

# GENCONT-2

**A program for calculation of Optimal Genetic Contributions**

**User manual**

*Version 1.0*

Binyam S. Dagnachew and T.H.E. Meuwissen

Department of Animal and Aquacultural Sciences  
Norwegian University of Life Sciences (NMBU)  
N-1432 Ås  
Norway

Contact: [binyam.dagnachew@nmbu.no](mailto:binyam.dagnachew@nmbu.no)

*©Copyright 2015*

*Last update: June 2015*

## 1. Introduction

### a. Purpose

GENCONT-2 is a program to calculate optimal genetic contributions for selection candidates so that it maximizes the genetic gain whilst controlling the rate of inbreeding (managing genetic diversity). Inbreeding is managed by restricting the average relationship among selected parents because in a population inbreeding increases by 0.5 times the average relationship. GENCONT-2 uses a method that optimizes genetic contributions of selection candidates constraining on a predefined rate of inbreeding as described in Meuwissen (1997). GENCONT-2 is based on an iterative algorithm implementing the aforementioned method as presented in Dagnachew and Meuwissen (2014).

### b. Features

GENCONT-2 consists of a single program and all the required information to run the program, *input files* and their layout is specified in a *parameter file*. GENCONT-2 also incorporates calculation of optimum genetic contributions in the case of population with overlapping generation as described in Meuwissen and Sonesson (1998) and breeding schemes with multiple selection stages. The latter feature allows optimizing more than one selection stages simultaneously (for example, pre-selection of young animals to enter test station and final selection of sire after progeny testing).

The basic inputs are data file containing EBV of selection candidates and a pedigree file for computation of relationship among selection candidates. Some advance running options are described in the parameter file section. In the case of genomic selection, where animals are selected based on GEBVs instead of traditional EBVs, genomic relationship (G) matrix should be used in place of pedigree relationship (A) matrix (Sonesson, et.al., 2012). GENCONT-2 can take genomic relationship matrix (G) constructed outside the program and will use it to constrain relationship. In some cases, some candidate groups are genotyped and have GEBVs, and other groups are ungenotyped and have traditional EBVs. In situation like this, a hybrid of genomic and pedigree relationship matrices (i.e. H matrix) can be constructed (Legarra et al., 2009) and could be used in GENCONT-2 to constrain inbreeding.

## 2. Availability

More to come ...

## 3. Installation and running GENCONT-2

GENCONT-2 is distributed as pre-compiled executable file. The main target operating system is Linux. In addition, executable files for Windows environments can be available.

GENCONT-2 is expected to run from a command line interface, i.e. you need to open a ‘terminal’ window and type in the appropriate command, hitting return to start the program running. The general form of the command (under Linux) is:

**gencont\_2**    *parameter\_file*    *output\_file*

with the *parameter\_file* is the name of the ‘parameter file’ and the *output\_file* is the name of the ‘output file’ to which GENCONT-2 will write the output. If the *output\_file* is not provided, the default output name is ‘gencont\_2.out’.

## 4. Input files

The minimal requirement of GENCONT-2 are *parameter file*, *data file* and *pedigree file*.

*Parameter file*: this file contains all information that specifies everything GENCONT-2 needs to know about the input files and their outline and user running preferences (see parameter file section).

*Data file*: this file should at least contains ID and EBV or GEBV of the selection candidates. If more than one group (sex) of animals are included in the data file, then ‘sex’ of the individuals should be specified. Age-class and availability of the animals should be included in the file in the case of overlapping generation and multiple selection stages. Further, some maximum and minimum contributions that a specific animal can have, and fraction of offspring already allocated to a specific animal could be also included in the data file. The way to read the data file into GENCONT-2 is described under the parameter file section.

*Pedigree file:* Almost in all cases (the exception is when ‘*readin*’ option is activated, see parameter file under ‘*Amatrix*’ keyword), a pedigree file is required to calculate relationship among selection candidates. GENCONT-2 requires the pedigree file to be renumbered from 1 to N. The individuals ID provided in the data file should match the ID in the pedigree file.

## 5. Parameter file

GENCONT-2 uses similar parameter format as its previous version, Gencont. The parameter file is read line by line. Any line beginning with a ‘!’ and ‘%’ is considered to be a comment line and therefore skipped. GENCONT-2 relies on specific keywords (codes) at the beginning of each line to distinguish between different types of information given. The ordering of the lines in the ‘parameter file’ is arbitrary. The format of a line in the ‘parameter file’ looks like:

*<keyword> <parameter1 parameter2 ... >*

The keywords (codes) are:

### **ncand value**

*value* specifies the number of candidates in the ‘data\_file’. It is only used to allocate memory for the arrays and may thus be an overestimate of the actual number of candidates.

### **nped value**

*value* specifies the number of animals in the ‘pedigree file’. It is only used to allocate memory for the arrays and may thus be an overestimate of the actual number of individuals in the ‘pedigree file’.

### **ngroup value**

specifies the number of groups that are considered (e.g. if only male and female are presented in the data file then ‘ngroups’=2 and only males are included in the data then ‘ngroups’=1). In the case of multiple stages selection option, groups of selection candidates at one stage are considered belonging to similar group.

### **cmax value1 value2 value3 ... valueN**

specifies the maximum fraction of the offspring that could be contributed by anyone in group 1 (*value1*), anyone in group 2 (*value2*) and so on. *value1* = 0.0 or *value1* = 1.0 indicates that any fraction between 0 and 1 could be achieved. These are the default values.

**cmin** *value1 value2 ... valueN*

specifies the minimum number of the offspring that could be contributed by any individual in a group 1 (*value1*), anyone in group 2 (*value2*) and so on. Thus, in order for an individual (in any group) to be selected, it has to get at least a fraction of *value1*, *value2*, ... *valueN* (according to the groups requirement). Otherwise, it will get zero offspring (contribution). *value1*, *value2*, ... *valueN* = 0 indicates that any fraction between 0 and 1 could be achieved. This is the default value.

**n\_group** *value1/opt value2/opt value3/opt ... valueN/opt*

specifies (fixes) the number of selected individuals per group to respective *value*, and each individual in a specific group contributes a fraction of offspring of  $1/\textit{value}$ . If 'n\_group' is specified and no restriction is not implied in one of groups, use 'opt' instead of 'value'. For instance, we have two groups of candidates (1 and 2). We want to select  $N_2$  individuals from the 2<sup>nd</sup> group of candidates, but optimize group 1 candidates. We can specify this in the parameter file as 'n\_group opt  $N_2$ '. This forces to optimize candidates in group 1 and selects  $N_2$  individuals from group 2.

**constraint** *deltaf/relat/minim value*

This keyword specifies the constraint applied to the optimization process. The option *deltaf* indicates that the constraint should be on  $\Delta F$  (rate on inbreeding) and the desired rate of inbreeding  $\Delta F = \textit{value}$ . The option *relat* indicates that the constraint should be on the average relationship and that the desired average relationship is equal to *value*. The option *minim* indicates that the average relationship should be minimized (*values* is not needed for this option)

**overlap** *yes/no*

indicates overlapping generations/ no overlapping generations. Default is **overlap no**, i.e. no overlapping gene rations.

### **multistage run/post\_run**

This keyword specifies the presence of multiple selection stages. Options are:

**run:** this option indicates optimization of multiple selection stages simultaneously.

**post\_run:** this option indicates, one or more of the selection stage/s is/are completed beforehand and the information is used while optimizing the other selection stage/s.

### **Amatrix henderson/vanraden/readin filename full/half**

This keyword specifies a method to construct the numeric relationship matrix among selection candidates.

Options are:

**henderson:** specifies the use of Henderson's method (Henderson, 1976) to construct the relationship matrix.

**vanraden:** specifies the use VanRaden's method (VanRaden, 1992) to calculate relationship among selection candidates. This option requires including birth year of the animals in the pedigree file. For animals with missing (unknown) parents, the method set average inbreeding of their birth year as their inbreeding coefficient.

**readin:** this option indicates to read a relationship matrix from a file given by the '**filename**'. The '**filename**' is only required for the **readin** option. The relations matrix can be read as either a full-saved or half-saved matrix. The default option is full-saved, which can also be selected by specifying **full** after the '**filename**'. A half-saved matrix can be read into the program by specifying **half** after the '**filename**'.

In the case of genomic selection (when GEBVs are available instead of traditional EBVs), genomic relationship (i.e. G) matrix should be used in place pedigree relationship (i.e. A) matrix. If there are more than one group of selection candidates, one or more group/s is/are genotyped and has/have GEBV, and the other group/s is/are ungenotyped, then the pedigree-genomic relationship matrix (i.e. H matrix) could be used. The only

way to use G and H matrix in GENCONT-2 is by reading in the matrix to the program through '*readin*' option.

If the 'Amatrix' keyword is not set in the parameter file, the default option to construct the numeric relationship matrix is *henderson*.

**data file** *datafile input1 input2 ... inputN*

this code assigns the input file name to '*datafile*'. *input1, input2, ..., inputN* describes the N columns of the '*datafile*'.

**name:** name of the candidate (up to 25 characters)

**id:** ID number of the candidates. It must be similar to the ID number provided in the 'pedigreefile'.

**avail:** indicates whether the candidate is available for selection or not. It must be given as either 0 or 1. It is compulsory if **overlap** is *yes*.

**age:** age class<sup>1</sup> to which the candidate belongs

**sex:** indicates the sex of the candidate (compulsory if **nsex** 2)

**ebv:** ebv of the candidate

**c\_max:** specifies the maximum fraction of offspring that could be contributed by a candidate. This overrides any global values defines by the **cmax** keyword

**c\_min:** specifies the minimum fraction of offspring that could be contributed by a candidate. This overrides any global values defines by the **cmin** keyword

**c\_prev:** fraction of offspring already allocated to this candidate. In other word it specifies the compulsory usage of this candidate. Note that for each sex  $\text{SUM}(\text{c\_prev}) + \text{SUM}(\text{optimal contribution}) = 1$ . For example, if  $\text{c\_prve}(i) = 0.1$ , then the remaining candidates can only contribute 90% of the offspring. The candidate i may still be used on top of its *c\_prev* usage if *avail(i)=1*)

**ped file** *pedigreefile*

---

<sup>1</sup> age classes are user defined, and typically represent the time between 2 rounds of selection. If this time is 3 months, animals of age 1, 2, 3 months belong to age class 1, those of age 4,5, 6 months to age class 2 etc. If the age classes are very short, we should start counting ages from conception instead of from birth of the animals.



this code assigns the names of the pedigree file to *pedigreefile*. *pedigreefile* should contains 3 columns (i.e. animal\_ID, sire\_ID and dam\_ID)

## 6. Output

At the end of each run, GENCONT-2 generates at least two outputs, but produce more file depending on the running options. The output files are:

*Output\_file*: This main output file will contain summary of the run and assigned contributions to the candidates. The user can specify the name of the output file while running GENCONT-2 (see the installation and running section). However, the user can also choose to accept the default output name. If an *output\_file* is no specified, GENCONT-2 automatically generate an output file called '*gencont\_2.out*'.

*Error.log*: This additional output is also automatically produced for each run of GENCONT-2. The 'Error.log' is an empty file for un-problematic run; however, if the run encounter fatal error, it lists errors that terminate the program.

*Inbcoef.out*: This file will contain inbreeding coefficient of all animals in the pedigree file. This output file is only generated if the 'vanraden' option is selected for constructing of A matrix (refer parameter file under 'Amatrix' key word).

*Aveinb.out*: This file will contain the average inbreeding coefficient for the birth year of animals in the pedigree file. This output file is only generated if the 'vanraden' option is selected for constructing of A matrix (refer parameter file under 'Amatrix' key word).

## References

- Dagnachew, B.S. and T.H. Meuwissen. 2014. An Iterative Algorithm for Optimal Contribution Selection in Large Scale Breeding Program. *Proceeding of the 10<sup>th</sup> WCGALP*, 2014 Vancouver August, 023.
- Henderson, C. R. 1976. Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics* 32:69-83.
- Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *Journal of dairy science* 92(9):4656-4663.
- Meuwissen, T. H. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *Journal of animal science* 75(4):934-940.
- Meuwissen, T. H. and A. K. Sonesson. 1998. Maximizing the response of selection with a predefined rate of inbreeding: overlapping generations. *Journal of animal science* 76(10):2575-2583.
- Sonesson AK, Woolliams JA, and Meuwissen TH. 2012. Genomic selection requires genomic control of inbreeding. *Genetics, selection, evolution*: 44:27.
- VanRaden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large population. *Journal of Dairy Science* 75:3136-3144.

## Examples