka*	=	xxx
va	=	xxx	partial va	=	xxx	kva	=	xxx
va*	=	xxx	partial va*	=	xxx	kva*	=	xxx
R	=	xxx	partial R	=	xxx	kr	=	xxx

RESULTS OF ADDITIVE MEANS FOR GENERATION AND LINE

additive mean, generation, line

RESULTS OF ADDITIVE VARIANCE MEANS FOR GENERATION AND LINE

variance additive mean, generation, line

=====
STATISTICS OF SYSTEMATIC EFFECTS FOR THE MEAN
=====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
STATISTICS OF SYSTEMATIC EFFECTS FOR THE VARIANCE
=====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
STATISTICS OF COVARIABLE FOR THE MEAN
=====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
STATISTICS OF COVARIABLE FOR THE VARIANCE
=====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
STATISTICS OF GENETIC VARIANCE FOR THE MEAN
=====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
 STATISTICS OF GENETIC VARIANCE FOR THE VARIANCE
 =====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
 STATISTICS OF CORRELATION BETWEEN GENETIC VALUES
 =====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
 STATISTICS OF THE NO GENETIC VARIANCE FOR THE MEAN
 =====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
 STATISTICS OF THE NO GENETIC VARIANCE FOR THE VARIANCE
 =====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
 STATISTICS OF THE GENETIC VALUES FOR THE MEAN
 =====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
 STATISTICS OF THE GENETIC VALUES FOR THE VARIANCE
 =====

Effects	Min	1Quant	Median	Mean	3Quant	Max	SD
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

=====
 Normality-test for A
 =====

One-sample Kolmogorov-Smirnov test
 data: genetic values for the mean model
 alternative hypothesis: two-sided
 D= xxx p-value= xxx

=====
Normality-test for AR
=====

One-sample Kolmogorov-Smirnov test
data: genetic values for the mean model
alternative hypothesis: two-sided
D= xxx p-value= xxx

7.2 BREEDINGV

The BREEDINGV file contains, for all animals in the pedigree file, the animal number and the mean of his posterior additive genetic values for the mean a and the environmental variance \tilde{a} , respectively. Example:

```
1    0.12  0.02
2    0.05 -0.05
   :
470  0.1   0.01
```

7.3 SYSTEMATICS

The program give two output files for estimation of systematic effects. The first one, called SYSTEMATICS, contains samples of the marginal posterior of systematic effects and of the covariable of the mean model. The second one, called SYSTEMATICSR, contains samples of the marginal posterior of systematic effects and of the covariable for the residual variance model. The samples are in columns and follow the order of the systematic effects defined in the parameter file in model section followed by the covariable.

7.4 VARIANCES

This file contains the samples of the marginal posterior distribution of the variance components. The order of the columns are:

σ_a^2 , $\sigma_{\tilde{a}}^2$, σ_p^2 , $\sigma_{\tilde{p}}^2$

7.5 MEGELI

If the user gives a GENERATION file, the program generates one file with the average of the posterior additive values of the mean trait and his residual variance for each generation and each line in the corresponding iteration. These effects are not computed, instead of they stem from the posterior distribution of the genetic values.

7.6 GRAPHICS

This pdf file contains the graphical outputs of the analysis for all parameters, the order of the graphics are:

- historic diagram.
- posterior density.
- autocorrelation between each saved iteration (lack).
- boxplot.
- posterior distributions of all levels together of each systematic effect (10 by graphic).

- boxplot of all levels together.
- for the genetic values this file contains also the normality test.

APENDIX

A Algorithm and implementation

The model analysed is a non standard linear model which has considerable computational problems. In order to obtain efficient samples, different MCMC (Markov chain Monte Carlo) strategies are implemented. In this section it will be made a brief description of the algorithm and his implementation. However, a more detailed description can be found in [3, 4]

- Gibbs sampling: a conventional blocking Gibbs sampling [3] was used for β since it is the only full conditional distribution known.
- Adaptative Metropolis Hasting: allows to sampler from a unknown conditional distribution. This algorithm is used to sampler all parameters except β . The algorithm generates a Markov chain a^1, a^2, \dots as follows. Given the current value a^i of the Markov chain a proposal a^{prop} is generated from a proposal density $q(a^{prop} | y)$, with probability

$$(1) \quad \min \left\{ 1, \frac{p(a^{prop} | y)q(a^i | a^{prop})}{p(a^i | y)q(a^{prop} | y)} \right\}$$

the new state a^{i+1} is given by a^{prop} , otherwise $a^{i+1} = a^i$. The target of this algorithm is to choose a suitable proposal density q that allows obtain efficient samples from the posterior density. In this software two different proposal have been choosen:

- Random walk proposal: is a multivariate normal density centered in the current value x^i of the Markov chain and with covariance hI where h is a user-specified proposal variance and I is the identity matrix, i.e. $q(a^{prop} | a^i)$ is the density of $N(a^i, hI)$. This proposal is commonly used, however, for high-dimensional problems the convergence can be very slow and produce highly autocorrelated samples. In this software this proposal is used to sampler the variances $\sigma_a, \sigma_{\tilde{a}}, \sigma_p, \sigma_{\tilde{p}}$, and the correlation coefficient ρ .
- Langevin Hastings proposal: unlike the random walk, this proposal use a gradient information in the proposal density. The proposal density of a Langevin Hastings algorithm is given by $N(a^i + h \nabla \log p(a | y)/2, hI)$ where $\nabla \log p(a | y)$ is the gradient of the posterior density. Intuitively, the use of gradient information helps to direct the algorithm towards regions of high posterior density.

B Reparametrization

Simulation studies in Gustafson et al. [5] show that Langevin Hastings updates may not work well if the components of a have a very different posterior variances. In applications in quantitative genetics, the different animals may contribute with different number of observations and hence the posterior variances may be very different. It may be useful to transform a and \tilde{a} into quantities whose components are less correlated a posteriori. Using the factorisation $A = TDT'$ [1], where D is a diagonal matrix, one may let $a = \sigma_a \gamma D^{\frac{1}{2}} T'$ and $\tilde{a} = \rho \sigma_{\tilde{a}} \tilde{\gamma} D^{\frac{1}{2}} T' + \sqrt{1 - \rho^2} \sigma_{\tilde{a}} \tilde{\gamma} D^{\frac{1}{2}} T'$ where γ and $\tilde{\gamma}$ are the priors standard normal $N(0, I)$. The posterior correlation matrix of γ given y is then closer to the correlation matrix I of the Langevin Hastings proposal.

C Theoretical background

C.1 The model

Consider the model which assumes that the environmental variance is heterogeneous and partly under genetic control:

$$y_i = x'_i \beta + z'_i a + w'_i p + f(\tilde{x}'_i \tilde{\beta}, z'_i \tilde{a}, \tilde{w}'_i \tilde{p}) \epsilon_i$$

for the exponential model,

$$f(\tilde{x}'_i\tilde{\beta}, z'_i\tilde{a}, \tilde{w}'_i\tilde{p}) = \exp\left(\frac{\tilde{x}'_i\tilde{\beta} + z'_i\tilde{a} + \tilde{w}'_i\tilde{p}}{2}\right)$$

where, y_i is the i^{th} observation of a given individual, β and $\tilde{\beta}$ are the vectors associated with the systematic effects, a and \tilde{a} the random vectors associated with the genetic effects, p and \tilde{p} are the random vectors associated with the no genetic random effects and $x_i, \tilde{x}_i, z_i, \tilde{z}_i, w_i, \tilde{w}_i$ are the incidence vectors.

For the other two models, we proceed in the same way as bellow.

Prior following prior distributions were assigned to the parameters:

The systematic effects β and $\tilde{\beta}$ are assumed to be Uniform distribution,

$$\beta \sim U(l, u), \tilde{\beta} \sim U(l, u)$$

The genetic effects a and \tilde{a} are assumed to be Gaussian:

$$\begin{pmatrix} a \\ \tilde{a} \end{pmatrix} | G \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, G \otimes A\right)$$

where A is the additive genetic relationship matrix, and

$$G = \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_{\tilde{a}} \\ \rho\sigma_a\sigma_{\tilde{a}} & \sigma_{\tilde{a}}^2 \end{bmatrix}$$

with, σ_a^2 and $\sigma_{\tilde{a}}^2$ the additive genetic variance for a and \tilde{a} respectively, and ρ the coefficient of genetic correlation in the joint distribution of a and \tilde{a}

$$\sigma_k^2 \sim v_k \chi_{v_k}^{-2}, \text{ with } k = a, \tilde{a}$$

$$\rho \sim U(-1, 1)$$

Vector p and \tilde{p} are assumed to be independent with distribution,

$$\begin{pmatrix} p \\ \tilde{p} \end{pmatrix} | G_p \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, G_p \otimes I\right)$$

where I is the identity matrix with order equal to number of no genetic random effects and

$$G_p = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_{\tilde{p}}^2 \end{bmatrix}$$

σ_p^2 and $\sigma_{\tilde{p}}^2$ are the components of the variances due to no genetic random effects affecting the mean and the residual variance models respectively.

$$\sigma_m^2 \sim v_m \chi_{v_m}^{-2}, \text{ with } m = p, \tilde{p}$$

To simplify let $\sigma_i^2 = \exp(\tilde{x}_i\tilde{\beta} + z_i\tilde{a} + \tilde{w}_i\tilde{p})$. So,

$$y | \beta, a, \sigma_i^2 \sim N(X\beta + Za + Wp, \text{diag}(\sigma_i^2))$$

The number of elements in y is n (number of records) and in a is q (number of animals in the pedigree). Posterior distribution is

(2)

$$p(\beta, a, \sigma_i^2, G, G_p | y) \propto p(y | \beta, a, \sigma_i^2, G, G_p) \cdot p(a, \tilde{a} | G) \cdot p(\sigma_a^2) p(\sigma_{\tilde{a}}^2) \cdot p(\rho) p(p, \tilde{p} | G_p) \cdot p(\sigma_p^2) p(\sigma_{\tilde{p}}^2)$$

$$\begin{aligned} p(\beta, a, \sigma_i^2, G, G_p | y) &\propto \left\{ \prod_{i=1}^n (\sigma_i^2)^{-\frac{1}{2}} \exp \left[-\frac{(y_i - x'_i \beta - z'_i a - w'_i p)^2}{2\sigma_i^2} \right] \right\} \cdot |2\pi A|^{-\frac{1}{2}} |G|^{-\frac{q}{2}} \\ &\cdot \exp \left[-\frac{1}{2} (a, \tilde{a})' (G^{-1} \otimes A^{-1}) (a, \tilde{a}) \right] \cdot (\sigma_a^2)^{-\left(\frac{v_a}{2}+1\right)} \exp \left[-\frac{v_a S_a}{2\sigma_a^2} \right] \cdot (\sigma_{\tilde{a}}^2)^{-\left(\frac{v_{\tilde{a}}}{2}+1\right)} \exp \left[-\frac{v_{\tilde{a}} S_{\tilde{a}}}{2\sigma_{\tilde{a}}^2} \right] \\ &\cdot \exp \left[-\frac{1}{2} (p, \tilde{p})' (G_p^{-1} \otimes I) (p, \tilde{p}) \right] (\sigma_p^2)^{-\left(\frac{v_p}{2}+1\right)} \exp \left[-\frac{v_p S_p}{2\sigma_p^2} \right] \cdot (\sigma_{\tilde{p}}^2)^{-\left(\frac{v_{\tilde{p}}}{2}+1\right)} \exp \left[-\frac{v_{\tilde{p}} S_{\tilde{p}}}{2\sigma_{\tilde{p}}^2} \right] \end{aligned}$$

Introducing the one-to-one transformation

$$\begin{aligned} \begin{pmatrix} \gamma \\ \tilde{\gamma} \end{pmatrix} &= L^{-1} \otimes D^{-\frac{1}{2}} T^{-1} \begin{pmatrix} a \\ \tilde{a} \end{pmatrix} \\ \begin{pmatrix} a \\ \tilde{a} \end{pmatrix} &= L \otimes T D^{\frac{1}{2}} \begin{pmatrix} \gamma \\ \tilde{\gamma} \end{pmatrix} \end{aligned}$$

where D is diagonal, T^{-1} is lower triangular and L is lower triangular in order of Cholesky decomposition of G matrix.

$$\begin{aligned} G &= L L' \\ L &= \begin{bmatrix} \sigma_a & 0 \\ \rho\sigma_{\tilde{a}} & \sigma_{\tilde{a}}\sqrt{1-\rho^2} \end{bmatrix} \\ \begin{pmatrix} \gamma \\ \tilde{\gamma} \end{pmatrix} &= \begin{pmatrix} \sigma_a T D^{\frac{1}{2}} \gamma \\ \rho\sigma_{\tilde{a}} T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}}\sqrt{1-\rho^2} T D^{\frac{1}{2}} \tilde{\gamma} \end{pmatrix} \end{aligned}$$

where

$$\begin{pmatrix} \gamma \\ \tilde{\gamma} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, I_2 \otimes I_q \right)$$

The sampling model of the data is now

$$p(y | \beta, \gamma, p, \sigma_i^2) \propto \prod_{i=1}^n (\sigma_i^2)^{-\frac{1}{2}} \exp \left[-\frac{(y_i - x'_i \beta - z'_i \sigma_a T D^{\frac{1}{2}} \gamma - w'_i p)^2}{2\sigma_i^2} \right]$$

with $l < \beta, \tilde{\beta} < u$, $-1 < \rho < 1$ and

$$\sigma_i^2 = \exp \left(\tilde{x}'_i \tilde{\beta} + z'_i \left(\rho\sigma_{\tilde{a}} T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}}\sqrt{1-\rho^2} T D^{\frac{1}{2}} \tilde{\gamma} \right) + \tilde{w}'_i \tilde{p} \right)$$

The posterior density to be used to estimate the parameters is proportional to

$$(3) \quad p(\beta, \tilde{\beta}, \gamma, \tilde{\gamma}, p, \tilde{p}, \sigma_a^2, \sigma_{\tilde{a}}^2, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2 | y) \propto \left\{ \prod_{i=1}^n (\sigma_i^2)^{-\frac{1}{2}} \exp \left[-\frac{(y_i - x'_i \beta - z'_i \sigma_a T D^{\frac{1}{2}} \gamma - w'_i p)^2}{2\sigma_i^2} \right] \right\}$$

$$\begin{aligned} & \cdot \exp \left[-\frac{1}{2} \left(\gamma' \gamma + \tilde{\gamma}' \tilde{\gamma} \right) \right] (\sigma_a^2)^{-\left(\frac{v_a}{2}+1\right)} \exp \left[-\frac{v_a S_a}{2\sigma_a^2} \right] (\sigma_a^2)^{-\left(\frac{v_a}{2}+1\right)} \exp \left[-\frac{v_a S_a}{2\sigma_a^2} \right] \\ & \cdot \exp \left[-\frac{1}{2} (p, \tilde{p})' (G_p^{-1} \otimes I) (p, \tilde{p}) \right] (\sigma_p^2)^{-\left(\frac{v_p}{2}+1\right)} \exp \left[-\frac{v_p S_p}{2\sigma_p^2} \right] (\sigma_p^2)^{-\left(\frac{v_p}{2}+1\right)} \exp \left[-\frac{v_p S_p}{2\sigma_p^2} \right] \end{aligned}$$

For numerical convenience, it is usually easier to compute the maximum in the log posterior density. This maximum will coincide with the maximum in the posterior density. However, because the calculation the log of the posterior density involves a summation rather than a product, there are computational advantages to the use the log.

The log-posterior density is then given by

$$\begin{aligned} (4) \quad \ln p \left(\beta, \tilde{\beta}, \gamma, \tilde{\gamma}, \sigma_a^2, \sigma_a^2, \rho, \sigma_p^2, \sigma_p^2 \mid y \right) &= -\left(\frac{v_p}{2} + 1\right) \ln \sigma_p^2 - \left(\frac{v_{\tilde{p}}}{2} + 1\right) \ln \sigma_{\tilde{p}}^2 \\ &\quad - \frac{v_p S_p}{2\sigma_p^2} - \frac{v_{\tilde{p}} S_{\tilde{p}}}{2\sigma_{\tilde{p}}^2} - \left(\frac{v_a}{2} + 1\right) \ln \sigma_a^2 - \left(\frac{v_{\tilde{a}}}{2} + 1\right) \ln \sigma_{\tilde{a}}^2 - \frac{v_a S_a}{2\sigma_a^2} - \frac{v_{\tilde{a}} S_{\tilde{a}}}{2\sigma_{\tilde{a}}^2} \\ &\quad - \frac{1}{2} \left(\frac{p' p}{\sigma_p^2} + \frac{\tilde{p}' \tilde{p}}{\sigma_{\tilde{p}}^2} \right) - \frac{1}{2} \left(\gamma' \gamma + \tilde{\gamma}' \tilde{\gamma} \right) - \frac{1}{2} \sum_{i=1}^n \ln \sigma_i^2 - \sum_{i=1}^n \frac{1}{2\sigma_i^2} \left(y_i - x_i' \beta - z_i' \sigma_a T D^{\frac{1}{2}} \gamma - w_i' \tilde{p} \right)^2 \end{aligned}$$

In next section is presented the way to obtain the estimated.

C.2 The MCMC algorithm

The model is fitted using the following algorithm:

- Gibbs update for β
- Langevin-Hastings proposal for $\tilde{\beta}$
- Langevin-Hastings proposal for $\gamma, \tilde{\gamma}, p$ and \tilde{p}
- Normal random walk proposal for $\sigma_a^2, \sigma_a^2, \rho$

C.2.1 Gibbs update for β

From standard theory,

$$\beta \mid \tilde{\beta}, \gamma, \tilde{\gamma}, p, \tilde{p}, \sigma_a^2, \sigma_a^2, \rho, \sigma_p^2, \sigma_p^2, y \sim N \left(\hat{\beta}, \left(\sum_{i=1}^n \sigma_i^2 \right)^{-1} \right)$$

where

$$\hat{\beta} = \left(\sum_{i=1}^n \sigma_i^2 \right)^{-1} \left(\sum_{i=1}^n y_i \sigma_i^2 - 1' R^{-1} Z a \right)$$

where R^{-1} is a $n \times n$ diagonal matrix with i th element equal to

$$\sigma_i^{-2} = \exp \left(-\tilde{x}_i' \tilde{\beta} - z_i' \left(\rho \sigma_a T D^{\frac{1}{2}} \gamma + \sigma_a \sqrt{1 - \rho^2} T D^{\frac{1}{2}} \tilde{\gamma} \right) - \tilde{w}_i' \tilde{p} \right)$$

C.2.2 Langevin-Hastings proposal for $\tilde{\beta}$

This algorithm requires

$$\frac{\partial}{\partial \tilde{\beta}} \ln p \left(\beta \mid \tilde{\beta}, \gamma, \tilde{\gamma}, p, \tilde{p}, \sigma_a^2, \sigma_{\tilde{a}}^2, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right)$$

The conditional posterior distribution for $\tilde{\beta}$ is proportional to

$$(5) \quad -\frac{1}{2} \sum_{i=1}^n \ln \sigma_i^2 - \sum_{i=1}^n \frac{1}{2\sigma_i^2} \left(y_i - x_i' \beta - z_i' \sigma_a T D^{\frac{1}{2}} \gamma - w_i' p \right)^2$$

$$\sum_{i=1}^n \ln \sigma_i^2 = \sum_{i=1}^n \tilde{x}_i' \tilde{\beta} + \rho \sigma_{\tilde{a}} \sum_{i=1}^n z_i' T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}} \sqrt{1 - \rho^2} \sum_{i=1}^n z_i' T D^{\frac{1}{2}} \tilde{\gamma} + \sum_{i=1}^n \tilde{w}_i' \tilde{p}$$

$$\sigma_i^{-2} = \exp \left(-\tilde{x}_i' \tilde{\beta} - z_i' \left(\rho \sigma_{\tilde{a}} T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}} \sqrt{1 - \rho^2} T D^{\frac{1}{2}} \tilde{\gamma} \right) - \tilde{w}_i' \tilde{p} \right)$$

Taking logarithms in (5) and keeping only terms in $\tilde{\beta}$ we obtain

$$\frac{\partial}{\partial \tilde{\beta}} \ln p \left(\beta \mid \tilde{\beta}, \gamma, \tilde{\gamma}, p, \tilde{p}, \sigma_a^2, \sigma_{\tilde{a}}^2, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right) = -\frac{n}{2} + \sum_{i=1}^n \left(y_i - x_i' \beta - z_i' \sigma_a T D^{\frac{1}{2}} \gamma - w_i' p \right)^2$$

$$\cdot \exp \left(-\tilde{x}_i' \tilde{\beta} - z_i' \left(\rho \sigma_{\tilde{a}} T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}} \sqrt{1 - \rho^2} T D^{\frac{1}{2}} \tilde{\gamma} \right) - \tilde{w}_i' \tilde{p} \right)$$

The mean of the proposal distribution for drawing at state $t + 1$ is

$$(6) \quad \left[\tilde{\beta}^{[t]} - \frac{kn}{4} + \frac{k}{2} \sum_{i=1}^n \frac{1}{2} \left(y_i - x_i' \beta - z_i' a - w_i' p \right)^2 \exp \left(-\tilde{x}_i' \tilde{\beta} - z_i' \tilde{a} - \tilde{w}_i' \tilde{p} \right) \right]$$

Let Y_β and $Y_{\tilde{\beta}}$ the proposed values for β and $\tilde{\beta}$, so that

$$Y = \left(Y_\beta, Y_{\tilde{\beta}} \right)$$

and the state of the chain at stage t be denoted by

$$H_t = \left(\beta^{[t]}, \tilde{\beta}^{[t]} \right) = h$$

Then from (5)

$$\ln \pi(Y) = -\frac{1}{2} \sum_{i=1}^n \ln Y \sigma_i^2 - \sum_{i=1}^n \frac{1}{2Y \sigma_i^2} \left(y_i - x_i' \beta - z_i' \sigma_a T D^{\frac{1}{2}} \gamma - w_i' p \right)^2$$

where

$$\sum_{i=1}^n \ln Y \sigma_i^2 = n Y_{\tilde{\beta}} + \rho \sigma_{\tilde{a}} \sum_{i=1}^n z_i' T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}} \sqrt{1 - \rho^2} \sum_{i=1}^n z_i' T D^{\frac{1}{2}} \tilde{\gamma} + \tilde{w}_i' \tilde{p}$$

and

$$Y \sigma_i^2 = \exp \left(Y_{\tilde{\beta}} + z_i' \left(\rho \sigma_{\tilde{a}} T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}} \sqrt{1 - \rho^2} T D^{\frac{1}{2}} \tilde{\gamma} \right) + \tilde{w}_i' \tilde{p} \right)$$

$$\ln \pi(h) = -\frac{1}{2} \sum_{i=1}^n \ln \sigma_i^2 - \sum_{i=1}^n \frac{1}{2\sigma_i^2} \left(y_i - x_i' \beta - z_i' \sigma_a T D^{\frac{1}{2}} \gamma - w_i' p \right)^2$$

where

$$\sum_{i=1}^n \ln \sigma_i^2 = \sum_{i=1}^n \tilde{x}'_i \tilde{\beta} + \rho \sigma_{\tilde{a}} \sum_{i=1}^n z'_i T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}} \sqrt{1 - \rho^2} \sum_{i=1}^n z'_i T D^{\frac{1}{2}} \tilde{\gamma} + \sum_{i=1}^n \tilde{w}'_i \tilde{p}$$

and

$$\sigma_i^2 = \exp \left(\tilde{x}'_i \tilde{\beta} + z'_i \left(\rho \sigma_{\tilde{a}} T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}} \sqrt{1 - \rho^2} T D^{\frac{1}{2}} \tilde{\gamma} \right) + \tilde{w}'_i \tilde{p} \right)$$

Then we have

$$q(Y_{\beta}; h) = \exp \left(-\frac{1}{2} \left(\tilde{\beta}^{[t]} - Y_{\tilde{\beta}} - \left(\frac{kn}{4} - \frac{k}{4} \sum_{i=1}^n (y_i - x'_i \beta - z'_i a - w'_i p)^2 \exp \left(-\tilde{x}'_i \tilde{\beta} - z'_i \tilde{a} - \tilde{w}'_i \tilde{p} \right) \right) \right)^2 \right)$$

where k is a user-specified proposal variance.

Similarly

$$q(h; Y) = q(h; Y_{\tilde{\beta}})$$

These expressions are used to compute the acceptance ratio:

$$(7) \quad \alpha_j(h; Y) = \left\{ \frac{\pi(Y) q_j(Y; h_j)}{\pi(h) q_j(h; Y_j)} \right\}$$

C.2.3 Langevin-Hastings proposal for $\gamma, \tilde{\gamma}$

This algorithm requires

$$(8) \quad \begin{aligned} & \frac{\partial}{\partial \gamma} \ln p \left(\gamma, \tilde{\gamma} \mid \beta, \tilde{\beta}, p, \tilde{p}, \sigma_a^2, \sigma_{\tilde{a}}^2, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right) \\ & \frac{\partial}{\partial \tilde{\gamma}} \ln p \left(\gamma, \tilde{\gamma} \mid \beta, \tilde{\beta}, p, \tilde{p}, \sigma_a^2, \sigma_{\tilde{a}}^2, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right) \end{aligned}$$

Taking equation (4) and keeping only terms in $\gamma, \tilde{\gamma}$ we obtain

$$\begin{aligned} \ln p \left(\gamma, \tilde{\gamma} \mid \beta, \tilde{\beta}, \sigma_a^2, \sigma_{\tilde{a}}^2, p, \tilde{p}, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right) & \propto -\frac{1}{2} \left(\gamma' \gamma + \tilde{\gamma}' \tilde{\gamma} \right) \\ & - \frac{1}{2} \sum_{i=1}^n \ln \sigma_i^2 - \sum_{i=1}^n \frac{1}{2} \left(y_i - x'_i \beta - z'_i \sigma_a T D^{\frac{1}{2}} \gamma - w'_i p \right)^2 \end{aligned}$$

Differentiating with respect to $\gamma, \tilde{\gamma}$ we obtain:

$$\begin{aligned} & \frac{\partial}{\partial \gamma} \ln p \left(\gamma, \tilde{\gamma} \mid \beta, \tilde{\beta}, \sigma_a^2, \sigma_{\tilde{a}}^2, p, \tilde{p}, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right) = -\gamma - \frac{1}{2} \rho \sigma_{\tilde{a}} \sum_{i=1}^n D^{\frac{1}{2}} T' z_i \\ & + \sigma_a \sum_{i=1}^n D^{\frac{1}{2}} T' z_i \left(y_i - x'_i \beta - z'_i \sigma_a T D^{\frac{1}{2}} \gamma - w'_i p \right) \exp \left(-\tilde{x}'_i \tilde{\beta} - \rho \sigma_{\tilde{a}} z'_i T D^{\frac{1}{2}} \gamma - \sigma_{\tilde{a}} \sqrt{1 - \rho^2} z'_i T D^{\frac{1}{2}} \tilde{\gamma} - \tilde{w}'_i \tilde{p} \right) \\ & + \rho \sigma_{\tilde{a}} \sum_{i=1}^n D^{\frac{1}{2}} T' z_i \frac{1}{2} \exp \left(-\tilde{x}'_i \tilde{\beta} - \rho \sigma_{\tilde{a}} z'_i T D^{\frac{1}{2}} \gamma - \sigma_{\tilde{a}} \sqrt{1 - \rho^2} T D^{\frac{1}{2}} \tilde{\gamma} - \tilde{w}'_i \tilde{p} \right) \left(y_i - x'_i \beta - z'_i \sigma_a T D^{\frac{1}{2}} \gamma - w'_i p \right)^2 \end{aligned}$$

to oversimplify the computation

$$\begin{aligned} \frac{\partial}{\partial \gamma} \ln p \left(\gamma, \tilde{\gamma} \mid \beta, \tilde{\beta}, \sigma_a^2, \sigma_{\tilde{a}}^2, p, \tilde{p}, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right) &= -\gamma + \sum_{i=1}^n D^{\frac{1}{2}} T' z_i \cdot \\ &\left(-\frac{1}{2} \rho \sigma_{\tilde{a}} + \sigma_a (y_i - x'_i \beta - z'_i a - w'_i p) \exp \left(-\tilde{x}'_i \tilde{\beta} - z'_i \tilde{a} - \tilde{w}'_i \tilde{p} \right) \right. \\ &\left. + \frac{1}{2} \rho \sigma_{\tilde{a}} \exp \left(-\tilde{x}'_i \tilde{\beta} - z'_i \tilde{a} - \tilde{w}'_i \tilde{p} \right) (y_i - x'_i \beta - z'_i a - w'_i p)^2 \right) \end{aligned}$$

Call now

$$\begin{aligned} W_1 &= \sum_{i=1}^n z_i \left(-\frac{1}{2} \rho \sigma_{\tilde{a}} + \frac{1}{2} \rho \sigma_{\tilde{a}} \exp \left(-\tilde{x}'_i \tilde{\beta} - z'_i \tilde{a} - \tilde{w}'_i \tilde{p} \right) (y_i - x'_i \beta - z'_i a - w'_i p)^2 \right. \\ &\quad \left. + \sigma_a (y_i - x'_i \beta - z'_i a - w'_i p) \exp \left(-\tilde{x}'_i \tilde{\beta} - z'_i \tilde{a} - \tilde{w}'_i \tilde{p} \right) \right) \end{aligned}$$

a simple computational strategy for obtain $T'W_1 = W_1^*$ is the following. Write

$$T^{-1}(TW_1) = W_1$$

$$\frac{\partial}{\partial \gamma} \ln p \left(\gamma, \tilde{\gamma} \mid \beta, \tilde{\beta}, p, \tilde{p}, \sigma_a^2, \sigma_{\tilde{a}}^2, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right) = -\gamma + D^{\frac{1}{2}} W_1^*$$

Differentiating with respect to $\tilde{\gamma}$ gives

$$\begin{aligned} \frac{\partial}{\partial \tilde{\gamma}} \ln p \left(\gamma, \tilde{\gamma} \mid \beta, \tilde{\beta}, p, \tilde{p}, \sigma_a^2, \sigma_{\tilde{a}}^2, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right) &= -\tilde{\gamma} - \frac{\sigma_{\tilde{a}} \sqrt{1 - \rho^2}}{2} \sum_{i=1}^n D^{\frac{1}{2}} T' z_i \cdot \\ &+ \sigma_{\tilde{a}} \sqrt{1 - \rho^2} \sum_{i=1}^n D^{\frac{1}{2}} T' z_i \left(\frac{1}{2} (y_i - x'_i \beta - \sigma_a z'_i T D^{\frac{1}{2}} \gamma - w'_i p)^2 \right. \\ &\quad \left. \exp \left(-\tilde{x}'_i \tilde{\beta} - \rho \sigma_{\tilde{a}} z'_i T D^{\frac{1}{2}} \gamma + \sigma_{\tilde{a}} \sqrt{1 - \rho^2} z'_i T D^{\frac{1}{2}} \tilde{\gamma} - \tilde{w}'_i \tilde{p} \right) \right) \end{aligned}$$

to oversimplify the computation, we can also write

$$W_2 = \sum_{i=1}^n z_i \left(-\frac{\sigma_{\tilde{a}} \sqrt{1 - \rho^2}}{2} + \frac{\sigma_{\tilde{a}} \sqrt{1 - \rho^2}}{2} (y_i - x'_i \beta - z'_i a - w'_i p)^2 \exp \left(-\tilde{x}'_i \tilde{\beta} - z'_i \tilde{a} - \tilde{w}'_i \tilde{p} \right) \right)$$

Let call $W_2^* = T'W_2$. So,

$$T^{-1}(TW_2) = W_2$$

$$\frac{\partial}{\partial \tilde{\gamma}} \ln p \left(\gamma, \tilde{\gamma} \mid \beta, \tilde{\beta}, p, \tilde{p}, \sigma_a^2, \sigma_{\tilde{a}}^2, \rho, \sigma_p^2, \sigma_{\tilde{p}}^2, y \right) = -\tilde{\gamma} + D^{\frac{1}{2}} W_2^*$$

The mean of the proposal distribution for drawing at state $t + 1$ is

$$\begin{bmatrix} \gamma^{(t)} + \frac{k}{2} \left(D^{\frac{1}{2}} W_1^* - \gamma^{(t)} \right) \\ \tilde{\gamma}^{(t)} + \frac{k}{2} \left(D^{\frac{1}{2}} W_2^* - \tilde{\gamma}^{(t)} \right) \end{bmatrix}$$

Remark 6 Langevin Hastings proposal for p and \tilde{p} are computed in the same way as for γ and $\tilde{\gamma}$.

C.2.4 Normal random walk proposal for $\sigma_a, \sigma_{\bar{a}}, \sigma_p, \sigma_{\bar{p}}$ and ρ

Joint updating of $\ln \sigma_a$ and $\ln \sigma_{\bar{a}}$ using Metropolis-Hastings algorithm can be implemented as follows. Let $Y_{\sigma_a^2}, Y_{\sigma_{\bar{a}}^2}$ and Y_ρ the proposed values generated from the two dimensional normal distribution $N(m, Ik)$ centered at the previous realizations $\ln \sigma_a^2, \ln \sigma_{\bar{a}}^2, \rho$; that is, $m = (\ln \sigma_a^2, \ln \sigma_{\bar{a}}^2, \rho)$. Matrix I is the identity of dimension $3 \otimes 3$ and k is a user supplied turning scalar. Thus $\ln Y_{\sigma_a^2}$ is generated from

$$\begin{aligned} q \ln \sigma_a^2 &= (2\pi k)^{-\frac{1}{2}} \exp \left[-\frac{(\ln Y_{\sigma_a^2} - \ln \sigma_a^2)^2}{2k} \right] \frac{1}{Y_{\sigma_a^2}} \\ q \ln \sigma_{\bar{a}}^2 &= (2\pi k)^{-\frac{1}{2}} \exp \left[-\frac{(\ln Y_{\sigma_{\bar{a}}^2} - \ln \sigma_{\bar{a}}^2)^2}{2k} \right] \frac{1}{Y_{\sigma_{\bar{a}}^2}} \\ q\rho &= (2\pi k)^{-\frac{1}{2}} \exp \left[-\frac{(Y_\rho - \rho)^2}{2k} \right] \end{aligned}$$

The acceptance ratio involves the density of the proposal distribution for $\ln \sigma_a$ and $\ln \sigma_{\bar{a}}, \rho$ given by $q \ln \sigma_a^2, q \ln \sigma_{\bar{a}}^2, q\rho$ as well as all terms in equation (3) except $\exp \left[-\frac{1}{2} (\gamma' \gamma + \tilde{\gamma}' \tilde{\gamma}) \right]$

Remark 7 Normal random walk proposal for σ_p^2 and $\sigma_{\bar{p}}^2$ are obtained in the same way as for σ_a^2 and $\sigma_{\bar{a}}^2$.

References

- [1] C. R. Henderson. 1976. A simple method for the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics*, 32:69-83.
- [2] M. SanCristobal-Gaudy, J.M. Elsen, L. Bodin and C. Chevalet. 1998. Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genetics Selection Evolution*, 30:423-451.
- [3] D. Sorensen and D. Gianola. 2002. *Likelihood, Bayesian and Markov chain Monte Carlo methods in quantitative genetics*, Springer-Verlag.
- [4] D. Sorensen and R. Waagepetersen. 2003. Normal linear models with genetically structured variance heterogeneity: a case of study. *Genetical Research*, 82:207-222.
- [5] P. Guftanson, Y. C. Macnab and S. Wen. 2004. On the value of derivative evaluations and random walk suppression in Markov chain Monte Carlo algorithms. *Statistics and computing*, 14:23-28.