

# Estimating “social” genetic effects using WOMBAT

## 1 Introduction

Over the last decade, mixed models incorporating indirect genetic effects due to individuals’ “social” or competitive effects on its neighbours or group members have become popular, especially for the analysis of plant breeding data and livestock kept in relatively small groups or housed closely together, such as pigs and poultry. Models suggested have progressed from initial implementations comprising a simple, additional genetic effect to more detailed models allowing for some degree of dilution of competitive effects depending on group size and non-genetic components through a random, independently distributed group effect (Muir, 2005; Bijma, 2010). Alternatively, a random regression model has been suggested to separate between components of social effects independent and dependent of group size (Hadfield and Wilson, 2007). There have been concerns about the identifiability of “social” genetic effects and estimability of the pertaining covariance components when the degree of competition between individuals is assumed to be constant, and differential weighting accounting for the degree of exposure or time spent in the same pen has been proposed as a more flexible alternative (Cantet and Cappa, 2008; Cappa and Cantet, 2008).

## 2 Fitting “social” genetic effects using WOMBAT

WOMBAT allows for the model suggested by Bijma (2010) to be fitted for univariate analyses. Upon specification of a group effect for each individual the program identifies the relevant competitors. “Social” genetic effects can be fitted as independent or as correlated with direct, additive genetic effects. In addition, the model can account for non-genetic competition effects by fitting an independently distributed random group effect. Both group and residual variances can be estimated separately for different group sizes.

Analyses are carried out estimating (co)variances for direct and “social” genetic effects for an assumed dilution factor,  $d$ . Estimation of (a non-zero)  $d$  thus requires several runs: Estimates of variance components together with the corresponding (log) likelihood value for given values of  $d$  represent points on the profile (log) likelihood for  $d$ . Hence a quadratic approximation of the profile likelihood, as described in Smith and Graser (1986) can be used to determine the value of  $d$  which maximizes the likelihood. At convergence, the quadratic coefficients also directly yield estimates of the sampling errors of  $\hat{d}$ .

### 2.1 What needs to be in the data file

- a) To fit both direct and “social” additive genetic effects, the data file should contain two identical columns with a code for the individuals’ identity<sup>1</sup>
- b) To identify individuals’ competitors, the data file has to contain a column with a code for the “social” group.
- c) To fit a random regression type model, the data file needs to contain a column to be used as ‘control variable’. This can be a dummy (integer) variable or, if heterogeneous

<sup>1</sup>Two columns are required as WOMBAT assumes a one-to-one mapping between a column in the data file and a random effect

residual or group variances are to be estimated a variable specifying group size or group size class.

## 2.2 Specifying a “social” effects model in the parameter file

### 2.2.1 Analysis type

The “social” effects model is implemented as a special type of random regression analysis. Hence this analysis type needs to be specified:

```
ANAL RR
```

### 2.2.2 Model block

To fit ‘standard’ model in a random regression setting, a single regression coefficient is fitted for each random effect, and the corresponding basis function is comprised simply of a coefficient of unity – WOMBAT provides the basis function option ONE to specify this.

The model needs to specify:

- The control variable to be used, in a line starting with RRC, e.g. size.
- The random effect modelling individuals’ *direct*, additive genetic effects, which needs to have covariance option NRM, e.g. RAN animal(1,ONE) NRM.
- The random effect modelling individuals’ *competitive*, additive genetic effects, which needs to have covariance option NRM, e.g. RAN compet(1,ONE) NRM.
- If non-genetic competitive effects are to be modelled by fitting a random group effect, this needs to be fitted as a random effect with covariance option IDE. In the simplest case, the corresponding variance is homogeneous and a single regression coefficient is fitted as for the genetic effects, e.g. RAN group(1,ONE) IDE. Alternatively, to estimate different variances for different levels of size, a separate regression coefficient needs to be fitted for each level, e.g. RAN group(3,IDE) IDE if size has 3 levels, where the basis function option IDE provides the coefficient of unity for each level and zeros otherwise.
- The effect specifying the “social” group. This can be a fixed effect in the model or an ‘extra’ effect (i.e. an effect which needs to be coded but is not fitted in the model). Typically, this would be an ‘extra’ effect, e.g. EXT group.

```
MODEL
  ...
  RRC size
  RAN animal(1,ONE) NRM
  RAN compet(1,ONE) NRM
  RAN group(1,ONE) IDE
  EXT group
  ...
END
```

### 2.2.3 Variance components

Starting values for variance components are given as for other analyses. Direct and competitive genetic effects can be estimated allowing for a covariance between them. When heterogeneous group variances are to be estimated, these should be specified as a diagonal covariance matrix of the appropriate dimension, e.g. VAR group 3 DIAG.

### 2.2.4 Special statement

Finally, an entry in a SPECIAL block is needed to instruct WOMBAT which parts of the model represent the “social” effects. This should consist of a single line, with the following, space separated entries:

- a) The code SOCIAL at the start of the line.
- b) The name of the random effect representing the “social” genetic effect. This can be abbreviated by omitting the (1, ONE) part.
- c) The name of the fixed or extra effect representing the “social” group.
- d) An INTEGER variable giving the maximum group size.
- e) Optional: A REAL variable giving the dilution factor  $d$  to be used. If omitted, a default value of  $d = 0$  (i.e. no dilution) is used.
- f) Optional: If a non-zero dilution factor has been given, an additional, four-letter code is recognised which specifies the “target group size” for which competitive genetic variance components are to be estimated. Codes available are NTWO which targets a group size of 2 (by using an entry of  $[1/(n-1)]^d$  in the corresponding design matrix) and NBAR which targets the average group size (using coefficients  $[(\bar{n}-1)/(n-1)]^d$  in the design matrix; see Bijma (2010)). If omitted, the default used is NTWO.

```

SPECIAL
    ...
    SOCIAL  compet  group  5  0.d0
    ...
END

```

## 2.3 Quadratic approximation

For each run, WOMBAT writes the value of the dilution factor used together with the corresponding maximum log likelihood to a file in the working directory named LogL4Quapprox.dat, appending the current pair of values to the end of the file. Note that runs which have not converged and require a continuation may also write such lines which need to be deleted manually.

Running WOMBAT with the option --quapp, these values are read and a quadratic approximation of the profile likelihood curve is attempted. Note that this relies on finding a valid triplet of pairs of values, i.e. the dilution factor with the highest likelihood should be bracketed by factors either side with lower values. If successful, WOMBAT copies the parameter file for the run to a new file wmbnew.par, substituting the new optimum value for  $d$  estimated for the original dilution factor, so that this file can be directly used

for the new estimation run. Typically, once the initial triplet has been established, 2 to 3 quadratic approximation steps suffice to estimate  $d$ .

## 2.4 Simulating data

WOMBAT can simulate data for “social” genetic effects models. However, a dilution factor, if specified, is only applied to the “social” genetic effects sampled (by multiplying with  $1/(n - 1)^d$ ).

## 3 Worked example

Use of WOMBAT with “social” effects models is illustrated in Example 16. Data used are simulated records for 1000 independent families of size 8, sampled assuming Bondari’s design (Bondari et al., 1978). Records are distributed over 2000 groups of size 3, 4 and 5.

Part A shows the simplest case of an analysis with uncorrelated direct and “social” genetic effects and homogeneous variances. Part B allows for non-zero direct-social genetic covariances, and part C illustrates how to fit heterogeneous variances for groups and residuals. Records used for these examples were simulated with a non-zero genetic covariance, but no dilution of competitive effects. Part D demonstrates the estimation of the dilution factor for records simulated with a factor of 0.8.

## References

- Bijma P. Multilevel selection 4: Modeling the relationship of indirect genetic effects and group size. *Genetics* 186 (2010) 1029–1031. doi: [10.1534/genetics.110.120485](https://doi.org/10.1534/genetics.110.120485).
- Bondari K., Willham R.L., Freeman A.E. Estimates of direct and maternal genetic correlations for pupa weight and family size of *Tribolium*. *J. Anim. Sci.* 47 (1978) 358–365.
- Cantet R.J.C., Cappa E.P. On identifiability of (co)variance components in animal models with competition effects. *J. Anim. Breed. Genet.* 126 (2008) 371–381. doi: [10.1111/j.1439-0388.2008.00743.x](https://doi.org/10.1111/j.1439-0388.2008.00743.x).
- Cappa E.P., Cantet R.J.C. Direct and competition additive effects in tree breeding: Bayesian estimation from an individual tree mixed model. *Silvae Genetica* 57 (2008) 45–55.
- Hadfield J.D., Wilson A.J. Multilevel Selection 3: Modeling the Effects of Interacting Individuals as a Function of Group Size. *Genetics* 177 (2007) 667–668. doi: [10.1534/genetics.107.075622](https://doi.org/10.1534/genetics.107.075622).
- Muir W.M. Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics* 170 (2005) 1247–1259. doi: [10.1534/genetics.104.035956](https://doi.org/10.1534/genetics.104.035956).
- Smith S.P., Graser H.U. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.* 69 (1986) 1156–1165.