"DXMRR" - A PROGRAM TO ESTIMATE COVARIANCE FUNCTIONS FOR LONGITUDINAL DATA BY RESTRICTED MAXIMUM LIKELIHOOD

Karin Meyer

Animal Genetics and Breeding Unit, University of New England,
Armidale, NSW 2351, Australia

OBJECTIVE

Covariance functions (CF) are, in essence, the 'infinite-dimensional' equivalent to covariance matrices for traits measured repeatedly, potentially infinitely many times, along some continuous scale such as time or age. CFs produce a continuous description of the covariance structure over time, giving the covariance between any two ages as a function of the ages. DxMRR allows the estimation of genetic, environmental and phenotypic CFs directly from the data by Restricted Maximum Likelihood (REML), fitting an animal model.

COMPUTING METHODS

Full details of the model of analysis and computational procedure are given by Meyer (1997); see also Meyer (1998), this conference.

CFs can be estimated by fitting a mixed 'linear' model with regression coefficients on some function of time as *random* effects. A set of random regression (RR) coefficients is fitted for each source of variation to be modelled. The coefficients of the CFs are estimated as the covariances among the RR coefficients, using a 'standard' multivariate REML algorithm; options in DxMRR include a derivative-free (Simplex or Powell) search and an 'average information' Newton-Raphson type procedure.

Regressing on orthogonal (Legendre) polynomials of time yields the CF model described by Kirkpatrick *et al.* (1990). This is the default in DxMRR, but other, user-defined RR functions can be fitted. It is assumed that the order of polynomial fit is reduced, i.e., less than the number of time points in the data. This allows temporary and permanent environmental variances to be estimated. It implies that at least two sets of RR coefficients, modelling additive genetic and permanent environmental effects due to the animal, are fitted and corresponding CF, $\mathcal{A}$ and $\mathcal{R}$ are estimated. For animal i with records y_{ij} taken at time t_{ij} this gives model of analysis

$$y_{ij} = F + \sum_{m=0}^{k_A-1} \alpha_{im}\phi_m(t_{ij}) + \sum_{m=0}^{k_R-1} \gamma_{im}\phi_m(t_{ij}) + \epsilon_{ij} \tag{1}$$

with F denoting some fixed effects, ϵ_{ij} the temporary environmental effect on y_{ij}, α_{im} and γ_{im} representing the $m-$th additive genetic and permanent environmental random regression coefficients for animal i, respectively, and k_A and k_R denoting the respective orders of fit. Additional random effects can be taken into account by fitting corresponding RR coefficients. Temporary environmental or measurement error variances are assumed to be independently distributed; DxMRR allows for single or multiple error variances to be estimated.

DxMRR accommodates different orders of fit for different, uncorrelated CF, thus facilitating the minimum degree of polynomial fit for each CF to be selected. Estimation can be carried out on the original scale or by reparameterising to the Cholesky decompositions of the covariance matrices of RR coefficients. The latter allows the estimation of reduced rank covariance matrices (of

RR) and thus CFs. For a CF with order of fit k and rank m this reduces the number of parameters to be estimated from $k(k+1)/2$ to $km - m(m-1)/2$. It allows parsimonious modelling when a high degree (k) polynomial is needed to describe the shape of trajectory adequately, but only the first m eigenvalues and eigenfunctions of the CF are important.

Output from DxMrr includes the estimated covariances among RR coefficients, the resulting CF and their eigenvalues and eigenfunctions. Optionally, covariance and correlation matrices for the ages in the data are calculated from the estimated CF. Run options are provided to generate a grid of predicted covariances for selected time points, and to obtain backsolutions for the RR coefficients fitted. For the genetic CF, these are predicted breeding values for the shape parameters of the trajectory, and predicted breeding values for selected time points can be calculated from them.

USE

DxMrr can be used for the analysis of any repeated records taken along some continuous scale. The most common 'meta-meter' is time or age, but other scales such as size or weight are possible. Currently, DxMrr performs 'univariate' analyses only (e.g. weight measured at different ages), but an extension to multivariate models (e.g., height and weight for age) is planned. To date, DxMrr has been used in the analysis of test day records in dairy cattle (Van der Werf *et al.*1997) and to analyse records for mature weights of beef cows (Meyer 1997).

AVAILABILITY

DxMrr is part of DfReml (version 3.0). Release is planned for late 1997. Both DxMrr and DfReml $v3.0$ are available free of charge to the scientific community, provided their use is being credited. As for previous versions of DfReml, source code for the programs (complete with Makefiles for a Unix environment), documentation and a worked example can be downloaded from the DfReml home page, `http://metz.une.edu.au/~kmeyer/dfreml.html`. Alternatively, all material can be obtained via anonymous ftp from `metz.une.edu.au` (129.180.1.4), directory `/pub/agbu/DFREML` (see `README` file for details).

COMPUTING ENVIRONMENT

DxMrr is written in standard Fortran 90 and is self-contained, except for calls to system routines to monitor CPU time and dates. It has been developed under an Unix environment, using the DEC f90 compiler, and also been compiled on Sparc stations using f90 (Solaris) and epcf90 (Edinburgh Portable Compiler) compilers. A PC version is planned.

REFERENCES

Kirkpatrick, M., Lofsvold, D. and Bulmer, M. 1990. *Genetics* **124** : 979-993.
Meyer, K. 1997. *Livest. Prod. Sci.*(submitted).
Meyer, K. 1998. 6th *World Congr. Genet. Appl. Livest. Prod.* (submitted).
Van der Werf, J., Goddard, M. and Meyer, K. 1997. *J. Dairy Sci.* (submitted).