# gamevar.f90

**A small and practical software for (co)variance of gametic diversity and others components of the coefficient of relative variation of the genetic additive values estimation**

Daniel Jordan de Abreu Santos[1], John B. Cole[2], Paul M. VanRaden[2] and Li Ma[1]

[1]Department of Animal and Avian Sciences, University of Maryland, College Park, 20742

[2]Henry A. Wallace Beltsville Agricultural Research Center, Animal Genomics and Improvement Laboratory, ARS, USDA, Beltsville, MD 20705-2350

July 12, 2019

## License

gamevar.f90 was developed by University of Maryland and partnership with USDA in order to benefit the U.S. dairy cattle genetic evaluations and has been partially financed by … This is a free Fortran software: you may use, redistribute, modify and improve the code. Please report any bugs to daniel_jordan2008@hotmail.com.

# Contents

# 1 - Introduction

The gamevar.90 is a software for individual genetic variability parameters estimation useful for selection strategies, mating and progeny testing designs. Traditionally, the selection and mating design aiming a genetic additive improvement were based on the estimated breeding values (EBVs) of the individuals. The EBV represents the sum of the average effects of the genes, which is independent on the variability of meiotic recombination events (i.e. separation of homologous chromosomes and crossover). This variability though can be computed as a function of the binomial probabilities of transmission of DNA variants (markers) with known effects of an individual to its gametes, originating a normal distribution variance, as well as the covariance between two traits, of all possible additive gamete values, the (co)variance of gametic diversity ($\sigma^2_{\text{gamete}}$). The standard deviation of gametic diversity can be combined with EBV as a new selection index, RPTA, which aid to control the genetic diversity and improve the genetic gain in the long term (Santos et al., 2019). Thus, the gamavar.90 estimates the (co) variance caused by meiotic events using data on phased genotypes, allelic substitution effects and recombination rates between the variants from a chromosome.

The gamevar.f90 also calculates per chromosome the component $\sum_i^{NHom} \alpha_i^2$ (HOM – sum of squared effects of the homozygous loci from an individual) and coefficient of relative variation (CRV) as described by Santos et al., (2019). The CRV measures of variability in percentage of additive genetic values transmitted from an individual to its gametes. The useful statistics $\sigma^2_{\text{gamete}}$ and CRV shall to include all chromosome used in the analysis to predict the Genomic Breeding Values. The gamevar.f90 does not use all chromosome simultaneously but separately. The total statistics can obtained in a second

step by the user as simple sum by chromosome of the $\sigma^2_{\text{gamete}}$ and the component $\sum_i^{NHom} \alpha_i^2$ of CRV, applying the formula $CRV = \frac{\sigma_{gamete}}{\sqrt{0.5 \sum_i^{NHom} \alpha_i^2 + \sigma^2_{\text{gamete}}}}$.

The gamevar.f90 does not calculate the RPTA, since it depends on the future intensity of selection and can be also adjusted by the average of individual percentage of variation in a future progeny group (given as function of number of progeny, the critical value associated with the degree of confidence, and the average of the population CRV).

# 2 - Compiling

This is a standard code composed by some subroutines written in FORTRAN language, without use of any additional library. The dynamic allocable components are used thorough whole code. The standard compilers for Fortran90 and 95, such as gfortran, are suitable and recommended. A version already compiled in Linux environment is also available, in case of compiling problems. A compiling example in gfortran on an Unix terminal follows:

```
gfortran -o gamevar gamevar.f90
```

# 3 - Input File Formats

## 3.1 – Parameter File

A parameter file is required to run gamevar.f90. The parameter file provides some user-specific controls and the name of this file is determined by the user. The user can not add or delete any line in the parameter file. The number of traits specified must equal to the number of allele effects columns. The name of the input files must be specified

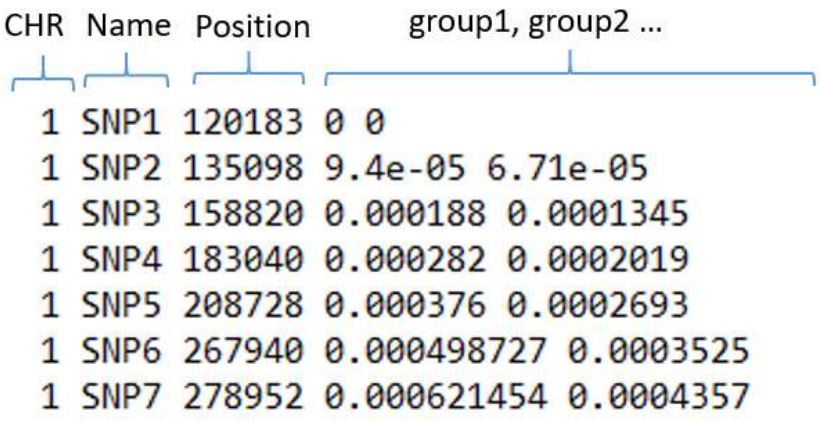
correctly bellow the "*_FILE" options. The option GENETIC_DISTANCE_UNIT the user must specified if the genetic distance between the variants is in centiMorgan (morgans) or recombination rate (reco). If the option is TRUE the software creates an output file with both, variance and covariance of the gametic diversity. If the option CRV is TRUE the software estimates the variance of gametic diversity internally but it does not output the results. An option with a desired name for the output files is also required. An example of a typical example of parameter file follows:

```
NUMBER_OF_TRAITS
5
MAP_FILE
chromosome_inf
ALLELE_EFFECTS_FILE
allele_effect.txt
PHASED_GENOTYPE_FILE
phas_chr_1
RECOMBINATION_GROUP_FILE
group.txt
GENETIC_DISTANCE_UNIT
reco              ! reco or morgans
GAMETIC_VAR
TRUE              !T or TRUE
GAMETIC_(CO)VAR
T                 !T or TRUE
HOM
T                 !T or TRUE
CRV
T                 !T or TRUE
GEBV
T                 !T or TRUE
OUTPUT_NAME
out_name
```

### 3.2 - Chromosome information (MAP_FILE)

This file is determined by the user and specified in the parameter file. The file has to follow the format without header and delimited by single space: Chromosome (CHR),

SNP name (Name), SNP position (Position) and recombination rate/ genetic distance information (group1, group2 …) for the SNP. Only one chromosome can be specified in this file. The recombination rate/genetic distance information among the SNPs shall be as an accumulative function through the chromosome of SNP position, so that the number of recombination information in the file is equal to the number of SNP. At least one column with recombination information (only group1) shall be provided. The maximum number of recombination group columns is given by the number of animals, case the user has individual recombination rates. The number of columns for recombination groups shall be the same specified in the recombination group file as well the order of the columns has to be as the first in time descending direction that the group labeled appears in that file.




### 3.3 – Recombination group (RECOMBINATION_GROUP_FILE)

This is a reference to which information on recombination rate or genetic distance between the markers will be used per individual. The number of labeled group shall to match with the number columns in chromosome information file. The order of the first

appearance of the labeled groups shall be the same order of the columns with recombination rates/genetic distance in the chromosome information file. This file has the follow format: Label of group (group) up to 10 characters and ID.

```
group  ID
   F 1777663
   F 2030712
   F 2045243
   F 2053683
   M 2063056
   M 2083329
   M 2091842
```

### 3.4 – Allele effect information (ALLELE_EFFECTS_FILE)

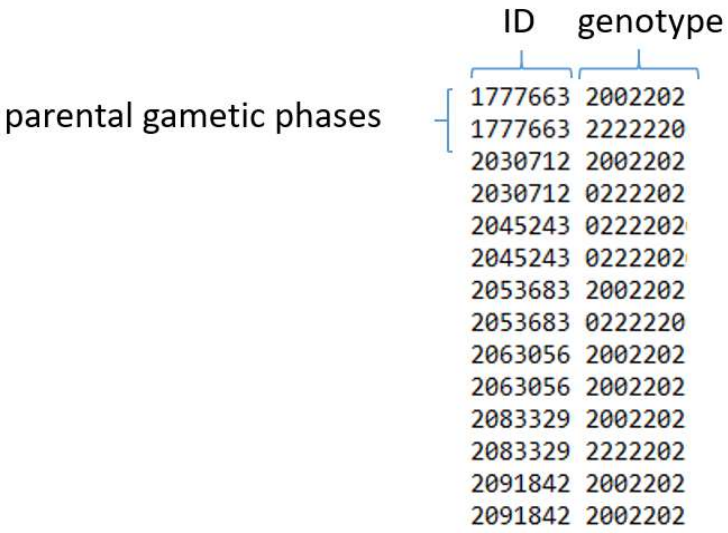
gamevar.f90 does not estimate the allele effects for the markers. The allele effects have to be estimated previously, using GBLUP or any differential shrinkage model (Santos et al., 2019). The file shall be without header with the maker solutions per trait in columns delimited for single space.

```
          Trait1                  Trait2                  Trait3                  Trait4                  Trait5
SNP1  -1.17343127381017 1.667909764491027E-002 -1.191073600074208E-003 2.412725221446055E-004 1.484545239407006E-004
SNP2  1.05663772239830 4.853192005618667E-002 1.457270709747225E-002 -6.512366982153934E-007 -1.055275140309635E-004
SNP3  3.17454541550206 8.815913254097595E-002 5.865056034425038E-002 -6.533357177128664E-005 -1.061936459157088E-004
SNP4  -2.60824702534779 -3.237894784948252E-002 -6.319524076100827E-002 2.147743764504157E-004 6.945128500048745E-005
SNP5  -4.19715575780482 -6.678589050138468E-002 -9.676841283230546E-002 2.335312875752295E-004 9.191707060728553E-005
SNP6  -1.12354005258381 -2.312092007178172E-003 -5.133673270317784E-002 1.480620193295882E-004 -7.090941288012509E-005
SNP7  0.632052607757951 4.311777918551202E-002 2.179373037200113E-002 7.813928006309311E-005 1.615337098252752E-005
```

### 3.5 – Phased genotype information (PHASED_GENOTYPE_FILE)

The genotype data must to be phased previously and the maternal and paternal gametic phase have to be ordered in rows tagged with same ID. The phased genotype

file follow the format ID and haplotype phase delimited by a single space. The haplotype phase has no space between the alleles coded as 0 for A1 and 2 for A2. Since the genotype data was already phased, no missing allele/genotyped code is required, therefore accepted by gamevar.f90.




## 4 – Run

gamevar.90 can be run on a terminal using a simple Linux command line with the parameter file name as an argument:

```
./gamevar {parameter file name}
```

The software runs only one chromosome per time. For run all chromosomes, or some groups of them, as well as split the running, the user can use a looping in Linux command like this:

```
for i in {1..29};  do
```

```
echo "NUMBER_OF_TRAITS
5
MAP_FILE
reco_chr$i.txt
ALLELE_EFFECTS_FILE
alle_effec_$i
PHASED_GENOTYPE_FILE
phas_chr_$i
RECOMBINATION_GROUP_FILE
group.txt
GENETIC_DISTANCE_UNIT
reco
GAMETIC_VAR
TRUE
GAMETIC_(CO)VAR
T
CRV
T
GEBV
T
OUTPUT_NAME
output_chr_$i" > parameter.txt
echo parameter.txt |./gamevar
done
```

# 5 - Outputs

## 5.1 - Log

A internal checking is printed on the screen, such as the options defined by user in the parameter file, initial data descriptions, warnings, stoppings and output mensages, as follows:

```
  *------------------------------------------------------------------------------*
  *----------------------   gamevar.f90 Version: 1.0  -------------------------*
  *------------------------------------------------------------------------------*
  * Authorship: Daniel Jordan de Abreu Santos                                    *
  * Algorithm basead on Santos, D.J.A et al., Journal of Dairy Sciences, 2019 *
  *------------------------------------------------------------------------------*

Current Data and Time: 17/06/2019 18:41:19


PARAMETE FILE:

OPTIONS:

Number of Traits:    5
Map File: chromosome_inf
Allele Effects File: allele_effect.txt
Phased Genotype File: phas_chr_1
Recombination Group File: group.txt
Genetic Distance Unit: Recombination rate

Estimate Gametic Variance = TRUE
Estimate Gametic (Co)Variance = TRUE
Estimate HOM = TRUE
Estimate CRV = TRUE
Estimate EBV = TRUE

output file: saidaT


DESCRIPTION:

Parameter File Readed:      param.txt
Number of SNPs:                    58
Number of Allele Effects:          58
Number of Genotypes:               58
Number of Reco Groups:              2
Number of Traits:                   5
Number of Animals:                 29

Calculating recombination matrix ........... group           1
Calculating recombination matrix ........... group           2

Calculating (co)variance of gametic diversity CHR:           1

Animal  ....................................          1
Animal  ....................................         10

Number of Haplotypes analysed:.....................          58

Outputs:
----------------------------------------------------------
  saidaT_COVAR has been created!
  saidaT_HOM has been created!
  saidaT_CRV has been created!
  saidaT_EBV has been created!
----------------------------------------------------------


The solutions have been calculated!!! Have a nice day !!!      ;)
```

## 5.2 - Output files

Five output files can be created by gamevar.90:

{output_name}_EBV = solutions for EBVs;

{output_name}_VAR = output with variances of gametic diversity

{output_name}_COVAR = output with variance and covariance of gametic diversity

{output_name}_CRV = output with CRV;

{output_name}_HOM = output with $\sum_i^{NHom} 2\alpha_i^2$.

All of them have the solutions per individual in rows and header with ID and trait number, i.e.TRAIT_{1,2,3..}. When the option GAMETIC_(CO)VAR isTRUE an extra number of column between the traits *i* and *j* with headers TRAIT_{$i_1$,$i_2$,$i_3$..$i_n$}_{$j_1$,$j_2$,$j_3$...$j_n$} are also written.

## Reference

Santos, D.J.A., J.B. Cole, J. B., T.J. Lawlor Jr, VanRaden, P.M., Tonhati, H. and Ma, L. Variance of gametic diversity and its application in selection programs. Journal of Dairy Science, 102(6): 5279-5294, 2019.