

“WOMBAT” – DIGGING DEEP FOR QUANTITATIVE GENETIC ANALYSES BY RESTRICTED MAXIMUM LIKELIHOOD

Karin Meyer

Animal Genetics and Breeding Unit, University of New England (a joint venture with NSW Department of Primary Industries), Armidale, NSW 2351, Australia

INTRODUCTION

Availability of efficient software to estimate the (co)variance components in linear mixed models is an essential prerequisite for estimation of genetic parameters for traits of interest in livestock improvement schemes. One of the oldest programs used for restricted maximum likelihood (REML) estimation is DfREML, originally written in 1987 and updated until 2000 (Meyer, 1988, 1998). While reliable (and still quite useful), today the range of models accommodated by DfREML is too limited, the ‘interactive’ user interface of DfREML is outdated, and improved strategies for some computational techniques used have become available in the meantime.

WOMBAT replaces DfREML. Like its predecessor, WOMBAT is aimed at the analysis of data from animal breeding programmes. It accommodates most models commonly fitted for such data, and employs up-to-date methods for ordering the mixed model equations and maximising the likelihood. It is particularly suitable for analyses of large data sets fitting simple models.

COMPUTING METHODS

WOMBAT consists of a single program. Data and pedigree files, record layout, analysis type, the model to be fitted and starting values are specified in a parameter file. Numerous command line options allow different tasks to be selected or run time behaviour to be modified.

Analysis types. WOMBAT performs uni- and multivariate analyses for both standard and random regression models. For the latter, a choice of bases is available, including polynomial and B-spline functions. WOMBAT allows reduced rank estimation, by fitting the leading principal components for random effects only, but, at present, does not provide any facility to impose any other structure on estimated covariance matrices.

Models. WOMBAT allows multiple cross-classified fixed effects to be fitted, and one-way interactions between fixed effects to be defined. Covariables can be fitted crossclassified or within levels of a fixed effect, choosing between ordinary and orthogonal polynomials, and simple spline functions. Similarly, multiple random effects are accommodated. These can be distributed proportional to an identity matrix, the numerator relationship matrix (NRM) between animals or an arbitrary, user-supplied symmetric matrix.

Numerator relationship matrix. WOMBAT sets up the inverse of the NRM from a list of pedigree records, for either an animal or a sire-maternal grandsire model. For the animal model, pedigrees are first ‘pruned’, i.e. any animal without an expression of their phenotype and a link to only one other animal is deleted from the analysis. If a second, uncorrelated genetic effect is fitted, the pedigree for this effect is pruned separately, and a second, separate NRM is set up. Inbreeding coefficients are calculated using Tier (1990)’s algorithm.

Ordering strategies. REML analyses require repeated factorisation or inversion of the coeffi-

cient matrix in the mixed model equations. Re-ordering rows and columns in this large, sparse matrix to minimise the operations involved can have dramatic impact on computational requirements; see Meyer (2005) for details. WOMBAT provides multiple minimum degree, approximate minimum degree and multilevel nested dissection procedures for this task. Multiple orderings are readily trialled, choosing the method to be used explicitly or quantifying the degree of effort to be spent on this step.

Algorithms. WOMBAT incorporates average information (AI), expectation maximisation (EM), parameter expanded EM (PX-EM) and derivative-free (DF) algorithms. By default, analyses are begun using a small number of PX-EM iterates, then switching to the AI REML algorithm. This exploits the stability and good performance of the PX-EM algorithm in early iterates together with the fast convergence rates of the AI algorithm sufficiently close to the maximum of the likelihood (Meyer, 2006a). Run time options are available to choose between these algorithms, and specify the number of iterates allowed and criteria to define convergence.

Output. WOMBAT provides estimates of covariance components and corresponding correlations and variance ratios, together with approximate lower bounds sampling errors, derived from the average information matrix. At convergence, solutions for all fixed and random effects fitted are listed, and residuals for all observations are calculated and written to file for further analysis.

Extras. WOMBAT includes options to carry out a simple BLUP analysis only, and to replace observations for a given data set and covariance components by simulated values.

USE

WOMBAT has been used for multi-trait analyses of data on beef cattle, for example, a 14-trait analyses of carcass traits, estimating the leading principal components (Meyer, 2006a,b).

COMPUTING ENVIRONMENT

WOMBAT is written in standard FORTRAN 95 and is self-contained, except for public domain routines to reorder equations and generate random samples from the normal distribution. An executable program for Linux operating systems is available.

AVAILABILITY

WOMBAT is available for downloading at : <http://agbu.une.edu.au/~kmeyer/wombat.html>. Material available comprises the compiled Linux program, a manual and numerous worked examples. There are no license requirements, but it is a condition that a) use of WOMBAT is credited in any publications or reports, and b) that WOMBAT is not used commercially.

REFERENCES

- Meyer, K. (1988) *J. Dairy Sci.* **71** Supplement 2:33–34.
Meyer, K. (1998) DfREML version 3.0. CD-ROM Sixth World Congr. Genet. Appl. Livest. Prod.
Meyer, K. (2005) *Proc. Ass. Advan. Anim. Breed. Genet.* **16**:282–285.
Meyer, K. (2006a) *Proc. Eighth World Congr. Genet. Appl. Livest. Prod.* Comm. No. 24–00.
Meyer, K. (2006b) *Proc. Eighth World Congr. Genet. Appl. Livest. Prod.* Comm. No. 25–00.
Tier, B. (1990) *Genet. Select. Evol.* **22**:419–425.

ACKNOWLEDGEMENTS

This work was supported by the International Livestock Resources & Information Centre (www.ilric.com).