Original article

Genetic variation of traits measured in several environments. II. Inference on between-environment homogeneity of intra-class correlations

C Robert, JL Foulley, V Ducrocq

Institut national de la recherche agronomique, station de génétique quantitative et appliquée, centre de recherche de Jouy-en-Josas, 78352 Jouy-en-Josas cedex, France

Summary – This paper describes a further contribution to the problem of testing homogeneity of intra-class correlations among environments in the case of univariate linear models, without making any assumption about the genetic correlation between environments. An iterative generalized expectation-maximization (EM) algorithm, as described in Foulley and Quaas (1994), is presented for computing restricted maximum likelihood (REML) estimates of the residual and between-family components of variance and covariance. Three different parameterizations (cartesian, polar and spherical coordinates) are proposed to compute EM-REML estimators under the reduced (constant intra-class correlation between environments) model. This procedure is illustrated with the analysis of simulated data.

heteroskedasticity / parameterization / intra-class correlation / expectation-maximization / restricted maximum likelihood

Résumé – Variation génétique de caractères mesurés dans plusieurs milieux. II. Inférence relative à des corrélations intra-classe constantes entre milieux. *Cet article décrit une approche permettant d'estimer les composantes de variance-covariance entre milieux dans le cas de corrélation intra-classe homogènes entre milieux, sans faire d'hypothèse sur les corrélations génétiques entre milieux pris 2 à 2. Un algorithme itératif d'espérance-maximisation (EM), comparable à celui décrit par Foulley et Quaas (1994), est proposé pour calculer les estimations du maximum de vraisemblance restreinte (REML) des composantes résiduelles et familiales de variance covariance. Trois paramétrisations différentes (coordonnées cartésiennes, polaires et sphériques) sont proposées pour calculer les estimateurs EM-REML sous le modèle réduit (les corrélations intra-classe sont supposées toutes égales à une même constante). Cette procédure est illustrée par l'analyse de données simulées.*

hétéroscédasticité / paramétrisation / corrélation intra-classe / espérance-maximisation / maximum de vraisemblance restreinte

INTRODUCTION

Statistical procedures based on the theory of the generalized likelihood ratio, previously proposed by Foulley *et al* (1994), Shaw (1991) and Visscher (1992), have been applied to test the homogeneity of genetic and phenotypic parameters against Falconer's (1952) saturated model. In particular, Robert *et al* (1995) have described a procedure for estimating components of variance and covariance between environments and for testing the homogeneity of the following parameters: (a) a constant genetic correlation between environments; and (b) constant genetic and intra-class correlations between environments.

The objective of this article is to present a procedure for dealing with homogeneous intra-class correlations among environments without making any assumption about the genetic correlations between environments. The method is based on restricted maximum likelihood estimators (REML) and on a generalized expectation-maximization (EM) algorithms as proposed initially by Foulley and Quaas (1994) for heteroskedastic univariate linear models. Three parameterizations of variance-covariance components are suggested for solving this problem. A simulated example is presented to illustrate this procedure.

THEORY

A model often used to deal with genotypic variation in different environments is the 2-way crossed genotype (random) × environment (fixed) linear model with interaction. In particular, this model has been proposed as an alternative to a multiple-trait approach when variance and covariance components are homogeneous and genetic correlations between environments are positive (Foulley and Henderson, 1989). It has also been employed by Visscher (1992) to study the power of likelihood ratio tests for heterogeneity of intra-class correlations between environments when genetic correlations among them are assumed equal to unity. The aim of this paper is to go one step further in addressing the same problem with the same model but with a heterogeneous structure of variance-covariance components.

The full model

Let us assume that records are generated from a cross-classified layout. The model is defined as follows:

$$y_{ijk} = \mu + h_i + \sigma_{s_i} s_j^* + \sigma_{hs_i} hs_{ij}^* + e_{ijk} \qquad [1]$$

where μ is the mean, h_i is the fixed effect of the ith environment: $\sigma_{s_i} s_j^*$ is the random family j contribution such that $s_j^* \sim \text{NID}(0,1)$ and $\sigma_{s_i}^2$ is the family variance for records in the ith environment; $\sigma_{hs_i} hs_{ij}^*$ is the random family × environment interaction effect such that $hs_{ij}^* \sim \text{NID}(0,1)$ and $\sigma_{hs_i}^2$ is the interaction variance for records in the ith environment; e_{ijk} is the residual effect assumed $\text{NID}(0, \sigma_{e_i}^2)$. Remember that this model has been extensively used in factor analysis of psychological data (Lawley and Maxwell, 1963).

Model [1] can be written more generally using matrix notation as:

$$\mathbf{y}_i = \mathbf{X}_i\beta + \sigma_{u_{1,i}}\mathbf{Z}_{1i}\mathbf{u}_1^* + \sigma_{u_{2,i}}\mathbf{Z}_{2i}\mathbf{u}_2^* + \mathbf{e}_i \tag{2}$$

where $\mathbf{y}_i$ is a $(n_i \times 1)$ vector of observations in environment i; β is a $(p \times 1)$ vector of fixed effects with incidence matrix $\mathbf{X}_i$; $\mathbf{u}_1^* = \{s_j^*\}$ and $\mathbf{u}_2^* = \{hs_{ij}^*\}$ are 2 independent random normal components of the model with incidence matrices for standardized effects $\mathbf{Z}_{1i}$ and $\mathbf{Z}_{2i}$ respectively; $\sigma_{u_{1,i}}^2$ and $\sigma_{u_{2,i}}^2$ are the corresponding components of variance, pertaining to stratum i and $\mathbf{e}_i$ is the vector of residuals for stratum i assumed $N(\mathbf{0}, \sigma_{e_i}^2\mathbf{I}_{n_i})$.

The reduced model

The null hypothesis (H_0) consists of assuming homogeneous intra-class correlations between environments ($ie, \forall\, i, t_i = (\sigma_{s_i}^2 + \sigma_{hs_i}^2) / (\sigma_{s_i}^2 + \sigma_{hs_i}^2 + \sigma_{e_i}^2) = t$). The variance-covariance structure of the residual is assumed to be diagonal and heteroskedastic. Under model [1], this hypothesis is tantamount to assuming a constant ratio of variances between environments: $\forall\, i, \sigma_{e_i}^2 / (\sigma_{s_i}^2 + \sigma_{hs_i}^2) = \delta^2$, where δ is a constant.

Under this hypothesis, 3 different parameterizations will be considered to solve this problem.

Cartesian coordinates

$$\forall\, i, \sigma_{e_i}^2 = \delta^2(\sigma_{s_i}^2 + \sigma_{hs_i}^2)$$

where δ is a positive real number.

Polar coordinates

$$\forall\, i, \sigma_{s_i}^2 = \rho_i^2\cos^2\theta_i$$
$$\forall\, i, \sigma_{hs_i}^2 = \rho_i^2\sin^2\theta_i$$
$$\forall\, i, \sigma_{e_i}^2 = \delta^2\rho_i^2$$

where ρ_i and δ are positive real numbers.

Spherical coordinates

$$\forall\, i, \sigma_{s_i}^2 = \psi_i^2(\cos^2\alpha)(\cos^2\varphi_i)$$
$$\forall\, i, \sigma_{hs_i}^2 = \psi_i^2(\cos^2\alpha)(\sin^2\varphi_i)$$
$$\forall\, i, \sigma_{e_i}^2 = \psi_i^2(\sin^2\alpha)$$

where Ψ_i is a positive real number. Under this parameterization $\delta^2 = \tan^2\alpha$.

An EM-REML algorithm

A generalized expectation-maximization (EM) algorithm to compute REML estimators is applied (Foulley and Quaas, 1994). As in Robert *et al* (1995) and for heteroskedastic mixed models, the function to be maximized is:

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) = \text{Const} - (1/2) \sum_{i=1}^{p} n_i \ln(\sigma_{e_i}^2) - (1/2) \sum_{i=1}^{p} \sigma_{e_i}^{-2} E_c^{[t]}[\mathbf{e}_i'\mathbf{e}_i] \qquad [3]$$

where $\boldsymbol{\gamma}$ is the set of estimable parameters for each of the 3 models (under each parameterization considered). $E_c^{[t]}[.]$ represents the conditional expectation taken with respect to the distribution of fixed and random effects given the data vector and $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}$. $E_c^{[t]}[.]$ can be expressed as a function of bilinear forms and a trace of parts of the inverse coefficient matrix of the mixed-model equations (as described in Foulley and Quaas, 1994). So, for each parameterization, we derive function [3] with respect to each parameter of $\boldsymbol{\gamma}$ and we solve the resulting system $\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) / \partial \boldsymbol{\gamma} = 0$. After some algebra and using the method of 'cyclic ascent' (Zangwill, 1969), we obtain the 3 following algorithms.

For model [2] and using cartesian coordinates, the algorithm at iteration $[t, l+1]$ can be summarized as follows. Let $\delta^{2[t,l]}$, $\sigma_{u_{1,i}}^{[t,l]}$ and $\sigma_{u_{2,i}}^{[t,l]}$ be the values at iteration $[t, l]$. The next iterates are obtained as:

$$\bullet \ \delta^{2[t,l+1]} = \frac{\sum_{i=1}^{p} E_c^{[t]}[\mathbf{e}_i^{'[t,l]}\mathbf{e}_i^{[t,l]}] / (\sigma_{u_{1,i}}^{2[t,l]} + \sigma_{u_{2,i}}^{2[t,l]})}{\sum_{i=1}^{p} n_i}$$

where $\mathbf{e}_i^{[t,l]} = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{1,i}}^{[t,l]}\mathbf{Z}_{1i}\mathbf{u}_1^* - \sigma_{u_{2,i}}^{[t,l]}\mathbf{Z}_{2i}\mathbf{u}_2^*$

$\bullet$ $\sigma_{u_{1,i}}^{[t,l+1]}$ is the only positive root of the following cubic equation:

$$a\sigma_{u_{1,i}}^{3[t,l+1]} + b\sigma_{u_{1,i}}^{2[t,l+1]} + c\sigma_{u_{1,i}}^{[t,l+1]} + d = 0$$

with

$a = -n_i\delta^{2[t,l+1]}$

$b = -E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{2,i}}^{[t,l]}\mathbf{Z}_{2i}\mathbf{u}_2^*)'\mathbf{Z}_{1i}\mathbf{u}_1^*]$

$c = E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{2,i}}^{[t,l]}\mathbf{Z}_{2i}\mathbf{u}_2^*)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{2,i}}^{[t,l]}\mathbf{Z}_{2i}\mathbf{u}_2^*)] - \sigma_{u_{2,i}}^{2[t,l]} E_c^{[t]}[\mathbf{u}_1^{*'}\mathbf{Z}_{1i}'\mathbf{Z}_{1i}\mathbf{u}_1^*]$
$\qquad - n_i\delta^{2[t,l+1]}\sigma_{u_{2,i}}^{2[t,l]}$

$d = \sigma_{u_{2,i}}^{2[t,l]} E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{2,i}}^{[t,l]}\mathbf{Z}_{2i}\mathbf{u}_2^*)'\mathbf{Z}_{1i}\mathbf{u}_1^*]$

$\bullet$ $\sigma_{u_{2i}}^{[t,l+1]}$ is the only positive root of the following cubic equation:

$$a\sigma_{u_{2,i}}^{3[t,l+1]} + b\sigma_{u_{2,i}}^{2[t,l+1]} + c\sigma_{u_{2,i}}^{[t,l+1]} + d = 0$$

with

$$a = -n_i \delta^{2[t,l+1]}$$

$$b = -E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{1,i}}^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\mathbf{Z}_{2i}\mathbf{u}_2^*]$$

$$c = E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{1,i}}^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{1,i}}^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*)]$$
$$- \sigma_{u_{1,i}}^{2[t,l+1]}E_c^{[t]}[\mathbf{u}_2^{*'}\mathbf{Z}_{2i}'\mathbf{Z}_{2i}\mathbf{u}_2^*] - n_i\delta^{2[t,l+1]}\sigma_{u_{1,i}}^{2[t,l+1]}$$

$$d = \sigma_{u_{1,i}}^{2[t,l+1]}E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{1,i}}^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\mathbf{Z}_{2i}\mathbf{u}_2^*]$$

For model [2] and polar coordinates, the algorithm at iteration $[t, l+1]$ can be summarized as follows. Let $\delta^{2[t,1]}$, $\rho_i^{[t,l]}$ and $\theta_i^{[t,l]}$ be the values at iteration $[t, l]$. The next iterates are obtained as:

- $$\delta^{2[t,l+1]} = \frac{\sum_{i=1}^{p} E_c^{[t]}[\mathbf{e}_i^{'[t,l]}\mathbf{e}_i^{[t,l]}] / \rho_i^{2[t,l]}}{\sum_{i=1}^{p} n_i}$$

with $\mathbf{e}_i^{[t,l]} = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \rho_i^{[t,l]}\cos\theta_i^{[t,l]}\mathbf{Z}_{1i}\mathbf{u}_1^* - \rho_i^{[t,l]}\sin\theta^{[t,l]}\mathbf{Z}_{2i}\mathbf{u}_2^*$

- $\rho_i^{[t,l+1]}$ is the only positive root of the following quadratic equation:

$$a\rho_i^{2[t,l+1]} + b\rho_i^{[t,l+1]} + c = 0$$

with:

$$a = n_i$$

$$b = (1/\delta^{2[t,l+1]})E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\cos\theta_i^{[t,l]}\mathbf{Z}_{1i}\mathbf{u}_1^* + \sin\theta_i^{[t,l]}\mathbf{Z}_{2i}\mathbf{u}_2^*)]$$

$$c = -(1/\delta^{2[t,l+1]})E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})]$$

- $\theta_i^{[t,l+1]}$ is the solution of the equation $\tau_i^{[t,l+1]} = \tan(\theta_i^{[t,l+1]}/2)$

where $\tau_i^{[t,l+1]}$ is the only positive root of the quartic equation:

$$a\tau_i^{4[t,l+1]} + 2b\tau_i^{3[t,l+1]} - 6c\tau_i^{2[t,l+1]} + 2d\tau_i^{[t,l+1]} + e = 0$$

with:

$$a = E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} + \rho_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\rho_i^{[t,l+1]}\mathbf{Z}_{2i}\mathbf{u}_2^*]$$

$$b = E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} + \rho_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\rho_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*] - \rho_i^{2[t,l+1]}E_c^{[t]}[\mathbf{u}_2^{*'}\mathbf{Z}_{2i}'\mathbf{Z}_{2i}\mathbf{u}_2^*]$$

$$c = \rho_i^{2[t,l+1]}E_c^{[t]}[\mathbf{u}_1^{*'}\mathbf{Z}_{1i}'\mathbf{Z}_{2i}\mathbf{u}_2^*]$$

$$d = E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \rho_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\rho_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*] + \rho_i^{2[t,l+1]}E_c^{[t]}[\mathbf{u}_2^{*'}\mathbf{Z}_{2i}'\mathbf{Z}_{2i}\mathbf{u}_2^*]$$

$$e = -E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \rho_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\rho_i^{[t,l+1]}\mathbf{Z}_{2i}\mathbf{u}_2^*]$$

For model [2] and spherical coordinates, the algorithm at iteration $[t, l+1]$ can be summarized as follows. Let $\psi_i^{[t,l]}$, $\varphi_i^{[t,l]}$ and $\alpha^{[t,l]}$ the values at iteration $[t,l]$. The next iterates are obtained as:

- $\psi_i^{[t,l+1]}$ is the only positive root of the following quadratic equation:

$$a\psi_i^{2[t,l+1]} + b\psi_i^{[t,l+1]} + c = 0$$

with:

$a = n_i \sin^2 \alpha^{[t,l]}$

$b = \cos \alpha^{[t,l]} E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\cos \varphi_i^{[t,l]}\mathbf{Z}_{1i}\mathbf{u}_1^* + \sin \varphi_i^{[t,l]}\mathbf{Z}_{2i}\mathbf{u}_2^*)]$

$c = -E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})]$

- $\varphi_i^{[t,l+1]}$ is the solution of the equation $\tau_i^{[t,l+1]} = \tan(\varphi_i^{[t,l+1]}/2)$ where $\tau_i^{[t,l+1]}$ is the only positive root of the quartic equation:

$$a\tau_i^{4[t,l+1]} + 2b\tau_i^{3[t,l+1]} + 6c\tau_i^{2[t,l+1]} + 2d\tau_i^{[t,l+1]} + e = 0$$

with:

$a = E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \psi_i^{[t,l+1]}\cos \alpha^{[t,l]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\mathbf{Z}_{2i}\mathbf{u}_2^*]$

$b = E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} + \psi_i^{[t,l+1]}\cos \alpha^{[t,l]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\mathbf{Z}_{1i}\mathbf{u}_1^*]$
$\quad - \psi_i^{[t,l+1]}\cos \alpha^{[t,l]} E_c^{[t]}[\mathbf{u}_2^*\mathbf{Z}_{2i}'\mathbf{Z}_{2i}\mathbf{u}_2^*]$

$c = -\psi_i^{[t,l+1]}\cos \alpha^{[t,l]} E_c^{[t]}[\mathbf{u}_1^{*'}\mathbf{Z}_{1i}'\mathbf{Z}_{2i}\mathbf{u}_2^*]$

$d = E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \psi_i^{[t,l+1]}\cos \alpha^{[t,l]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\mathbf{Z}_{1i}\mathbf{u}_1^*]$
$\quad + \psi_i^{[t,l+1]}\cos \alpha^{[t,l]} E_c^{[t]}[\mathbf{u}_2^{*'}\mathbf{Z}_{2i}'\mathbf{Z}_{2i}\mathbf{u}_2^*]$

$e = -E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \psi_i^{[t,l+1]}\cos \alpha^{[t,l]}\mathbf{Z}_{1i}\mathbf{u}_1^*)'\mathbf{Z}_{2i}\mathbf{u}_2^*]$

- $\alpha^{[t,l+1]}$ is the solution of the equation $\kappa_i^{[t,l+1]} = \tan^2(\alpha^{[t,l+1]}/2)$ where $\kappa_i^{[t,l+1]}$ is the only positive root of the cubic equation:

$$a\kappa_i^{3[t,l+1]} - b\kappa_i^{2[t,l+1]} + c\kappa_i^{[t,l+1]} + d = 0$$

with:

$a = -\sum_{i=1}^{p}(1/\psi_i^{2[t,l+1]})E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} + \psi_i^{[t,l+1]}\cos \varphi_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*$
$\quad + \psi_i^{[t,l+1]}\sin \varphi_i^{[t,l+1]}\mathbf{Z}_{2i}\mathbf{u}_2^*)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} + \psi_i^{[t,l+1]}\cos \varphi_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*$
$\quad + \psi_i^{[t,l+1]}\sin \varphi_i^{[t,l+1]}\mathbf{Z}_{2i}\mathbf{u}_2^*)]$

$b = -4\sum_{i=1}^{p}n_i - \sum_{i=1}^{p}(1/\psi_i^{2[t,l+1]})E_c^{[t]}[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} + \psi_i^{[t,l+1]}\cos \varphi_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*$
$\quad + \psi_i^{[t,l+1]}\sin \varphi_i^{[t,l+1]}\mathbf{Z}_{2i}\mathbf{u}_2^*)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} + \psi_i^{[t,l+1]}\cos \varphi_i^{[t,l+1]}\mathbf{Z}_{1i}\mathbf{u}_1^*$
$\quad + \psi_i^{[t,l+1]}\sin \varphi_i^{[t,l+1]}\mathbf{Z}_{2i}\mathbf{u}_2^*)]$

$$c = -4 \sum_{i=1}^{p} n_i + \sum_{i=1}^{p} (1/\psi_i^{2[t,l+1]}) E_c^{[t]} [(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \psi_i^{[t,l+1]} \cos \varphi_i^{[t,l+1]} \mathbf{Z}_{1i}\mathbf{u}_1^*$$

$$- \psi_i^{[t,l+1]} \sin \varphi_i^{[t,l+1]} \mathbf{Z}_{2i}\mathbf{u}_2^*)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \psi_i^{[t,l+1]} \cos \varphi_i^{[t,l+1]} \mathbf{Z}_{1i}\mathbf{u}_1^*$$

$$- \psi_i^{[t,l+1]} \sin \varphi_i^{[t,l+1]} \mathbf{Z}_{2i}\mathbf{u}_2^*)]$$

$$d = \sum_{i=1}^{p} (1/\psi_i^{2[t,l+1]}) E_c^{[t]} [(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \psi_i^{[t,l+1]} \cos \varphi_i^{[t,l+1]} \mathbf{Z}_{1i}\mathbf{u}_1^*$$

$$- \psi_i^{[t,l+1]} \sin \varphi_i^{[t,l+1]} \mathbf{Z}_{2i}\mathbf{u}_2^*)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \psi_i^{[t,l+1]} \cos \varphi_i^{[t,l+1]} \mathbf{Z}_{1i}\mathbf{u}_1^*$$

$$- \psi_i^{[t,l+1]} \sin \varphi_i^{[t,l+1]} \mathbf{Z}_{2i}\mathbf{u}_2^*)]$$

The convergence of the EM-REML procedure is measured as the norm of the vector of changes in variance-covariance components between iterations. In our simulation and for the 3 parameterizations, convergence is assumed when the norm is less than 10^{-6}. In practice, the number of inner iterations is reduced to only one in the method of 'cyclic ascent'. The algebraic solution of quadratic, cubic or quartic equations, using the discriminant method, demonstrates that each time only one root is possible in the parameter space. In the simulated example, the polar parameterization converged the fastest.

Testing procedure

Let $L(\boldsymbol{\gamma}; \mathbf{y})$ be the log-restricted likelihood, Γ be the complete parameter space and Γ_0 a subset of it pertaining to the null hypothesis H_0. H_0 is rejected at the level α if the statistic $\zeta(y) = 2\text{Max}_\Gamma L(\boldsymbol{\gamma}; \mathbf{y}) - 2\text{Max}_{\Gamma_0} L(\boldsymbol{\gamma}; \mathbf{y})$ exceeds ζ_0 where ζ_0 corresponds to $\Pr[\chi_r^2 \geqslant \zeta_0] = \alpha$ (χ_r^2 is the chi-square distribution with r degrees of freedom given by difference between the number of parameters estimated under the full and the reduced models). Formulae to evaluate $-2\text{Max}L(\boldsymbol{\gamma}; \mathbf{y})$ can easily be made explicit:

$$-2L(\boldsymbol{\gamma}; \mathbf{y}) = \text{Const} + \ln|\mathbf{B}| + \sum_{i=1}^{p} \sigma_{e_i}^{-2} \mathbf{y}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{u_{1,i}} \mathbf{Z}_{1i}\mathbf{u}_1^* - \sigma_{u_{2,i}} \mathbf{Z}_{2i}\mathbf{u}_2^*)$$

where $\mathbf{B}$ is the coefficient matrix of the mixed-model equations.

NUMERICAL EXAMPLE

This procedure is illustrated from a hypothetical data set corresponding to a balanced, crossed design with 3 environments, 20 families per environment and 50 replicates per family ($p = 3$, $s = 20$ and $n = 50$). The 20 families were randomized within each environment. Basic ANOVA statistics for the between-family and within-family sums of squares and cross-products are given in table I. Table II presents the estimation of genetic and residual parameters under the full and reduced (hypothesis of a constant intra-class correlation between environments) models respectively, and the likelihood ratio test of the reduced model against the full model. The P values in table II indicate that there are no significant differences between intra-class correlations.

Table I. Sums of squares and cross-products between families (B) and within families (W) in 3 environments.

Cross product*	Sum
B-11	562 175.79
B-12	243 277.60
B-13	386 851.11
B-22	715 206.00
B-23	365 652.53
B-33	1 265 742.06
W-11	7 982 944.18
W-22	6 177 917.47
W-33	8 185 061.79

* $1, 2, 3$ = the 3 environments.

Sums of cross-products between families: $n \sum_{j=1}^{s}(y_{ij.} - y_{i..})(y_{i'j.} - y_{i'..})$

Sums of squares within families: $\sum_{j=1}^{s}\sum_{k=1}^{n}(y_{ijk} - y_{ij.})^2$

DISCUSSION AND CONCLUSION

In this paper, estimation and testing of homogeneity of intra-class correlations among environments have been studied with heteroskedastic univariate linear models. Another possible approach to account for 'genotype × environment' effects would be to consider the multiple-trait linear approach, defined by Falconer (1952). As described hereafter, these 2 approaches may or may not be equivalent. In this discussion, the conditions required to have equivalence between the multiple-trait and the univariate linear models will be established.

In Falconer's approach, expressions of the trait in different environments (i, i') are those of 2 genetically correlated traits, with a coefficient of correlation $\forall(i, i')$, $\rho_{ii'} = \sigma_{B_{ii'}} / \sigma_{B_i}\sigma_{B_{i'}}$. The model is defined as follows:

$$y_{ijk} = \mu_i + b_{ij} + e_{ijk} \qquad [4]$$

where y_{ijk} is the performance of the kth individual ($k = 1, 2, \ldots, n$) of the jth family ($j = 1, 2, \ldots, s$) evaluated in the ith environment ($i = 1, 2, \ldots, p$); b_{ij} is the random effect of the jth family in the ith environment, assumed normally distributed such that $\text{Var}(b_{ij}) = \sigma_{B_i}^2$, $\text{Cov}(b_{ij}, b_{i'j}) = \sigma_{B_{ii'}}$ for $i \neq i'$ and $\text{Cov}(b_{ij}, b_{i'j'}) = 0$ for $j \neq j'$ and any i and i'; e_{ijk} is a residual effect pertaining to the kth individual in the subclass ij, assumed normally and independently distributed with mean zero and variance $\sigma_{W_i}^2$.

Under the hypothesis of homogeneity of intra-class correlations between environments, the 2 approaches (multiple-trait and univariate) do not generate the same

Table II. Estimation of genetic and residual parameters between environments under the full and the reduced (constant intra-class correlation) models respectively and test of this last model against the full model.

Environment	Full model	Reduced model
Residual variances		
1	8 145.87	8 073.52
2	6 304.02	6 308.59
3	8 352.08	8 421.44
Family variances		
1	270.93	466.55
2	242.05	260.04
3	612.05	373.36
Interaction variances		
1	157.92	322.18
2	384.72	356.27
3	553.26	449.36
Genetic correlations		
12	0.49*	0.50
13	0.58*	0.52
23	0.45*	0.44
Intra-class correlations		
1	0.05*	
2	0.09*	0.09
3	0.12*	
−2Log-likelihood	3 043.73	3 047.23
Test		
ζ[a]		3.50
P-value[b]		0.17

[a] Likelihood ratio test; [b] degrees of freedom = 2; * same EM-REML estimates under the multiple trait approach.

number of parameters. Model [1] has $[2p + 1]$ genetic and residual parameters and model [4] has $[(p(p+1)/2) + 1]$ parameters.

For $p = 3$, whatever the hypotheses considered, even though these 2 models have the same number of estimable parameters, the parameter spaces are not exactly the same. Two conditions must be added to satisfy the equivalence between the multiple-trait and the univariate linear models. The univariate linear model does not allow the estimation of a negative genetic correlation between environments, since it is a ratio of variances. Thus, we have the following condition:

$$(1) \ \forall (i, j), \rho_{ij} \geqslant 0 \tag{5}$$

Furthermore, the relationships between the parameters of these 2 models are:

$$\begin{cases} \sigma_{B_i}^2 = \sigma_{s_i}^2 + \sigma_{hs_i}^2, \ \forall i \\ \sigma_{B_{ij}} = \sigma_{s_i} \sigma_{s_j}, \ \forall (i, j) \end{cases}$$

Then we have:
$$\sigma^2_{s_i} = \sigma_{B_{ij}}\sigma_{B_{ik}} \,/\, \sigma_{B_{jk}} = (\rho_{ij}\rho_{ik} \,/\, \rho_{jk})\sigma^2_{B_i}$$
and
$$\sigma^2_{hs_i} = \sigma^2_{B_i} - \sigma^2_{s_i} = (1 - (\rho_{ij}\rho_{ik} \,/\, \rho_{jk}))\sigma^2_{B_i}$$

By definition, $\sigma^2_{s_i}$ and $\sigma^2_{hs_i}$ are positive parameters, so the following relation must be satisfied:

$$(2) \; \forall\{i,j,k\} \in \{1,2,3\}, i \neq j, i \neq k \quad \text{and} \quad j \neq k, \rho_{jk} \geqslant \rho_{ij}\rho_{ik} \qquad [6]$$

It is worth noticing that the condition in [6] means that the partial genetic correlation between any pair (j,k) of environments for environments i fixed is also positive.

The problem of testing homogeneity of intra-class correlations between environments was finally solved under 3 different assumptions about the genetic correlations between environments: equal to one (Visscher, 1992); constant and positive (Robert *et al*, 1995); and just positive (this work).

For more than 3 traits, model [1] is no longer equivalent to the multiple trait approach of Falconer. As a matter of fact, it generates fewer parameters than [4], $2p$ *vs* $p(p+1)/2$ for [1] and [4] respectively.

This parsimony might be an interesting feature, because the difference in numbers of parameters increases with the number of traits considered (*eg*, 10 *vs* 15 parameters for 5 traits). Comparison of approaches on real genetic evaluation problems such as sire evaluation of dairy cattle in several countries would be of great interest.

REFERENCES

Falconer DS (1952) The problem of environment and selection. *Am Nat* 86, 293-298

Foulley JL, Henderson CR (1989) A simple model to deal with sire by treatment interactions when sires are related. *J Dairy Sci* 72, 167-172

Foulley JL, Quaas RL (1994) Statistical analysis of heterogeneous variances in Gaussian linear mixed models. *Proc 5th World Congress Genet Appl Livest Prod*, Univ Guelph, Guelph, ON, Canada, 18, 341-348

Foulley JL, Hébert D, Quaas RL (1994) Inference on homogeneity of between-family components of variance and covariance among environments in balanced cross-classified designs. *Genet Sel Evol* 26, 117-136

Lawley DN, Maxwell AE (1963) *Factor Analysis as a Statistical Method*. Butterworths Mathematical Texts, London, UK

Robert C, Foulley JL, Ducrocq V (1995) Genetic variation of traits measured in several environments. I. Estimation and testing of homogeneous and intra-class correlations between environments. *Genet Sel Evol* 27, 111-123

Shaw RG (1991) The comparison of quantitative genetic parameters between populations. *Evolution* 45, 143-151

Visscher PM (1992) On the power of likelihood ratio tests for detecting heterogeneity of intra-class correlations and variances in balanced half-sib designs. *J Dairy Sci* 73, 1320-1330

Zangwill (1969) *Non-Linear Programming: A Unified Approach*. Prentice-Hall, Englewood Cliffs, NJ, USA