

Inferences on homogeneity of between-family components of variance and covariance among environments in balanced cross-classified designs

JL Foulley¹, D Hébert², RL Quaas³

¹ Institut National de la Recherche Agronomique,
Station de Génétique Quantitative et Appliquée,
Centre de Recherches de Jouy-en-Josas, 78352 Jouy-en-Josas Cedex;

² Domaine Expérimental Agronomie d'Auzeville, Centre de Recherches de Toulouse,
BP 27, 31326 Castanet Tolosan Cedex, France;

³ Cornell University, Department of Animal Science, Ithaca, NY 14853, USA

(Received 4 August 1993; accepted 29 November 1993)

Summary – Estimation and testing of homogeneity of between-family components of variance and covariance among environments are investigated for balanced cross-classified designs. The variance-covariance structure of the residuals is assumed to be diagonal and heteroskedastic. The testing procedure for homogeneity of family components is based on the ratio of maximized log-restricted likelihoods for the reduced (hypothesis of homogeneity) and saturated models. An expectation-maximization (EM) algorithm is proposed for calculating restricted maximum likelihood (REML) estimates of the residual and between-family components of variance and covariance. The EM formulae to implement this are iterative and use the classical analysis of variance (ANOVA) statistics, *ie* the between- and within-family sums of squares and cross-products. They can be applied both to the saturated and reduced models and guarantee the solutions to be in the parameter space. Procedures presented in this paper are illustrated with the analysis of 5 vegetative and reproductive traits recorded in an experiment on 20 full-sib families of black medic (*Medicago lupulina* L) tested in 3 environments. Application to pure maximum likelihood procedures, extension to unbalanced designs and comparison with approaches relying on alternative models are also discussed.

genotype × environment interaction / heteroskedasticity / expectation-maximization / restricted maximum likelihood / likelihood ratio test

Résumé – Inférence relative à des composantes familiales homogènes de variance et de covariance entre milieux dans des dispositifs factoriels équilibrés. Cet article étudie les problèmes d'estimation et de test d'homogénéité des composantes familiales de variance et de covariance entre milieux dans des dispositifs factoriels équilibrés. La structure des variances et des covariances résiduelles est supposée diagonale et hétéroscédastique.

*La procédure de test d'homogénéité des composantes familiales repose sur le rapport des vraisemblances restreintes maximisées sous les modèles réduit (hypothèse d'homogénéité) et saturé. Un algorithme d'espérance-maximisation (EM) est proposé pour calculer les estimations du maximum de vraisemblance restreinte (REML) des composantes résiduelles et familiales de variance et de covariance. Les formules EM à appliquer sont itératives et utilisent les statistiques classiques de l'analyse de variance (ANOVA), c'est-à-dire les sommes de carrés et coproduits inter- et intrafamiliales. Elles s'appliquent à la fois aux modèles réduit et saturé et garantissent l'appartenance des solutions à l'espace des paramètres. Les méthodes présentées dans cet article sont illustrées par l'analyse de 5 caractères végétatifs et reproductifs mesurés lors d'une expérience portant sur 20 familles de pleins frères testées dans 3 milieux chez la minette (*Medicago lupulina* L). L'application au maximum de vraisemblance stricto sensu, la généralisation à des dispositifs déséquilibrés ainsi que la comparaison à des approches reposant sur d'autres modèles sont également discutées.*

interaction génotype \times milieu / hétéroscédasticité / espérance-maximisation / maximum de vraisemblance restreinte / rapport de vraisemblance

INTRODUCTION

There is a great deal of interest today in quantitative and applied genetics in heterogeneous variances. Ignoring such heterogeneity, as is usually done, may substantially affect the reliability of genetic evaluation and thus reduce the efficiency of selection (Hill, 1984; Visscher and Hill, 1992).

There is concern not only about estimating dispersion parameters for heteroskedastic models, but also about testing hypotheses for the real degree of heterogeneity which can be expected from experimental results. In this respect, Visscher (1992) investigated the statistical power of the likelihood ratio test in balanced half-sib designs for detecting heterogeneity of phenotypic variance and intra-class correlation between environments.

In that approach, the (family) correlation between environments (ρ) is assumed to be equal to 1, and heterogeneity of between-family components of covariance among environments is only due to scaling of variances.

The aim of this paper is to extend that approach to the case of true genotype by environment interactions ($\rho \neq 1$). Our attention will be focused on: i) cross-classified balanced designs; and ii) the null hypothesis involving homogeneity of between-family components of variance and covariance between environments. This variance-covariance structure has been widely used for analyzing family data recorded in different environments, in particular due to its close link with a 2-factor classification model (*ie* family and environment) with interaction (Mallard *et al*, 1983; Foulley and Henderson, 1989). Moreover, even for balanced designs, the estimation of the 2 parameters involved in this simple structure *via* maximum likelihood procedures has no analytical solution in the general case when no assumption is made about the residual variances. This motivated the proposal made in this study to use the expectation-maximization (EM) algorithm (Dempster *et al*, 1977) to solve the problem.

THEORY

Generalities

Let us assume that the records from the balanced cross-classified layout family (or genotype) \times environment can be written as:

$$y_{ijk} = \mu_i + b_{ij} + e_{ijk} \quad [1]$$

where y_{ijk} is the performance of the k th progeny (or individual) ($k = 1, 2, \dots, n$) of the j th family (or genotype) ($j = 1, 2, \dots, s$) evaluated in the i th environment ($i = 1, 2, \dots, p$); b_{ij} is the random effect of the j th family in the i th 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' ; and e_{ijk} is a residual effect pertaining to the k th progeny in the subclass ij , assumed $NIID(0, \sigma_{W_i}^2)$ *viz.*, normally and independently distributed with mean zero and variance $\sigma_{W_i}^2$.

Using vector notation, *ie* $\mathbf{y}_{jk} = \{y_{ijk}\}$, $\boldsymbol{\mu} = \{\mu_i\}$, $\mathbf{b}_j = \{b_{ij}\}$ and $\mathbf{e}_{jk} = \{e_{ijk}\}$ for $i = 1, 2, \dots, p$, the model [1] can alternatively be written as:

$$\mathbf{y}_{jk} = \boldsymbol{\mu} + \mathbf{b}_j + \mathbf{e}_{jk} \quad [2]$$

where $\mathbf{b}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma}_B)$ and $\mathbf{e}_{jk} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_W)$, with $\boldsymbol{\Sigma}_B = \{\sigma_{B_{ii'}}\}$, standing for the $(p \times p)$ matrix of between-family components of variance and covariance between environments and $\boldsymbol{\Sigma}_W = \text{Diag}\{\sigma_{W_i}^2\}$ for the $(p \times p)$ diagonal matrix of residual components of variance.

Actually, this approach consists of considering the expression of the trait in different environments (i, i') as that of 2 genetically related traits with a coefficient of correlation $\rho_{ii'} = \sigma_{B_{ii'}} / \sigma_{B_i} \sigma_{B_{i'}}$, (Falconer, 1952).

In a given environment (i), this 1-way linear model generates the classical ANOVA statistics, *ie* the between-family (S_{B_i}, B_i) and within-family (S_{W_i}, W_i) sums of squares and mean squares, respectively, whose distributions are proportional to chi-squares:

$$S_{B_i} = (s-1)B_i = n \sum_{j=1}^s (y_{ij.} - y_{i..})^2 \sim (\sigma_{W_i}^2 + n\sigma_{B_i}^2) \chi_{(s-1)}^2 \quad [3a]$$

$$S_{W_i} = s(n-1)W_i = \sum_{j=1}^s \sum_{k=1}^n (y_{ijk} - y_{ij.})^2 \sim \sigma_{W_i}^2 \chi_{s(n-1)}^2 \quad [3b]$$

Due to the cross-classified structure of the design, one also has to consider a sum ($S_{B_{ii'}}$) and a mean ($B_{ii'}$) between-family cross-product for each (ii') combination of environments:

$$S_{B_{ii'}} = (s-1)B_{ii'} = n \sum_{j=1}^s (y_{ij.} - y_{i..})(y_{i'j.} - y_{i'..}) \quad [4]$$

If we let $\mathbf{y}_j = \{y_{ij}\}$, $\mathbf{y}_{..} = \{y_{i..}\}$, then the matrix $\mathbf{S}_B = \{S_{B_{ii'}}\}$ with elements from [3a] and [4], such that:

$$\mathbf{S}_B = (s-1)\mathbf{B} = n \sum_{j=1}^s (\mathbf{y}_j - \mathbf{y}_{..})(\mathbf{y}_j - \mathbf{y}_{..})' \quad [5]$$

has a Wishart distribution, denoted $\mathbf{W}(\mathbf{\Gamma}, s-1)$, with parameters $(s-1)$ and $\mathbf{\Gamma} = \mathbf{\Sigma}_W + n\mathbf{\Sigma}_B$, thus generalizing to a matrix of between-family sums of squares and cross-products the $(\sigma_{W_i}^2 + n\sigma_{B_i}^2)\chi_{(s-1)}^2$ distribution arising in [3a].

In the 1-dimensional case, the set of S_{B_i} and S_{W_i} are independent, location invariant sufficient statistics for $\sigma_{B_i}^2$ and $\sigma_{W_i}^2$; similarly the matrices \mathbf{S}_B and $\mathbf{S}_W = \text{Diag}\{S_{W_i}\}$ have the same property for $\mathbf{\Sigma}_B$ and $\mathbf{\Sigma}_W$. Hence, one can write the density of [5] as

$$f(\mathbf{S}_B) \propto |\mathbf{\Gamma}|^{-(s-1)/2} \exp \left[-\frac{1}{2} \text{tr}(\mathbf{S}_B \mathbf{\Gamma}^{-1}) \right] \quad [6a]$$

Similarly,

$$f(\mathbf{S}_W) \propto |\mathbf{\Sigma}_W|^{-s(n-1)/2} \exp \left[-\frac{1}{2} \text{tr}(\mathbf{S}_W \mathbf{\Sigma}_W^{-1}) \right] \quad [6b]$$

Using [6a] and [6b] in the expression for the log likelihood,

$$L(\mathbf{\Sigma}_B, \mathbf{\Sigma}_W; \mathbf{y}) = L(\mathbf{\Sigma}_B, \mathbf{\Sigma}_W; \mathbf{S}_B) + L(\mathbf{\Sigma}_W; \mathbf{S}_W) + Ct$$

where Ct is a constant. This leads to:

$$\begin{aligned} L(\mathbf{\Sigma}_B, \mathbf{\Sigma}_W; \mathbf{y}) = & -\frac{1}{2}(s-1) [\ln|\mathbf{\Gamma}| + \text{tr}(\mathbf{B}\mathbf{\Gamma}^{-1})] \\ & - \frac{1}{2}s(n-1) [\ln|\mathbf{\Sigma}_W| + \text{tr}(\mathbf{W}\mathbf{\Sigma}_W^{-1})] + Ct \quad [7] \end{aligned}$$

with $\mathbf{W} = \text{Diag}\{W_i\} = \mathbf{S}_W/s(n-1)$ and $\text{tr}(\cdot) =$ the trace operator.

Notice that maximization of [7] yields REML estimators of $\mathbf{\Sigma}_B$ and $\mathbf{\Sigma}_W$ because the marginally sufficient statistics \mathbf{S}_B and \mathbf{S}_W are used in the log-likelihood function. Under the saturated model ($\sigma_{W_i}^2 \neq \sigma_{W_{i'}}^2$ and $\sigma_{B_{ii'}} \neq \sigma_{B_{i''i'''}}$ for any i, i', i'' and i'''), the partial derivatives with respect to $\sigma_{B_{ii'}}$ and $\sigma_{W_i}^2$ of minus twice the log likelihood ($-2L$) are:

$$\frac{\partial(-2L)}{\partial\sigma_{B_{ii'}}} = (s-1) \text{tr} \left[\frac{\partial\mathbf{\Gamma}}{\partial\sigma_{B_{ii'}}} (\mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}) \right] \quad [8a]$$

and

$$\begin{aligned} \frac{\partial(-2L)}{\partial\sigma_{W_i}^2} = & (s-1) \text{tr} \left[\frac{\partial\mathbf{\Gamma}}{\partial\sigma_{W_i}^2} (\mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}) \right] \\ & + s(n-1) \text{tr} \left[\frac{\partial\mathbf{\Gamma}}{\partial\sigma_{W_i}^2} (\mathbf{\Sigma}_W^{-1} - \mathbf{\Sigma}_W^{-1}\mathbf{W}\mathbf{\Sigma}_W^{-1}) \right] \quad [8b] \end{aligned}$$

$\partial\mathbf{\Gamma}/\partial\sigma_{B_{ii'}}$ is a $(p \times p)$ matrix having n as the (i, i') element and 0 elsewhere, so that the equation [8a] = 0 gives $\widehat{\mathbf{\Gamma}} = \mathbf{B}$. Similarly, $\partial\mathbf{\Sigma}_W/\partial\sigma_{W_i}^2$ has 1 as the i th diagonal element and 0 elsewhere. Given that $\mathbf{\Sigma}_W$ and \mathbf{W} are diagonal matrices and that $\widehat{\mathbf{\Gamma}} = \mathbf{B}$, the solutions to equations [8a] = 0 and [8b] = 0 are

$$\widehat{\mathbf{\Sigma}}_W = \mathbf{W} \quad [9a]$$

$$\widehat{\mathbf{\Sigma}}_B = (\mathbf{B} - \mathbf{W})/n \quad [9b]$$

provided that $\mathbf{B} - \mathbf{W}$ is positive definite. The maximum of the log-likelihood function is then (apart from a constant):

$$L_{\max} = (-1/2)\{(s-1)[\ln|\mathbf{B}| + p] + s(n-1)[\ln|\mathbf{W}| + p]\} \quad [10]$$

Otherwise, REML estimates of $\mathbf{\Sigma}_B$ and $\mathbf{\Sigma}_W$ are no longer identical to ANOVA estimates and require the use of another algorithm for their calculation (see *Appendix A*).

The null hypothesis consists of assuming the homogeneity of the between-family components of variance ($\forall i, \sigma_{B_i}^2 = \sigma_B^2$) and covariance ($\forall i \neq i', \sigma_{B_{ii'}} = C_B$) as postulated in many analyses of genotype by environment experiments (Dickerson, 1962; Yamada, 1962; Mallard *et al.*, 1983). The approach presented in this paper allows us to test this simplified structure of $\mathbf{\Sigma}_B$ against Falconer's saturated model for any structure of the residual variances. The nulle hypothesis (H_0) considered here can be written as:

$$\begin{aligned} \mathbf{\Sigma}_B &= (\sigma_B^2 - C_B)\mathbf{I}_p + C_B\mathbf{J}_p \\ \mathbf{\Sigma}_W &= \text{Diag}\{\sigma_{W_i}^2\} \end{aligned} \quad [11]$$

where \mathbf{I}_p = identity matrix of order p and $\mathbf{J}_p = (p \times p)$ matrix of ones.

Under H_0 , REML estimation of $\mathbf{\Sigma}_B$ and $\mathbf{\Sigma}_W$ becomes much more complex. Here $\partial\mathbf{\Gamma}/\partial\sigma_B^2 = n\mathbf{I}_p$ and $\partial\mathbf{\Gamma}/\partial C_B = n(\mathbf{J}_p - \mathbf{I}_p)$ result in the following equations:

$$\text{tr}(\mathbf{\Gamma}^{-1}) = \text{tr}(\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}) \quad [12a]$$

$$\mathbf{1}'_p\mathbf{\Gamma}^{-1}\mathbf{1}_p = \mathbf{1}'_p\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}\mathbf{1}_p \quad [12b]$$

were $\mathbf{1}_p = (p \times 1)$ vector of ones.

Since $\mathbf{\Gamma}^{-1} \neq \mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}$, the REML solution for the residual components ([8b]) is no longer $\widehat{\mathbf{\Sigma}}_W = \mathbf{W}$ and the system of equations [8b] (see also [B11]), [12a] and [12b] has no analytical solutions in the general case. This was the reason motivating our search for another approach for computing REML solutions to $\mathbf{\Sigma}_B$ and $\mathbf{\Sigma}_W$ under H_0 .

An EM diagonalization approach

The expectation-maximization approach is a very efficient concept in maximum likelihood estimation (Dempster *et al.*, 1977). It has been widely used for calculating ML and REML estimates of variance components of linear models (Meyer, 1990;

Quaas, 1992). The basic principle is to treat the unobservable random variables b_{ij} and e_{ijk} as missing data. Actually, the EM algorithm will not be applied directly to the model described in [1] and [2] but after a spectral decomposition of Σ_B according to its eigenvalues and vectors, *ie*:

$$\Sigma_B = \mathbf{U}\mathbf{\Delta}\mathbf{U}' \quad [13]$$

In this formula, $\mathbf{\Delta} = \text{Diag}\{\delta_i\}$ is the $(p \times p)$ matrix of eigenvalues δ_i , with δ_i repeated as many times as its multiplicity order, and $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_i, \dots, \mathbf{U}_p)$ is the $(p \times p)$ matrix of the corresponding p normed eigenvectors \mathbf{U}_i of Σ_B ($\mathbf{U}'\mathbf{U} = \mathbf{I}_p$). Under the special form shown in [11], Σ_B has only 2 distinct eigenvalues:

$$\delta_1 = \sigma_B^2 + (p-1)C_B \quad [14a]$$

and

$$\delta_2 = \sigma_B^2 - C_B \quad [14b]$$

with multiplicity orders 1 and $(p-1)$ respectively. Moreover, the matrix \mathbf{U} of eigenvectors does not depend on the values of δ_1 and δ_2 , \mathbf{U}' being the Helmert matrix of order p , see for example Searle, 1982 (p 71 and 322) for more details about such matrices. For instance, for $p = 3$,

$$\mathbf{U}' = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix}$$

Due to the invariance property of (RE)ML estimators, the one-to-one transformation in [14a] and [14b] allows us to change the parameterization from (σ_B^2, C_B) to (δ_1, δ_2) , or more conveniently to (δ_1, τ) where $\tau = \delta_1 + (p-1)\delta_2$, the back transformation is:

$$\begin{bmatrix} \sigma_B^2 \\ C_B \end{bmatrix} = \begin{bmatrix} 0 & 1/p \\ 1/(p-1) & -1/p(p-1) \end{bmatrix} \begin{bmatrix} \delta_1 \\ \tau \end{bmatrix} \quad [15]$$

From the spectral decomposition of Σ_B , the model in [2] can be written as:

$$\mathbf{y}_{jk} = \boldsymbol{\mu} + \mathbf{U}\mathbf{f}_j + \mathbf{e}_{jk} \quad [16]$$

where \mathbf{U} is defined as before and the vector $\mathbf{f}_j = \{f_{ij}\}$ is such that $\mathbf{f}_j \sim \mathbf{N}(\mathbf{0}, \mathbf{\Delta})$.

Using the Dempster *et al* (1977) terminology, a complete data set \mathbf{x} can be constructed from $\boldsymbol{\mu}, \mathbf{f}_j, \mathbf{e}_{jk}$ for $j = 1, 2, \dots, s$ and $k = 1, 2, \dots, n$, whereas the incomplete data set is the vector \mathbf{y} of observations.

Let us first consider the case of Σ_B . If the f_{ij} 's were known, sufficient statistics for δ_1 and τ would be, under the normality assumption:

$$\sum_{j=1}^s f_{1j}^2 \quad \text{and} \quad \sum_{i=1}^p \sum_{j=1}^s f_{ij}^2 = \text{tr} \left(\sum_{j=1}^s \mathbf{f}_j \mathbf{f}_j' \right), \quad \text{respectively.}$$

REML would then be obtained by equating the expectation of these sufficient statistics, *ie*:

$$\begin{aligned} E \left(\sum_{j=1}^s f_{1j}^2 \right) &= s\delta_1 \\ E \left[\text{tr} \left(\sum_{j=1}^s \mathbf{f}_j \mathbf{f}_j' \right) \right] &= s\tau \end{aligned}$$

to their calculated values (M step). Actually, these sufficient statistics are not directly observable and the EM algorithm proceeds first by estimating them by taking their conditional expectation given the observed data set (E step). Since such an estimation depends on the value of the unknown parameters, the procedure is iterative and consists of implementing the 2 usual steps:

E step: at iteration $[t]$, calculate

$$\mathbf{\Lambda}^{[t]} = E \left(\sum_{j=1}^s \mathbf{f}_j \mathbf{f}_j' | \mathbf{y}, \delta_1^{[t]}, \tau^{[t]}, \mathbf{\Sigma}_W^{[t]} \right) \quad [17]$$

M step: compute $\delta_1^{[t+1]}$ and $\tau^{[t+1]}$ from the following equations:

$$\delta_1^{[t+1]} = \left(\mathbf{\Lambda}^{[t]} \right)_{11} / s \quad [18a]$$

$$\tau^{[t+1]} = \text{tr} \left(\mathbf{\Lambda}^{[t]} \right) / s \quad [18b]$$

As shown in *Appendix A*, the $(p \times p)$ matrix $\mathbf{\Lambda}^{[t]}$ can be expressed as:

$$\mathbf{\Lambda}^{[t]} = s\mathbf{C}^{[t]} + n\mathbf{C}^{[t]}\mathbf{U}' \left(\mathbf{\Sigma}_W^{[t]} \right)^{-1} \left[(s-1)\mathbf{B} + \mathbf{\Gamma}^{[t]} \right] \left(\mathbf{\Sigma}_W^{[t]} \right)^{-1} \mathbf{U}\mathbf{C}^{[t]} \quad [19]$$

where \mathbf{U} , \mathbf{B} and $\mathbf{\Gamma}$ are defined as above (see [13], [5a] and [5b] respectively) and $\mathbf{C}^{[t]}$ is the matrix of variance of prediction errors of $\widehat{\mathbf{f}}_j^{[t]} = E(\mathbf{f}_j | \mathbf{y}, \delta_1^{[t]}, \tau^{[t]}, \mathbf{\Sigma}_W^{[t]})$, the best predictor of \mathbf{f}_j at iteration $[t]$ such that:

$$\mathbf{C}^{[t]} = \left[n\mathbf{U}' \left(\mathbf{\Sigma}_W^{[t]} \right)^{-1} \mathbf{U} + \left(\mathbf{\Delta}^{[t]} \right)^{-1} \right]^{-1} \quad [20]$$

Similarly, sufficient statistics for $\mathbf{\Sigma}_W$ under the complete data set \mathbf{x} are:

$$\text{Diag} \left\{ \sum_{j=1}^s \sum_{k=1}^n e_{ijk}^2 \right\} \text{ for } i = 1, 2, \dots, p$$

and the E and M steps are as follows:

For the E step, at iteration $[t]$, calculate:

$$\mathbf{\Omega}^{[t]} = \text{E} \left(\sum_{j=1}^s \sum_{k=1}^n \mathbf{e}_{jk} \mathbf{e}'_{jk} | \mathbf{y}, \delta_1^{[t]}; \tau^{[t]}, \mathbf{\Sigma}_W^{[t]} \right) \quad [21]$$

using the following formula based on the same reasoning as previously (see *Appendix A*):

$$\begin{aligned} \mathbf{\Omega}^{[t]} = & (ns - 1)\mathbf{T} + \mathbf{\Gamma}^{[t]} + \mathbf{M}^{[t]} \left[(s - 1)\mathbf{B} + \mathbf{\Gamma}^{[t]} \right] \mathbf{M}^{[t]'} \\ & - \left[(s - 1)\mathbf{B} + \mathbf{\Gamma}^{[t]} \right] \mathbf{M}^{[t]'} - \mathbf{M}^{[t]} \left[(s - 1)\mathbf{B} + \mathbf{\Gamma}^{[t]} \right] + ns\mathbf{UC}^{[t]}\mathbf{U}' \end{aligned} \quad [22]$$

where $(ns - 1)\mathbf{T} = \sum_{j=1}^s \sum_{k=1}^n (\mathbf{y}_{jk} - \mathbf{y}_{..})(\mathbf{y}_{jk} - \mathbf{y}_{..})'$ is the $(p \times p)$ matrix of total sums of squares and cross-products ($\mathbf{y}_{..}$ being defined as $(\sum_{j=1}^s \sum_{k=1}^n \mathbf{y}_{jk})/ns$), and $\mathbf{M}^{[t]} = n\mathbf{UC}^{[t]}\mathbf{U}' (\mathbf{\Sigma}_W^{[t]})^{-1}$.

For the M step compute the next value of $\mathbf{\Sigma}_W$ from:

$$\mathbf{\Sigma}_W^{[t+1]} = \text{Diag}\{\mathbf{\Omega}_{ii}^{[t]}/ns\} \quad [23]$$

for $i = 1, 2, \dots, p$.

Let us define $\mathbf{P} = \sum_{j=1}^s \mathbf{b}_j \mathbf{b}'_j$ and $\mathbf{\Pi}^{[t]} = \text{E} \left(\sum_{j=1}^s \mathbf{b}_j \mathbf{b}'_j | \mathbf{y}, \mathbf{\Sigma}_W^{[t]}, \mathbf{\Sigma}_B^{[t]} \right)$. From [17], one has $\mathbf{\Pi}^{[t]} = \mathbf{U}\mathbf{\Lambda}^{[t]}\mathbf{U}'$, and conversely $\mathbf{\Lambda}^{[t]} = \mathbf{U}'\mathbf{\Pi}^{[t]}\mathbf{U}$. Equations [18a] and [18b] can be written as:

$$s\delta_1^{[t+1]} = s\delta_B^{2[t+1]} + s(p - 1)C_B^{[t+1]} = (\mathbf{1}'\mathbf{\Pi}^{[t]}\mathbf{1})/p$$

$$s\tau^{[t+1]} = ps\sigma_B^{2[t+1]} = \text{tr}(\mathbf{\Pi}^{[t]})$$

ie,

$$\sigma_B^{2[t+1]} = \text{tr}(\mathbf{\Pi}^{[t]})/ps \quad [24a]$$

$$C_B^{[t+1]} = \left[(\mathbf{1}'\mathbf{\Pi}^{[t]}\mathbf{1}) - \text{tr}(\mathbf{\Pi}^{[t]}) \right] / p(p - 1)s \quad [24b]$$

Since $\mathbf{UCU}' = \mathbf{\Sigma}_B\mathbf{\Gamma}^{-1}\mathbf{\Sigma}_W$, an alternative expression for $\mathbf{\Pi}^{[t]} = \mathbf{U}\mathbf{\Lambda}^{[t]}\mathbf{U}'$ is

$$\mathbf{\Pi}^{[t]} = s\mathbf{\Sigma}_B^{[t]} + n(s - 1)\mathbf{\Sigma}_B^{[t]} (\mathbf{\Gamma}^{[t]})^{-1} (\mathbf{B} - \mathbf{\Gamma}^{[t]}) (\mathbf{\Gamma}^{[t]})^{-1} \mathbf{\Sigma}_B^{[t]} \quad [25]$$

Formulae [25] and [24a]–[24b] define the E and M steps, respectively, of an EM procedure equivalent to that described previously but applied to untransformed

parameters. Notice that in this scheme $\text{tr}(\mathbf{P})/p$ (average diagonal element of \mathbf{P}) and $[(\mathbf{1}'\mathbf{P}\mathbf{1}) - \text{tr}(\mathbf{P})]/p(p-1)$ (average off-diagonal element of \mathbf{P}) behave as sufficient statistics for σ_B^2 and C_B with respect to the complete data set.

Formulae for the residual components are unchanged with $\mathbf{UC}^{[t]}\mathbf{U}' = \Sigma_B^{[t]}$ $(\Gamma^{[t]})^{-1} \Sigma_W^{[t]}$ and $\mathbf{M}^{[t]} = n\Sigma_B^{[t]} (\Gamma^{[t]})^{-1}$. For the saturated model, the formulae to apply are the same for Σ_W and, simply, $\Sigma_B^{[t+1]} = \mathbf{\Pi}^{[t]}/s$ for Σ_B .

Testing procedures

Hypotheses of interest concern the vector θ of parameters involved in the matrices of between-family (Σ_B) and within-family (Σ_W) components of variance and covariance between environments. The theory of the generalized likelihood ratio can be applied to that purpose, as already proposed by Foulley *et al* (1990, 1992), Shaw (1991) and Visscher (1992) among others.

Let $H_0 : \theta \in \Theta_0$ be the null hypothesis and $H_1 : \theta \in \Theta - \Theta_0$ its alternative, where Θ refers to the complete parameter space and Θ_0 , a subset of it pertaining to H_0 . Under H_0 , the statistic

$$\lambda(\mathbf{y}) = 2\text{Max}_{\Theta} L(\theta; \mathbf{y}) - 2\text{Max}_{\Theta_0} L(\theta; \mathbf{y}) \quad [26]$$

where $L(\theta; \mathbf{y})$ is defined as in [7], has an asymptotic chi-square distribution with r degrees of freedom, r being the difference in the numbers of estimable parameters involved in Θ and Θ_0 (Mood *et al*, 1974).

Here Θ contains $p(p+3)/2$ parameters corresponding to p residual components of variance and $p(p+2)/2$ between-family components of variance and covariance between environments whilst Θ_0 has $p+2$ parameters only (p residual components, σ_B^2 and C_B), so that $r = [p(p+1)/2] - 2$.

In the Neyman-Pearson approach of hypothesis testing, H_0 is rejected at the α level if the calculated value of $\lambda(\mathbf{y})$ exceeds a critical value λ_c such that $\Pr(\chi_r^2 \geq \lambda_c) = \alpha$. However, the likelihood ratio statistic $\lambda(\mathbf{y})$ in [24] can also be interpreted as the difference in degree of fit *via* maximum likelihood procedures by 2 models: a reduced model (R) with parameter vector $\theta \in \Theta_0$ and a full model (F) with $\theta \in \Theta$ encompassing both the null hypothesis and its alternative. In the theory of significance testing (Kempthorne and Folks, 1971), this statistic is also used as a measure of strength of evidence against the reduced model or the null hypothesis. The lower the probability under H_0 of exceeding this statistic evaluated from the data (also referred to as the P -value or significance level or size of the test), the stronger the evidence against H_0 .

Example

Data used here to illustrate the procedures are from an experiment carried out in Montpellier (south west of France) on 20 full-sib families of black medic (*Medicago lupulina* L.) tested in 3 different environments (control, harvesting and competition treatments).

The experimental design was described in detail by Hébert (1991). There were 2 replicates per environment and the 20 genotypes were randomly allocated to each

replicate. Thirty-six traits were recorded and the variable used was the mean of the 5 plants cultivated in each replicate so that $p = 3$, $s = 20$ and $n = 2$.

Basic ANOVA statistics for the between-family and within-family sums of squares and cross-products are given in table I for a subset of 5 traits.

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

	1 <i>Days to flowering (d)</i>	2 <i>Days to 1st ripe pod (d)</i>	3 <i>Dry matter weight (g)</i>	4 <i>Dry matter weight/max plant size (dgc^m-¹)</i>	5 <i>Pod weight/ total weight (%)</i>
B-11	2 261.50	1 882.08	15 719.48	91.60	4 055
B-12	2 648.14	1 271.12	21 703.85	120.10	3 060
B-13	2 598.80	1 323.58	7 775.58	42.10	3 259
B-22	4 402.50	1 823.80	49 838.22	256.40	3 390
B-23	3 860.76	1 330.16	18 403.41	93.50	2 761
B-33	4 058.80	1 501.10	8 132.36	45.40	2 891
W-11	279.22	233.84	3 204.06	17.30	679
W-22	972.28	431.90	14 014.01	77.90	220
W-33	331.76	160.32	1 037.41	5.50	545

^a 1, 2, 3 = harvesting, control, competition environments, respectively.

Firstly, the null assumption that the diagonal terms of Σ_W were equal was tested *via* a Bartlett's test based on ANOVA mean squares statistics. *P* values were 0.007, 0.08, 1.4×10^{-7} , 8×10^{-8} and 0.04 so that this assumption can be reasonably rejected (except perhaps for trait 2).

Test statistics about Σ_B and estimates of Σ_B and Σ_W under both the reduced and saturated models are given in table II. *P*-values for vegetative yield traits, represented here by dry matter weight (trait No 3) and dry matter weight/max plant size diameter (trait No 4), were very low, indicating a large heterogeneity in genetic variation between environments with full-sib variances substantially reduced in the harvesting ($i = 1$) and competition ($i = 3$) environments compared with the control ($i = 2$). In contrast, the harvesting and competition environments do not generate a meaningful level of stress compared with the control for the expression of genetic variation of days to 1st ripe pod (trait No 2) and relative pod weight (trait No 5). These 3 environments then behave as 'exchangeable', as statisticians would say. In this example, genetic correlations between environments were rather high and it would have been interesting to test for some traits (*eg* No 1 and 5) using the assumption that these correlations are equal to unity by Visscher's (1992) procedures.

Table II. Estimates of variance-covariance components among environments^a under the saturated and reduced models and test of fit for this last model.

Model	1 Days to flowering (d)	2 Days to 1st ripe pod (d)	3 Dry matter weight (g)	4 Dry matter weight/ max plant size (dgc m ⁻¹)	5 Pod weight/ total weight (%)	
Saturated	Residual	11	156.04	0.83	27.28	
	components	22	512.06	2.98	10.89	
		33	46.35	0.26	19.57	
		11	337.86	2.01	96.59	
		22	1155.03	5.74	83.82	
	Between-family	35.49	193.76	1.07	70.36	
	components	12	556.35	3.07	80.98	
		13	207.16	1.12	81.98	
		23	467.24	2.40	72.17	
	Minus 2L	540.52	487.78	774.76	170.02	534.31
Reduced	Residual	11	182.46	0.95	32.55	
	components	22	856.07	4.44	13.97	
		33	49.70	0.27	21.07	
	Between-family	38.31	271.37	1.62	79.13	
	components	12	240.67	1.36	77.67	
		ρ^b	(0.89)	(0.89)	(0.84)	(0.98)
	Minus 2L	550.20	489.58	796.94	189.19	540.14
	Test	λ^c	9.69	22.19	19.17	5.83
		P-value	0.046	2×10^{-4}	7×10^{-4}	0.21

^a1, 2, 3 = harvesting, control, competition environments, respectively; $\rho = (12)/(11) = C_B/\sigma_B^2$; ^c likelihood ratio statistic.

DISCUSSION

This paper describes a further contribution to the solution of the problem of testing homogeneity of between-family components of variance and covariance between environments in the case of balanced cross-classified designs. The testing procedure is based on the likelihood ratio test as already advocated by Shaw (1987) in quantitative genetics. This study extends that of Visscher (1992), which was restricted to the case of pure scaling effects between environments (*ie* all genetic correlations between environments equal to one).

The choice of an EM algorithm for computing REML estimates of Σ_B and Σ_W under the null hypothesis allows us to make explicit the equations of the iterative process to implement *via* formulae based on the usual ANOVA statistics. This algorithm does not require any constraint on the value of these ANOVA statistics (*eg* $\mathbf{B} - \mathbf{W}$ can be non-positive definite) provided the starting values for Σ_B are within the parameter space. A simple reason for this is that the E phase under the restricted model involves the conditional expectation (given the data) of sums of squares, *eg* $\sum_j f_{ij}^2$ and $\sum_j \sum_k e_{ijk}^2$ as estimators of variance. Because $E\left(\sum_j \mathbf{f}_j \mathbf{f}'_j | \mathbf{y}, \boldsymbol{\theta} = \boldsymbol{\theta}^{[t]}\right)$ is always positive definite (Foulley *et al*, 1987), this property of the EM algorithm is also true under the saturated model; it can then be used to provide REML estimates of Σ_B and Σ_W when ANOVA estimators are not permissible (*eg* for traits 1, 3, 4 and 5 in the example).

Some authors such as Anderson (1984) and Shaw (1987) advocate the use of ML rather than REML procedures to test hypotheses about variance covariance matrices. Our EM algorithm can be easily adapted to obtain such ML estimates of Σ_B and Σ_W . It suffices to replace $(s - 1)\mathbf{B} + \mathbf{\Gamma}$ in formulae [19] and [21] for $\mathbf{\Lambda}$ and $\mathbf{\Omega}$ respectively by $(s - 1)\mathbf{B}$ corresponding to the change in the conditional expectation (given the data) of $n \sum_j (\mathbf{y}_j - \boldsymbol{\mu})(\mathbf{y}_j - \boldsymbol{\mu})'$ according to whether $\boldsymbol{\mu}$ is considered as a parameter of interest (ML) or a nuisance factor to be integrated out (REML).

This algorithm can also retrieve the usual ANOVA estimates for Σ_B and Σ_W using the 2-way crossed-mixed model:

$$y_{ijk} = \mu + h_i + s_j + hs_{ij} + e_{ijk} \tag{27}$$

involving fixed environmental effects (h_i), random family effects [$s_j \sim \text{NIID}(0, \sigma_s^2)$], random interactions [$hs_{ij} \sim \text{NIID}(0, \sigma_{hs}^2)$] and residuals [$e_{ijk} \sim \text{NIID}(0, \sigma_e^2)$]. In fact, Foulley and Henderson (1989) showed that this is an equivalent model for a simplified version of the multiple trait model in [1] restricted to $\Sigma_B = (\sigma_B^2 - C_B)\mathbf{I}_p + C_B\mathbf{J}_p$ and $\Sigma_W = \sigma_W^2\mathbf{I}_p$ with $\sigma_B^2 = \sigma_s^2 + \sigma_{hs}^2$, $C_B = \sigma_s^2$ and $\sigma_W^2 = \sigma_e^2$. Notice however that this simplified multiple trait model differs from that considered throughout this study (see [11]) not only by the assumption of homoskedastic residual variances but also by its restriction to a positive covariance (C_B) between environments.

For instance, for trait 1 (days to flowering), EM-REML estimates of variance and covariance components obtained from the algorithm described in [17] to [22] with $\sigma_W^{2[t+1]} = \text{tr}(\boldsymbol{\Omega}^{[t]})/nsp$ in [23] are: $\hat{\sigma}_B^2 = 80.86$; $\hat{C}_B = 79.89$; and $\hat{\sigma}_W^2 = 26.39$. These

values can easily be checked with ANOVA estimates: $\hat{\sigma}_s^2 = 79.89$; $\hat{\sigma}_{h,s}^2 = 0.97$; and $\hat{\sigma}_e^2 = 26.39$. Here again, the EM algorithm provides estimates within the parameter space, which is not always the case with ANOVA estimators as shown for instance with trait 5 : $\hat{\sigma}_B^2 = 79.50$, $\hat{C}_B = 79.43$ and $\hat{\sigma}_W^2 = 23.23$ with EM *versus* $\hat{\sigma}_s^2 = 79.65$, $\hat{\sigma}_{h,s}^2 = -1.02$ and $\hat{\sigma}_e^2 = 24.07$ with ANOVA.

The EM algorithm is a first-order procedure and therefore has close relationships with a maximization procedure based on zeroed first derivatives. As shown in *Appendix B* in the case of Σ_B , the difference between the formulae to implement in the 2 iterative procedures consists of replacing $(s-1)\mathbf{B} + \Gamma^{[t]}$ in [19] by $s\mathbf{B}$ in the 1st derivative algorithm (\mathbf{B} being a sufficient statistic for Γ in the saturated model). Again, the use of EM guarantees staying in the parameter space whereas there is no obvious proof of that for the 1st derivative algorithm. Moreover, the EM approach turns out to be easier to understand and to use than the other one, which relies mainly on algebraic tricks. As far as REML estimates of Σ_W are concerned, a functional iteration algorithm based on first derivatives was also proposed in *Appendix B* due to the lack of an obvious analogue of the EM formulae.

Finally, this EM reasoning can be extended to an unbalanced structure of data and to additional nuisance fixed effects cross-classified with family effects, *ie* to an extended version of model [1] such as

$$y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + b_{ij} + e_{ijk} \quad [28]$$

where $\boldsymbol{\beta}$ is a vector of fixed effects including the i th classification for environment and effects of other factors to be adjusted for, and \mathbf{x}'_{ijk} is the corresponding incidence (row) vector. Under that model, the E and M steps defined in [17] and [18a] and [18b] for Σ_B and in [21] and [23] for Σ_W are still valid. However, the E-statistics in [17], [21] and [24b] are not evaluated as functions of ANOVA statistics but directly from the numerical values of the BLUP and the variance of prediction errors of the \mathbf{f}_j or \mathbf{b}_j 's. In this situation, also, the EM algorithm should be applied systematically to both the reduced and saturated model for Σ_W . Another approach would be to write [28] under its equivalent form (for $C_B > 0$):

$$y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + h_i + s_j + hs_{ij} + e_{ijk} \quad [29]$$

where h_i, s_j and hs_{ij} are defined as in [25] and $e_{ijk} \sim \text{NID}(0, \sigma_{W_i}^2)$. Under such a mixed-model structure, one can then use the methods developed by Foulley *et al* (1990, 1992) and San Cristobal *et al* (1993) for calculating REML estimates of variances in the presence of heterogeneous residual components. However, the procedure derived in this paper remains definitively more general, for instance, it can also be easily applied to a non-diagonal structure of Σ_W using formulae [17] to [22] unchanged and [23] slightly modified into $\Sigma_W^{[t+1]} = \Omega^{[t]}/ns$.

This paper deals with a null hypothesis of constant between-family variance and covariance. In some instances, a more appropriate null hypothesis would be a constant between-family correlation (ρ) between environments ($\sigma_{B_{i,i'}} = \rho\sigma_{B_i}\sigma_{B_{i'}}$) and/or of constant intraclass correlation [$\sigma_{B_i}^2 = t(\sigma_{B_i}^2 + \sigma_{W_i}^2)$]. Testing procedures for these assumptions will be reported in a separate article.

ACKNOWLEDGMENTS

The authors are grateful to I Olivieri (INRA, Montpellier) for stimulating discussions which motivated this study, to J Ruane for the English revision of the manuscript and to the anonymous referees for their valuable comments. Special thanks are expressed to C Robert and M San Cristobal who read the manuscript thoroughly and checked the numerical example.

REFERENCES

- Anderson TW (1984) *An Introduction to Multivariate Statistical Analysis*. J Wiley and Sons, New York
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Statist Soc B* 39, 1-38
- Dickerson GE (1962) Implications of genetic-environmental interaction in animal breeding. *Anim Prod* 4, 47-63
- Falconer DS (1952) The problem of environment and selection. *Amer 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, Im S, Gianola D, Hoeschele I (1987) Empirical Bayes estimation of parameters for n polygenic binary traits. *Genet Sel Evol* 19, 197-224
- Foulley JL, Gianola D, San Cristobal M, Im S (1990) A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models. *J Dairy Sci* 73, 1612-1624
- Foulley JL, San Cristobal M, Gianola D, Im S (1992) Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Comput Stat Data Anal* 13, 291-305
- Hébert D (1991) Plasticité phénotypique et interaction génotype milieu chez *Medicago Lupulina*. Thèse Doctorat Sciences. Univ Sciences Techniques du Languedoc, Montpellier
- Hill WG (1984) On selection among groups with heterogeneous variance. *Anim Prod* 39, 473-477
- Kempthorne O, Folks L (1971) *Probability, statistics and data analysis*. The Iowa State University Press, Ames (IO)
- Mallard J, Masson JP, Douaire M (1983) Interaction génotype \times milieu et modèle mixte. I. Modélisation. *Genet Sel Evol* 15, 379-394
- Meyer K (1990) Present status of knowledge about statistical procedures and algorithms to estimate variance and covariance components. In: *4th World Congress on Genetics Applied to Livestock Production*, Edinburgh, 23-27 July 1990, vol 13. (WG Hill, R Thompson, JA Wooliams, eds), 407-418
- Mood A, Graybill FA, Boes DC (1974) *Introduction to the Theory of Statistics*. Mc Graw-Hill Inc, London
- Quaas RL (1992) REML NoteBook. Mimeo, Dep Anim Sci, Cornell Univ Ithaca (NY)
- San Cristobal M, Foulley JL, Manfredi E (1993) Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genet Sel Evol* 25, 3-30

Searle SR (1982) *Matrix Algebra Useful to Statistics*. J Wiley and Sons, New York
 Shaw RG (1987) Maximum likelihood approaches applied to quantitative genetics of natural populations. *Evolution* 41, 812-826
 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* 75, 1320-1330
 Visscher PM, Hill WG (1992) Heterogeneity of variance and dairy cattle breeding. *Anim Prod* 55, 321-329
 Yamada Y (1962) Genotype by environment interaction and genetic correlation of the same trait under different environments. *Jap J Genet* 37, 498-509

APPENDIX

An EM algorithm for REML estimations of Σ_B and Σ_W (part A)

Σ_B and Σ_W under H_0

An explicit formula for Λ can be obtained by successively conditioning and deconditioning the expression in [17] with respect to the mean vector μ , ie:

$$\Lambda^{[t]} = E_{\mu} \left[E \left(\sum_{j=1}^s \mathbf{f}_j \mathbf{f}'_j | \mathbf{y}, \mu, \theta = \theta^{[t]} \right) \right] \tag{A1}$$

where θ stands for the vector of parameters involved in the matrices of between-family (Σ_B) and within-family (Σ_W) components of variance and covariance between environments and $\theta^{[t]}$ is the current estimate of θ at iteration $[t]$.

Now, conditionally on \mathbf{y}, μ and $\theta = \theta^{[t]}$, the expectation of $\sum_{j=1}^s \mathbf{f}_j \mathbf{f}'_j$ can be decomposed into:

$$\begin{aligned} E \left(\sum_{j=1}^s \mathbf{f}_j \mathbf{f}'_j | \mathbf{y}, \mu, \theta = \theta^{[t]} \right) &= \sum_{j=1}^s E(\mathbf{f}_j | \mathbf{y}, \mu, \theta = \theta^{[t]}) E(\mathbf{f}'_j | \mathbf{y}, \mu, \theta = \theta^{[t]}) \\ &+ \sum_{j=1}^s \text{Var}(\mathbf{f}_j | \mathbf{y}, \mu, \theta = \theta^{[t]}) \end{aligned} \tag{A2}$$

This decomposition is especially helpful because it allows us to introduce the usual statistics of Gaussian models, ie the conditional mean $\hat{\mathbf{f}}_j = E(\mathbf{f}_j | \mathbf{y}, \mu, \theta)$ and its prediction error variance $\mathbf{C}_j = \text{Var}(\mathbf{f}_j | \mathbf{y}, \mu, \theta)$. Here, we have:

$$\hat{\mathbf{f}}_j = (n\mathbf{U}'\Sigma_W^{-1}\mathbf{U} + \Delta^{-1})^{-1} \mathbf{U}'\Sigma_W^{-1}n(\mathbf{y}_j - \mu) \tag{A3a}$$

$$\mathbf{C}_j = \mathbf{C} = (n\mathbf{U}'\Sigma_W^{-1}\mathbf{U} + \Delta^{-1})^{-1} \tag{A3b}$$

The next step is to specify the expression for:

$$\mathbf{E}_\mu \left(\sum_{j=1}^s \widehat{\mathbf{f}}_j \widehat{\mathbf{f}}_j' \right) = n \mathbf{C} \mathbf{U}' \Sigma_W^{-1} \left\{ \mathbf{E}_\mu \left[\sum_{j=1}^s n(\mathbf{y}_j - \boldsymbol{\mu})(\mathbf{y}_j - \boldsymbol{\mu})' \right] \right\} \Sigma_W^{-1} \mathbf{U} \mathbf{C}$$

with respect to the distribution of $\boldsymbol{\mu} | \mathbf{y}, \boldsymbol{\theta}$. On account of assumptions made in [1], the distribution of this random variable is $N[\mathbf{y}_{..}, (\Sigma_B/s) + (\Sigma_W/ns)]$, so that:

$$\begin{aligned} \mathbf{E}_\mu \left[\sum_{j=1}^s n(\mathbf{y}_j - \boldsymbol{\mu})(\mathbf{y}_j - \boldsymbol{\mu})' \right] &= n \sum_{j=1}^s (\mathbf{y}_j - \mathbf{y}_{..})(\mathbf{y}_j - \mathbf{y}_{..})' \\ &+ ns \text{Var}(\boldsymbol{\mu} | \mathbf{y}, \boldsymbol{\theta}) = (s-1)\mathbf{B} + \boldsymbol{\Gamma} \end{aligned} \quad [\text{A4}]$$

The formula for the variance of prediction error in [A3b] does not depend on $\boldsymbol{\mu}$, so that the expression of [A1] reduces to:

$$\boldsymbol{\Lambda}^{[t]} = n \mathbf{C}^{[t]} \mathbf{U}' \left(\Sigma_W^{[t]} \right)^{-1} \left[(s-1)\mathbf{B} + \boldsymbol{\Gamma}^{[t]} \right] \left(\Sigma_W^{[t]} \right)^{-1} \mathbf{U} \mathbf{C}^{[t]} + s \mathbf{C}^{[t]} \quad [\text{A5}]$$

where $\mathbf{C}^{[t]}$ is defined as in [A3b] using $\boldsymbol{\theta}^{[t]}$ as a current estimate of $\boldsymbol{\theta}$ in Σ_W and Σ_B , the matrices \mathbf{B} and \mathbf{U} being constant over rounds of iteration.

The next values of the unknown components of Σ_B (ie σ_B^2 and C_B or δ_1 and τ) are computed at the M step from the diagonal elements of $\boldsymbol{\Lambda}^{[t]}$ ([18a] and [18b]).

The diagonal terms of the matrix $\mathbf{Q} = \sum_{j=1}^s \sum_{k=1}^n \mathbf{e}_{jk} \mathbf{e}_{jk}'$ are sufficient statistics for Σ_W under the complete data set \mathbf{x} since:

$$\mathbf{E}(Q_{ii}) = ns\sigma_{W_i}^2 \quad \text{for } i = 1, 2, \dots, p \quad [\text{A6}]$$

Because this statistic is not observable, it is replaced by its conditional expectation given the data \mathbf{y} and $\boldsymbol{\theta} = \boldsymbol{\theta}^{[t]}$. As for B-components, this expectation is calculated after conditioning and deconditioning with respect to the mean vector $\boldsymbol{\mu}$, ie:

$$\boldsymbol{\Omega}^{[t]} = \mathbf{E}(\mathbf{Q} | \mathbf{y}, \boldsymbol{\theta} = \boldsymbol{\theta}^{[t]}) = \mathbf{E}_\mu \left[\mathbf{E}(\mathbf{Q} | \mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta} = \boldsymbol{\theta}^{[t]}) \right] \quad [\text{A7}]$$

By definition of a quadratic form, one has:

$$\begin{aligned} \mathbf{E} \left(\sum_{j=1}^s \sum_{k=1}^n \mathbf{e}_{jk} \mathbf{e}_{jk}' | \mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta} \right) &= \sum_{j=1}^s \sum_{k=1}^n \mathbf{E}(\mathbf{e}_{jk} | \mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta}) \mathbf{E}(\mathbf{e}_{jk}' | \mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta}) \\ &+ \sum_{j=1}^s \sum_{k=1}^n \text{Var}(\mathbf{e}_{jk} | \mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta}) \end{aligned} \quad [\text{A8}]$$

Now:

$$\mathbf{E}(\mathbf{e}_{jk} | \mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta}) = \widehat{\mathbf{e}}_{jk} = \mathbf{y}_{jk} - \boldsymbol{\mu} - \mathbf{U} \widehat{\mathbf{f}}_j \quad [\text{A9a}]$$

$$\text{Var}(\mathbf{e}_{jk}|\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta}) = \sum_{j=1}^s \sum_{k=1}^n \mathbf{U} \text{Var}(\mathbf{f}_j|\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\theta})\mathbf{U}' = ns\mathbf{UCU}', \quad [\text{A9b}]$$

where $\widehat{\mathbf{f}}_j$ and \mathbf{C} are as before.

From [A9a], the quadratic $\sum_{jk} \widehat{\mathbf{e}}_{jk} \widehat{\mathbf{e}}'_{jk}$ can be written as

$$\begin{aligned} \sum_{jk} \widehat{\mathbf{e}}'_{jk} \widehat{\mathbf{e}}_{jk} &= \sum_{jk} (\mathbf{y}_{jk} - \boldsymbol{\mu})(\mathbf{y}_{jk} - \boldsymbol{\mu})' + \mathbf{M} \left[\sum_{jk} (\mathbf{y}_j - \boldsymbol{\mu})(\mathbf{y}_j - \boldsymbol{\mu})' \right] \mathbf{M}' \\ &\quad - \mathbf{M} \sum_{jk} (\mathbf{y}_j - \boldsymbol{\mu})(\mathbf{y}_{jk} - \boldsymbol{\mu})' - \sum_{jk} (\mathbf{y}_{jk} - \boldsymbol{\mu})(\mathbf{y}_j - \boldsymbol{\mu})' \mathbf{M}' \quad [\text{A10}] \end{aligned}$$

with $\mathbf{M} = n\mathbf{UCU}'\boldsymbol{\Sigma}_W^{-1}$.

The next step is to take the expectation of each element in the right-hand side with respect to the distribution of $\boldsymbol{\mu}|\mathbf{y}, \boldsymbol{\theta}$. Then

$$\begin{aligned} E_{\boldsymbol{\mu}} \left[\sum_{jk} (\mathbf{y}_{jk} - \boldsymbol{\mu})(\mathbf{y}_{jk} - \boldsymbol{\mu})' \right] &= \sum_{jk} (\mathbf{y}_{jk} - \mathbf{y}_{..})(\mathbf{y}_{jk} - \mathbf{y}_{..})' \\ &\quad + \sum_{jk} \text{Var}(\boldsymbol{\mu}|\mathbf{y}, \boldsymbol{\theta}) = (ns - 1)\mathbf{T} + \boldsymbol{\Gamma} \quad [\text{A11}] \end{aligned}$$

where \mathbf{T} designates the matrix of total mean squares and mean cross-products, Similarly,

$$\begin{aligned} E_{\boldsymbol{\mu}} \left[\sum_{jk} (\mathbf{y}_j - \boldsymbol{\mu})(\mathbf{y}_j - \boldsymbol{\mu})' \right] &= \sum_{jk} (\mathbf{y}_j - \mathbf{y}_{..})(\mathbf{y}_j - \mathbf{y}_{..})' \\ &\quad + \sum_{jk} \text{Var}(\boldsymbol{\mu}|\mathbf{y}, \boldsymbol{\theta}) = (s - 1)\mathbf{B} + \boldsymbol{\Gamma} \quad [\text{A12}] \end{aligned}$$

$$\begin{aligned} E_{\boldsymbol{\mu}} \left[\sum_{jk} (\mathbf{y}_j - \boldsymbol{\mu})(\mathbf{y}_{jk} - \boldsymbol{\mu})' \right] &= \sum_{jk} (\mathbf{y}_j - \mathbf{y}_{..})(\mathbf{y}_{jk} - \mathbf{y}_{..})' \\ &\quad + \sum_{jk} \text{Var}(\boldsymbol{\mu}|\mathbf{y}, \boldsymbol{\theta}) = n \sum_j (\mathbf{y}_j - \mathbf{y}_{..})(\mathbf{y}_j - \mathbf{y}_{..})' \\ &\quad + n\boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_W = (s - 1)\mathbf{B} + \boldsymbol{\Gamma} \quad [\text{A13}] \end{aligned}$$

Placing [A11], [A12] and [A13] in [A8] and noting that the conditional variance in [A9b] does not depend on $\boldsymbol{\mu}$, gives

$$\begin{aligned} \boldsymbol{\Omega} &= (ns - 1)\mathbf{T} + \boldsymbol{\Gamma} + \mathbf{M}[(s - 1)\mathbf{B} + \boldsymbol{\Gamma}]\mathbf{M}' \\ &\quad - \mathbf{M}[(s - 1)\mathbf{B} + \boldsymbol{\Gamma}] - [(s - 1)\mathbf{B} + \boldsymbol{\Gamma}]\mathbf{M}' + ns\mathbf{UCU}' \quad [\text{A14}] \end{aligned}$$

In fact, $\boldsymbol{\Omega}$ is evaluated conditionally on $\boldsymbol{\theta} = \boldsymbol{\theta}^{[t]}$, ie by taking $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{[t]}$, $\mathbf{C} = \mathbf{C}^{[t]}$ and $\mathbf{M} = \mathbf{M}^{[t]}$ in [A.14]. The next value of $\sigma_{W_i}^2$ is obtained from [A.6] (M step) with Ω_{ii} replacing Q_{ii} , ie by $\sigma_{W_i}^{2[t+1]} = \Omega_{ii}^{[t]}/ns$.

Σ_B and Σ_W under the saturated model

Actually, the EM algorithm described previously can be easily accomodated to deal with the saturated model. This is especially helpful when the ANOVA estimates of Σ_B fall outside the parameter space.

Nothing is changed with respect to Σ_W , which has the same diagonal structure with p different elements in both situations. As far as Σ_B is concerned, a sufficient statistic under the complete data set is now the $(p \times p)$ matrix $\sum_j \mathbf{b}_j \mathbf{b}'_j = \mathbf{U} \left(\sum_j \mathbf{f}_j \mathbf{f}'_j \right) \mathbf{U}'$. However, for a given \mathbf{U} , the general expression of the conditional expectation of $\sum_j \mathbf{f}_j \mathbf{f}'_j$ given the data was already derived (see [19] and [A5], so that the E step remains the same.

Because all the elements of Λ are now required, the changes to implement at the M step are the following: compute the next value $\Sigma_B^{[t+1]}$ of Σ_B (saturated model) by

$$\Sigma_B^{[t+1]} = (\mathbf{U}^{[t]} \Lambda^{[t]} \mathbf{U}'^{[t]}) / s \tag{A15}$$

where $\Lambda^{[t]}$ is obtained from [A5] with $\mathbf{U}^{[t]}$ being the matrix of normed eigenvectors of $\Sigma_B^{[t]}$. Notice that here \mathbf{U} is updated at each iteration from the equation $\Sigma_B^{[t]} \mathbf{U}^{[t]} = \mathbf{U}^{[t]} \Lambda^{[t]}$.

Algorithms based on first derivatives (part B)

Between family components Σ_B

Using the spectral decomposition in [13], *ie* $\Gamma = n \mathbf{U}' \Delta \mathbf{U} + \Sigma_W$, we obtain

$$\frac{\partial \Gamma}{\partial \delta_i} = n \sum_{\ell=1}^{r_i} \mathbf{U}_{i_\ell} \mathbf{U}'_{i_\ell} \tag{B1}$$

where $\mathbf{U}_{i_1}, \mathbf{U}_{i_2}, \dots, \mathbf{U}_{i_\ell}, \dots, \mathbf{U}_{i_{r_i}}$ are the r_i normed eigenvectors of Σ_B corresponding to the eigenvalue δ_i with multiplicity order r_i . Remember that under the reduced model, $r_1 = 1$ and $r_2 = p - 1$ for δ_1 and δ_2 defined in [14a] and [14b], respectively.

Substituting $\partial \Gamma / \partial \delta_i$ by its expression [B1] in the equation $\partial(-2L) / \partial \delta_i = 0$ leads to:

$$\text{tr} \left(\Gamma^{-1} \sum_{\ell=1}^{r_i} \mathbf{U}_{i_\ell} \mathbf{U}'_{i_\ell} \right) = \text{tr} \left(\Gamma^{-1} \mathbf{B} \Gamma^{-1} \sum_{\ell=1}^{r_i} \mathbf{U}_{i_\ell} \mathbf{U}'_{i_\ell} \right) \tag{B2}$$

Let $\mathbf{L} = \sqrt{n} \mathbf{U} \Delta^{1/2}$ with \mathbf{L} partitioned in the same way as \mathbf{U} , *ie* $\mathbf{L}_{i_\ell} = \sqrt{n \delta_i} \mathbf{U}_{i_\ell}$, the system [B2] is then equivalent to:

$$\sum_{\ell=1}^{r_i} \mathbf{L}'_{i_\ell} \Gamma^{-1} \mathbf{L}_{i_\ell} = \sum_{\ell=1}^{r_i} \mathbf{L}'_{i_\ell} \Gamma^{-1} \mathbf{B} \Gamma^{-1} \mathbf{L}_{i_\ell}$$

or

$$\sum_{\ell=1}^{r_i} (\mathbf{L}'\mathbf{\Gamma}^{-1}\mathbf{L})_{i_\ell i_\ell} = \sum_{\ell=1}^{r_i} (\mathbf{L}'\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}\mathbf{L})_{i_\ell i_\ell} \quad [\text{B3}]$$

where $(\mathbf{A})_{i_\ell i_\ell}$ stands for the i_ℓ th diagonal element of the \mathbf{A} matrix.

Now:

$$\mathbf{\Gamma} = \mathbf{L}\mathbf{L}' + \mathbf{\Sigma}_W$$

and

$$\begin{aligned} \mathbf{\Gamma}\mathbf{\Sigma}_W^{-1}\mathbf{L} &= (\mathbf{L}\mathbf{L}' + \mathbf{\Sigma}_W)\mathbf{\Sigma}_W^{-1}\mathbf{L} \\ &= \mathbf{L}(\mathbf{L}'\mathbf{\Sigma}_W^{-1}\mathbf{L}) + \mathbf{L} \\ &= \mathbf{L}(\mathbf{\Psi} + \mathbf{I}_p) \end{aligned}$$

with

$$\mathbf{\Psi} = \mathbf{L}'\mathbf{\Sigma}_W^{-1}\mathbf{L} = \mathbf{L}'\mathbf{\Gamma}^{-1}\mathbf{L}(\mathbf{\Psi} + \mathbf{I}_p) \quad [\text{B4}]$$

Then, from [B4]

$$\mathbf{L}'\mathbf{\Gamma}^{-1}\mathbf{L} = \mathbf{\Psi}(\mathbf{\Psi} + \mathbf{I}_p)^{-1} \quad [\text{B5}]$$

and

$$\mathbf{L}'\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}\mathbf{L} = (\mathbf{\Psi} + \mathbf{I}_p)^{-1}\mathbf{L}'\mathbf{\Sigma}_W^{-1}\mathbf{B}\mathbf{\Sigma}_W^{-1}\mathbf{L}(\mathbf{\Psi} + \mathbf{I}_p)^{-1} \quad [\text{B6}]$$

Furthermore:

$$\mathbf{\Delta}^{-1/2}(\mathbf{\Psi} + \mathbf{I}_p)\mathbf{\Delta}^{-1/2} = \mathbf{\Delta}^{-1/2}\mathbf{L}'\mathbf{\Sigma}_W^{-1}\mathbf{L}\mathbf{\Delta}^{-1/2} + \mathbf{\Delta}^{-1} = n\mathbf{U}'\mathbf{\Sigma}_W^{-1}\mathbf{U} + \mathbf{\Delta}^{-1} = \mathbf{C}^{-1}$$

or

$$(\mathbf{\Psi} + \mathbf{I}_p)^{-1} = \mathbf{\Delta}^{-1/2}\mathbf{C}\mathbf{\Delta}^{-1/2}$$

so that

$$\mathbf{L}'\mathbf{\Gamma}^{-1}\mathbf{L} = \mathbf{I}_p - (\mathbf{\Psi} + \mathbf{I}_p)^{-1} = \mathbf{\Delta}^{-1/2}(\mathbf{\Delta} - \mathbf{C})\mathbf{\Delta}^{-1/2} \quad [\text{B7}]$$

and

$$\mathbf{L}'\mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{\Gamma}^{-1}\mathbf{L} = n\mathbf{\Delta}^{-1/2}\mathbf{C}\mathbf{U}'\mathbf{\Sigma}_W^{-1}\mathbf{B}\mathbf{\Sigma}_W^{-1}\mathbf{U}\mathbf{C}\mathbf{\Delta}^{-1/2} \quad [\text{B8}]$$

Using [B7] and [B8], the system [B3] reduces to

$$\sum_{i_\ell}^n (\mathbf{\Delta} - \mathbf{C})_{i_\ell i_\ell} = \sum_{i_\ell}^{r_i} n(\mathbf{C}\mathbf{U}'\mathbf{\Sigma}_W^{-1}\mathbf{B}\mathbf{\Sigma}_W^{-1}\mathbf{U}\mathbf{C})_{i_\ell i_\ell} \quad [\text{B9}]$$

The EM procedure described in [18a], [18b] and [19] can be alternatively written as

$$\sum_{i_\ell}^{r_i} (\mathbf{\Delta})_{i_\ell i_\ell} = \sum_{i_\ell}^{r_i} (\mathbf{C})_{i_\ell} + n(\mathbf{C}\mathbf{U}'\mathbf{\Sigma}_W^{-1} \frac{(s-1)\mathbf{B} + \mathbf{\Gamma}}{s} \mathbf{\Sigma}_W^{-1}\mathbf{U}\mathbf{C})_{i_\ell i_\ell} \quad [\text{B10}]$$

The parallel between [B9] and [B10] is straightforward. Here \mathbf{B} replaces $[(s-1)\mathbf{B} + \mathbf{\Gamma}]/s$, since these 2 quantities have the same expectation $\mathbf{\Gamma}$ under the saturated model.

Residual components Σ_W

Here, $\frac{\partial \mathbf{\Gamma}}{\partial \sigma_{W_i}^2} = \frac{\partial \Sigma_W}{\partial \sigma_{W_i}^2}$, $\left(\frac{\partial \Sigma_W}{\partial \sigma_{W_i}^2}\right)_{ii} = 1$ and $\left(\frac{\partial \Sigma_W}{\partial \sigma_{W_i}^2}\right)_{ii'} = 0$ for any $i \neq i'$. Then, the equation [8b] = 0 can be written as

$$(s-1) [(\mathbf{\Gamma}^{-1})_{ii} - (\mathbf{\Gamma}^{-1} \mathbf{B} \mathbf{\Gamma}^{-1})_{ii}] + s(n-1) \left[\frac{1}{\sigma_{W_i}^2} - \frac{W_i}{\sigma_{W_i}^4} \right] = 0 \quad [\text{B11}]$$

Equation [B11] defines a non-linear system that can be solved iteratively using, for example, the functional iteration approach

$$\sigma_{W_i}^{2[t+1]} = W_i + \frac{s-1}{s(n-1)} \varphi_i^{[t]} \sigma_{W_i}^{4[t]} \quad [\text{B12a}]$$

with

$$\varphi_i^{[t]} = \left[(\mathbf{\Gamma}^{[t]})^{-1} \mathbf{B} (\mathbf{\Gamma}^{[t]})^{-1} \right]_{ii} - \left[(\mathbf{\Gamma}^{[t]})^{-1} \right]_{ii} \quad [\text{B12b}]$$