DMU - A Package For Analyzing Multivariate Mixed Models

P. Madsen[*], G. Su[*], R. Labouriau[*], O. F. Christensen[*]

Introduction

DMU is a package directed towards applications in quantitative genetics, and it implements powerful tools to estimate variance components and fixed effects (BLUE) and to predict random effects (BLUP). Most of the development of DMU has been in response to the needs in research projects in applied quantitative animal genetics along a period extending for more than 25 years. Indeed, DMU has been the main tool for statistical inference (both estimation and prediction) in the Danish animal breeding research. Moreover, DMU has been used in routine genetic evaluation of cattle, sheep, mink and horse in Denmark, and in several other countries. This led to a package implementing an ample range of statistical methods and computation algorithms, which not only implement high performance methods for specific project-related problems, but also are of general applicability in quantitative genetics.

The name DMU originates from a working name used for initial procedures in the package. These procedures were used to perform a **D**erivative-free approach to **MU**ltivariate analysis by Restricted Maximum Likelihood (REML), therefore the name DMU. As the current release of DMU does not include a module for DF-REML, the D now stands for DJF (the Danish abbreviation of the Faculty of Agricultural Sciences).

DMU modules

A DMU analysis always starts by executing the initializing DMU1 module followed by one of the other modules in the package. The current version contains the following modules:

DMU1: This is the initializing module for DMU analyses. The main tasks here are: To read model and the data description; to check the model and the data; and to write a re-coded data and model information onto disk for use by subsequent modules.

DMUAI:

This module performs single and multiple trait REML estimation of variance components using either Average Information (AI) or Expectation Maximization (EM) algorithms. If an AI maximization step evaluates the parameter vector outside the parameter space, a combined AI-EM or a pure EM step is performed (Jensen et al. (1997)).

DMUAI can handle random regression models, models with direct and maternal effects and models with dominance effects. It can also be used for multivariate QTL analysis (Sørensen

[*]Aarhus University, Faculty of Agricultural Sciences, Dept. Genetics and Biotechnology, Research Centre Foulum, P.O. Box 50, DK-8830 Tjele, Denmark

et al. (2003)). The variance component for each QTL is estimated based on IBD matrices provided by the user. It is possible to fit different types of Identical By Descent (IBD) models (gametic, genotypic and cluster), IBD methods (linkage, linkage disequilibrium and combined linkage and linkage disequilibrium) and genetic models (pleiotropic, close linkage, epistatic). Random effect with a specific correlation structure can be handled via a user supplied inverse correlation matrix. This can be used for genomic prediction based on the genomic relationship matrix.

The module also includes the possibility of working with multivariate generalized linear mixed models (GLMM), which allows to treat many cases of non-normally distributed data and non-linear responses. The distribution of the response variable for each dimension of a multivariate model can be specified to be an exponential dispersion model. The inference is based on penalized quasi-likelihood (Breslow and Clayton (1993); Schall (1991) and Wolfinger and O'Connell (1993)) computed by fitting iteratively suitable weighted REML estimations.

The implementation of the multivariate GLMM has the following distributions pre-specified: normal, gamma, inverse gaussian, binomial and Poisson. Moreover, the user can specify other exponential dispersion models by supplying the variance function and the deviance. Non-linear responses can also be specified (independently of the distribution) by defining a link function. The following pre-defined link functions are available: identity, logarithm, exponential, reciprocal, logit, probit and complementary-log-log. It is possible to define other link functions by supplying functions to compute the link, the derivative of the link and the response function (i.e. the inverse of the link). DMUAI allows to define multivariate models where a different distribution and/or a different link can be used for each dimension (i.e. for each trait).

DMU4: Single and multiple trait estimation of fixed effects (BLUE) and prediction of random effects (BLUP). DMU4 can handle the same type of models as DMUAI.

DMU4 can be used for prediction of genomic breeding value based on linear random regression model and have facilities for marker assisted BLUP (MA-BLUP). Two different MA-BLUP approaches are implemented:

1. A direct method as descried by Ferando and Grossman (1989)
2. An iterative method based on Jafarikia et al. (2006) extended to multi-variate and multi-QTL models.

The direct method requires a user supplied inverse IBD matrix, while the iterative approach requires a user supplied IBD matrix, which is much easier to compute than the inverse IBD matrix.

DMU4 is an in-core solver, where the Mixed Model Equation (MME) is built in core and solved either by:

1. Direct method based on FSPAK subroutines (Perez-Enciso et al. (1994))
2. Direct methods based on subroutines from the Intel® Math Kernel Library using parallel execution on SMP Linux workstations
3. Iteratively based on ITPACK subroutines (Kincaid et al. (1982))

Memory requirement for holding the MME and needed work space together set the limit on the size of problems that can be handled. The work space needed by the direct the methods is larger than for the iterative methods, so the iterative solver can handle larger problems then the direct solvers.

DMU5: Single and multiple trait BLUP based on Preconditioned Conjugated Gradient (PCG) iteration on data. DMU5 can handle the same type of models as DMU4 except MA-BLUP. Because the MME is not built and stored in core, DMU5 it can handle much larger problems (>300.000.000 equations) than DMU4.

DMU5 has a build-in data buffer system, where all data or as large a part of the data as possible are stored in the available physical memory. In this way the amount of physical input operations is reduced and computational speed is improved.

RJMC: Single and multiple trait Markov Chain Monte Carlo (MCMC) based estimation of location and dispersion parameters for Gaussian, binary, ordered categorical, 2 component mixture (Ødegård et al. (2005)), zero inflated sequential binary traits (Ødegård et al. (2009); Ødegård et al. (2010b)) and reaction norm models where the environmental gradient is unknown (Su et al. (2006)). RJMC can handle models with heterogeneous residual variance (in groups).

Sampling of genetic dispersion parameters can be restricted in a way that only information on parental animals are used. This approach has been shown very useful for cross-sectional binary traits (Ødegård et al. (2010a); Ødegård et al. (2010b)).

DmuTrace: An independent program for checking the consistency of a pedigree file and for extracting a subset of a large pedigree file. The program generated a full pedigree and a pruned pedigree, where non-informative individuals are removed. The output files are in a format so they can be used as input files for DMU. DmuTrace can also compute gene contribution from base population to each individual as well as each individuals base population heterozygosity.

Computer Environment and Availability

DMU is written in Fortran 90/95 and has been developed under Linux on 32 and 64 bit based workstations. It has been ported to a variety of UNIX systems and to Windows (32 and 64 bit).

R interface: An R interface to DMU has been developed with the purpose of facilitating the specification of models and data. Using the interface, the model is specified by in the R formula language, DMU is executed and the results are returned to R. Functions to summarize the results and to extract various components of the analysis are also provided. In this way the R interface to DMU makes possible to combine the computational power of DMU with the various and flexible tools available within R.

As an example, the scripting facilities of R makes it easy to use DMU to fit mixed models within a loop (for example to detect QTLs) and the graphical facilities of R makes it easy to visualize results.

At the moment only the DMUAI module is supported (with some restrictions).

Availability: The DMU packaged is distributed as executable files (http://dmu.agrsci.dk) and is free of charge for research purposes. There are no charges for using DMU for research purposes, but DMU should be referred in publications by citing the "DMU Users Guide". The terms of conditions for commercial use (*e.g.* routine genetic evaluation) can be obtained from the Department of Genetics and Biotechnology at the Faculty for Agricultural Sciences (DJF), Aarhus University (email:per.madsen@agrsci.dk).

References

Breslow, N. E. and Clayton, D. G. (1993). *J. American Stat. Assoc.*, 88:9–25.

Ferando, R. L. and Grossman, M. (1989). *Genet. Sel. Evol.*, 21:467–477.

Jafarikia, M., Susanto, A., Robinson, J. A. B. et al. (2006). In *CD communication 8th WCGALP*, volume 22, page 2.

Jensen, J., Mantysaari, E. A., Madsen, P. et al. (1997). *J. Indian Soc. Agr. Stat.*, 49:215–236.

Kincaid, D. R., Respess, J. R., Young, D. M. et al. (1982). *ACM TOMS*, 8:302–322.

Ødegård, J., Madsen, P., Gianola, D. et al. (2005). *J. Dairy Sci.*, 88:2652–2659.

Ødegård, J., Madsen, P., Labouriau, R. et al. (2009). *J. Anim. Sci.*, (submitted).

Ødegård, J., Meuwissen, T. H. E., Heringstad, B. et al. (2010a). *Genet. Sel. Evol.*, (submitted).

Ødegård, J., Yazdi, M. H., Gitterle, T. et al. (2010b). *This Conference*.

Perez-Enciso, M., Misztal, I., and Elzo, M. A. (1994). In *Proc. 5th WCGALP*, volume XXII, pages 85–85.

Schall, R. (1991). *Biometrika*, 78(4):719–727.

Sørensen, P., Lund, M. S., Guldbrandtsen, B. et al. (2003). *Genet. Sel. Evol.*, 35(6):605–622.

Su, G., Madsen, P., Lund, M. S. et al. (2006). *J. Anim. Sci.*, 84:1651–1657.

Wolfinger, R. and O'Connell, M. (1993). *J. Stat. Computation and Simulation*, 48:233–343.