Mixed Model Formulations for Multi-Environment Trials

K.E. Basford\textsuperscript{1*}, W.T. Federer\textsuperscript{2}, and I.H. DeLacy\textsuperscript{1}

\textsuperscript{1}The University of Queensland, Brisbane Qld 4072, Australia

\textsuperscript{2}Cornell University, Ithaca NY 14850, USA

*Corresponding author: k.e.basford@uq.edu.au
Mixed Model Formulations for Multi-Environment Trials

Abstract

When studying genotype by environment (GxE) interaction in multi-environment trials, plant breeders and geneticists often consider one of the effects, environments or genotypes, to be fixed and the other to be random. However, there are two main formulations for variance component estimation for the mixed model situation, referred to as the unconstrained-parameters and constrained-parameters formulations. These formulations give different estimates of genetic correlation and heritability, as well as different tests of significance for the random effects factor. The definition of main effects and interactions and the consequences of such definitions should be clearly understood, and any formulation selected should be consistent for both fixed and random effects. A discussion of the practical outcomes of using the two formulations in the analysis of data from multi-environment trials is presented. It is recommended that the constrained-parameters formulation be used because of the meaning of its parameters and the corresponding variance components. When managed (fixed) environments are considered, users will have more confidence in prediction for them, but will not be overconfident in prediction in the target (random) environments. On the other hand, the genetic gain (predicted response to selection in the target environments from the managed environments) is independent of formulation.

Abbreviations: GxE Genotype by Environment; MET Multi-Environment Trial
When studying genotype by environment (GxE) interaction, breeders and geneticists often consider one of the two factors to be fixed and the other to be random. This results in what statisticians have dubbed the mixed model situation. For the fixed effect, all the levels in the population of parameters are present, while for the random factor, only a random sample from the population of levels is obtained. The experimenter often wishes to obtain estimates of variance components in order to compute genetic correlation, heritability estimates, repeatability estimates, genetic advance estimates, and other related statistics. Several discussions of variance component estimation in the mixed model situation have appeared in the literature (Federer, 1955; Cornfield and Tukey, 1956; Scheffe, 1956, 1959; Hocking, 1973; Ayres and Thomas, 1990; Samuels, Casella and McCabe, 1991; Fry, 1992; Schwarz, 1993; Searle, Casella and McCulloch, 1992; Voss, 1999). Different formulations have been put forward with two of these being used most frequently. This poses a dilemma for the breeder and geneticist as to which formulation to use as they give different estimates of genetic correlation and heritability, as well as different tests of significance for the random effects factor.

Hocking (1973), Samuels et al. (1991), Fry (1992), Schwarz (1993) and others have attempted to deal with this dilemma by giving what they consider to be justifications for each of these two formulations. Despite the extensive literature and the well-written paper by Voss (1999) on resolving the controversy, it appears that confusion still reigns with regard to practical interpretation, particularly by plant breeders. We believe that the definition of main effects and interactions and the consequences of such definitions
should be clearly understood, and any formulation should be consistent for both fixed
and random effects.

There are several variants of the two main formulations (see for example, Scheffe, 1959;
Hocking, 1973; Searle et al., 1992) that Fry (1992) called the Scheffe and SAS
formulations (since one of the formulations has been programmed into the SAS software
package for Proc Mixed and Proc VarComp). These have been more informatively
called the constrained-parameters (CP) formulation and unconstrained-parameters (UP)
formulation, respectively, by Voss (1999). We shall use this terminology here, with our
focus being on the effect of these two formulations on the genetic inference.

The structure of this paper is such that some agricultural background material to the
situation in which this problem arises is initially presented, then the statistical issues and
genetic issues are discussed in turn. The application of the two formulations is
illustrated using an example from the wheat breeding literature. Finally, some
conclusions and recommendations are presented.

**Agricultural Background**

Data sets obtained from the study of genotype-environment systems are usually
generated by evaluating candidate breeding lines (the genotypes) in a set of
environments. The environments are often considered to have been sampled from some
target population of environments in a series of experiments, referred to as multi-
environment trials (METs). As the process of sampling environments is generally
associated with testing the genotypes at a number of sites for a number of years,
environments are commonly defined as particular site-year combinations. Genotype by
environment (G×E) interactions are detected as a significantly different pattern of
response among the genotypes across environments, i.e. there is a significant difference
in the relative performance of the genotypes when they are grown in different
environments. Clearly, if there were no G×E interactions associated with the genotype-
environment system relevant to a breeding objective, selection would be simple because
the ‘best’ genotype in one environment would also be the ‘best’ genotype in all target
environments. Experience has indicated that G×E interactions are the norm (certainly in
Australia), rather than the exception, and they have considerable impact on selection for
 genetic improvement.

Cooper et al. (1995) hypothesised that regional testing strategies could be improved by
accommodating the effects of G×E interactions to maximise the response to selection.
They argued that one way of doing this was to identify the set of selection environments
most relevant to the future production-environments. If these test environments can be
repeated from year to year, confidence in predicting response in future environments
would be increased. They therefore assessed the scope for managing environmental
conditions at a restricted number of sites to provide discrimination among wheat lines
for grain yield that matches that in target production-environments.

In analyzing data from such a multi-environment testing regime, the genotypes can be
considered to be a random sample of the lines from the relevant stage of the breeding
programme. It seems reasonable to consider the managed environments (M) to be fixed
as they can be repeated over years and locations. Hence, a mixed model for the
genotype-environment system will be appropriate for data from these managed
environments. However, the interpretation of experimental results and any inference
from selection will apply to the target or production environments (T) which could most
reasonably be considered to be random. Cooper et al. (1995) argued that a successful
breeding strategy is one that gives a high indirect response to selection for average yield
over the production-environments and quantified this using the genetic correlation which
measured the similarity of line discrimination between the managed-environment
selection regime and that for average performance in the production-environments.
Thus, a combination of statistical and biological approaches is needed.

Statistical Issues
Cornfield and Tukey (1956) wanted a single flexible model to obtain the average values
of mean squares in factorials, the simplest of which is the \( r \) replicate \( a \times b \) factorial
experiment. This was achieved by Tukey (1949) and independently by Cornfield (1953)
and Wilk (1953). They stated that the choice of assumptions depends on more than
empirical questions about the behaviour of the experimental material. It depends on the
nature of the sampling and randomization involved in obtaining the data. Moreover, it
often depends on the purpose of the analysis, as expressed by the situations or
populations to which one wishes to make statistical inference. These dependencies
imply diversity, and adequate treatment of diversity requires flexibility of assumptions.
Thus, even fifty years ago the importance of looking at these models from different
perspectives was noted.
Voss (1999) described the two main formulations for the two-factor mixed model and put forward a resolution. The material in this section follows his presentation. The different two-factor models will be denoted as the unconstrained-parameters (UP) formulation and the constrained-parameters (CP) formulation. These are equivalent to the SAS (SAS (1990) software) and Scheffe (Scheffe, 1956, 1959) formulations, respectively.

The UP formulation for the \( r \) replicate \( axb \) factorial experiment with factor \( A \) fixed and factor \( B \) random is based on the following model for \( y_{ijk} \), the response for the \( k^{th} \) replicate of the \( j^{th} \) level of factor \( B \) and the \( i^{th} \) level of factor \( A \):

\[
y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + E_{ijk} \quad (1)
\]

where

\[
B_j \sim N(0, \sigma^2_B),
\]

\[
(\alpha B)_{ij} \sim N(0, \sigma^2_{\alpha B}),
\]

\[
E_{ijk} \sim N(0, \sigma^2),
\]

all terms are mutually independent, and \( i=1, ..., a; j=1, ..., b; k=1, ..., r. \)

The CP formulation for this same experiment is based on the following model for \( y_{ijk} \):

\[
y_{ijk} = \mu + \tau_i + D_j + (\tau D)_{ij} + E_{ijk} \quad (2)
\]

where
A major distinction between these two models is the generality of the CP model in allowing the covariance between \( y_{ijk} \) and \( y_{ij'k'} \) to be negative (Harville, 1978; Schwarz, 1993). In our agricultural example where factor \( A \) corresponds to managed environments and factor \( B \) corresponds to genotypes, this would allow a negative correlation between the responses for the same genotype in different managed environments. While many authors have noted that one model is simply a reparameterisation of the other, this does not help a plant breeder decide which of these two formulations should be chosen and subsequently interpreted.

The heart of the problem is in the expected mean squares for the analysis of variance of models (1) and (2), as given in Table 1. It would appear that under the UP formulation one would test \( H_0: \sigma^2_B = 0 \) by MSB/MSAB, but under the CP formulation one would test \( H_0: \sigma^2_D = 0 \) by MSB/MSE. The relationship between these variance components

\[
\sigma^2_D = \sigma^2_B + \sigma^2_{ab}/a \quad (3)
\]
\[ \sigma^2_{tD} = \sigma^2_{alt} \]  

(4)

does not clarify things as the plant breeder still needs to interpret the particular parameters in models (1) or (2).

In order to better understand the parameters, Voss (1999) constructed superpopulation models from which the UP and CP models could be induced. In particular, he showed that each parameter in the CP model is a main effect or interaction effect in the usual sense of deviations amongst means. Although he did not say so, this provides consistency across fixed and random effects. The parameters in the UP formulation are not main effects or an interaction effect in the usual sense because there are no constraints on the effects.

Voss concluded that the “bottom line is ... that ... the parameters and corresponding variance components in the CP mixed model correspond to specific main effects or interaction effects, and the analysis of variance tests motivated by consideration of the corresponding expected mean squares under the CP formulation are appropriate for testing the corresponding main effects and interactions under both the CP and UP models”. This is because the expected value of \( MSB \) under both the CP and UP models measures error variability plus main effects of \( B \). Thus the appropriate test of main effects of \( B \) under both mixed models is \( MSB/MSE \).
Another way of thinking about the situation is that for a random sample of genotypes, one has all of the interaction terms across the fixed environments. It can be argued that, under those circumstances, the sum of interaction terms should be zero, as in the CP model. Thus the CP model is consistent for both fixed and random effects. The UP model sets the expectation of the interaction terms to be zero over the population. But as all of the interaction terms in this population are present, the UP model is not consistent for both fixed and random effects.

Genetic Issues

Much quantitative genetic theory has been developed from the two-way model, particularly when both factors (genotypes and environments) are assumed to be random. The two concepts on which this theory is based are heritability (in the broad sense) and predicted genetic gain (or predicted response to selection) (Falconer, 1981). To understand their meaning, it is important that certain other parameters are defined with respect to the parameters in the associated statistical model. In this instance, they will be defined with respect to the mixed model (for managed environments) and for the fully random model (for target environments). The estimators for fixed effects are called best linear unbiased estimators (BLUEs) and those for random effects are called best linear unbiased predictors (BLUPs) (Henderson, 1963, 1977).

Selection amongst genotypes is based on phenotypic variance and the phenotypic variance on a line mean basis is determined directly from the expected mean square for genotypes from the analysis of variance of the data (Table 1). Thus for the managed environments
\[ \sigma^2_{p(M)} = \sigma^2_B + \sigma^2_{aB} / a + \sigma^2 / (ar) \]  
for the UP formulation and

\[ \sigma^2_{p(M)} = \sigma^2_D + \sigma^2 / (ar) \]  
for the CP formulation.

On the other hand, response to selection is based on genotypic variance. Again, for the managed environments

\[ \sigma^2_{g(M)} = \sigma^2_B \]  
for the UP formulation and

\[ \sigma^2_{g(M)} = \sigma^2_D \]  
for the CP formulation.

The heritability of genotype means in the managed environments is defined as the ratio of the genetic variance to the phenotypic variance:

\[ h^2_M = \sigma^2_{g(M)} / \sigma^2_{p(M)} \quad (5) \]

using either the UP or CP formulation for both of these variances. The heritability in the targeted environments is defined similarly, but the fully random model is assumed.

The phenotypic correlation, \( r_{p(M,T)} \), is calculated between the means of the genotype performance in the managed and production environments. The genetic correlation, \( r_{g(M,T)} \), measures the similarity of line discrimination between the managed-environments.
selection regime and that for average performance in the production-environments.

When the environment correlation from managed (M) to production (T) environments can be assumed zero (Burdon, 1977), as in this case, the relationship between the phenotypic and genetic correlations is

\[ r_{g(M,T)} = r_{p(M,T)} / (h_T h_M) \]  

where \( h^2_T \) and \( h^2_M \) are the heritabilities in the target and managed environments, respectively.

The predicted response to selection (or genetic gain) in the environment \( l \) where selection is made, \( \Delta G_l \), is given by

\[ \Delta G_l = i h^2_l \sigma_p(l) \]  

where \( i \) is the standardized selection differential, \( h^2_l \) is the heritability on a line mean basis and \( \sigma_p(l) \) is the phenotypic standard deviation in environment \( l \). This equation can be applied to selection for specific traits, such as resistance or tolerance to disease, pest or soil toxicity factors, when genotypes are exposed to the appropriate screen. Error variation reduces genetic gain as can be seen from the definition of heritability on a genotype means basis as the ratio of the genotypic to phenotypic variance.
Extending this concept to the common case where the environments in which selection is made are a sample of the target environments (T), the predicted response to selection in those target environments, \( \Delta G_T \), is given by

\[
\Delta G_T = i h^2_T \sigma_{p(T)}
\]  

where \( i \) is the standardised selection differential, \( h^2_T \) is the heritability on a line mean basis and \( \sigma_{p(T)} \) is the phenotypic standard deviation in the target environments. Variation due to GxE interaction decreases genetic gain as it is incorporated in the denominator in the definition of heritability.

When prediction is desired from a test set of managed environments (M) to a target set of environments (T), the predicted response to selection (correlated genetic gain), \( \Delta G_{(T|M)} \), is given by

\[
\Delta G_{(T|M)} = i h_T h_M r_{g(M,T)} \sigma_{p(T)} = i r_{p(M,T)} \sigma_{p(T)}
\]

where there is no error correlation among the managed and target environments, \( h_T \) and \( h_M \) are the square roots of the heritabilities of line means in the target and test environments, respectively, \( r_{g(M,T)} \) and \( r_{p(M,T)} \) are the genetic and phenotypic correlations between mean performance in the test and target environments, respectively, and \( \sigma_{p(T)} \) is the phenotypic standard deviation in the target environments. A more detailed
description of the derivation and interpretation of these statistics is given in Cooper et al. (1996).

Example

The data being considered here arose from trials conducted in a set of managed environments by the Queensland wheat breeding programme in Australia (Cooper et al., 1995). Grain yield (t ha\(^{-1}\)) was measured on 15 sampled lines which included three local check cultivars, one line from the 11\(^{th}\) International Bread Wheat Screening Nursery and 11 lines from the 17\(^{th}\) International Bread Wheat Screening Nursery. The 15 lines were evaluated in 18 managed environments. These were made up of six managed-environments at each of three locations, Emerald, Kingsthorpe (in 1988) and Gatton (in 1987 and 1988), and involved manipulating nitrogen availability, water and sowing date. They were evaluated in a randomized complete block design with two replicates in each managed-environment. A mixed model was adopted where the lines were random effects (as they were considered to be a random sample of the lines from the preliminary testing stage of the Queensland programme) and the managed-environments were fixed (as it was assumed that they represented known challenges which could be repeated over years). The estimation of variance components and genetic parameters was conducted using both the UP and CP formulations. The 15 lines were also evaluated in 10 target or production-environments over four years (1985 to 1988) in randomized complete block designs with three replicates in each environment. These were considered to be a random subset of the regional trials used by the Queensland wheat breeding programme (Brennan et al. 1981). Thus a completely random model was adopted for the
production-environment trials. More details may be found in Cooper et al. (1995) where two series of managed-environments were considered.

The resultant mean squares for genotypes, environments, GxE interaction and error for the data from the managed-environments (M) are presented in Table 2. As the focus here is on the interpretation from the mixed model, the mean squares for the data from the target or production environments (T) are not listed. The genetic parameter estimates using both the UP and CP formulations are presented in Table 3. Given the difference in the expected mean squares (Table 1), the estimate of the variance component for genotypes (i.e., the genetic variance) is greater for the CP formulation than for the UP formulation and consequently the line mean heritability is larger and the genetic correlation from the managed environments to the production environments is smaller for the CP formulation than for the UP formulation (Table 3). Irrespective of the formulation used, predicted genetic gain from managed to production environments ($\Delta G_{(T|M)} = 0.003$) remains the same as the phenotypic correlation ($r_{p(M,T)} = 0.56$) remains the same. This is in spite of the change in the estimated heritability in the managed environments.

The CP formulation puts more confidence in an ability to distinguish lines which are genetically better in the managed environments ($h^2_M = 0.968$) than does the UP formulation ($h^2_M = 0.896$) at the price of less confidence in prediction to production environments ($r_{g(M,T)}$ of 0.72 for CP and 0.78 for UP) (Table 3). This is compatible with the fixed model assumption for environments.
Another consequence is that the calculation of the best linear unbiased predictors (BLUPs) for genotypes will be affected by the different models with those using the CP formulation likely to overestimate the prediction of performance to the production environments. This arises because, for the completely random model, the BLUP for genotype performance across environments \( (p_j) \) is, in its heritability form (DeLacy et al., 1996), given by

\[
p_j = h^2_M \left( \bar{y}_{..} - \bar{y}_{..} \right),
\]

where \( \bar{y}_{..} \) is the mean genotype response across replicates and environments and \( \bar{y}_{..} \) is the overall mean response.

The heritability from the CP formulation \( (h^2_M = 0.968) \) is larger than that from the UP formulation \( (h^2_M = 0.896) \) and shrinks the BLUPs less. Unless there are unequal numbers, the correlation between these BLUPs and the raw genotype means over environments is one. Here, the usual assumptions of homogeneity are made, i.e. the error variance in each environment is the same and the G×E interaction variance is the same in each environment. General mixed model theory allows both assumptions to be relaxed.

The advantage of using BLUPs for prediction is that the predicted range is near to the "actual" range, i.e. the range of performance in the target environments. The arithmetic
average gives too large a spread. BLUPs also allow for different adjustment of the means depending on the number of replications: those means calculated from a large number of observations are shrunk less. Check genotypes usually have more replications and it is reasonable to assume that their means are known more reliably, and in consequence should be adjusted less for prediction purposes. This shrinkage is, in one sense, what was meant with the phrase, "regression to the mean" – the means of a selected group, when they are re-evaluated, will be nearer the mean of the unselected group than their means from test data. The use of BLUPs in selection based on multi-environment trials is discussed by Gilmour et al. (1996).

Discussion and Conclusion

Breeders are setting up fixed "managed" environments with known variables that contribute to genotype by environment interaction. This is opposed to selecting a "random sample" of environments from the population of environments in which a genotype will be grown. It is doubtful if a truly random sample of environments could be obtained anyway. The finite set of managed environments leads directly to a mixed model situation when the genotypes represent a random sample from the population of genotypes.

If the definition of main effects and interactions universally used in factorial experiments is acceptable, then the CP formulation is the correct one for the breeder to use. The inconsistencies associated with the UP formulation in going from fixed to random effects makes this formulation undesirable. Samuels et al. (1991) also prescribe the CP formulation as the appropriate one but for different reasons than those given.
A number of authors (e.g., Ayres and Thomas, 1990, Fry, 1992, Schwarz, 1993) have attempted to justify each of the formulations based on their covariance structures. The nature of the covariance structure arises from the finiteness of the population and from the way the response model is formulated. The latter item is not related to the population structure and properties but to the mathematical properties of the manner in which the model is written. Regardless of the algebraic properties and mathematical generality, a model is uninformative if it does not have practical value.

As far as hypothesis testing is concerned, it is irrelevant whether data are balanced or not, i.e. the sampling procedure does not change the hypothesis. The population parameters that are being estimated are not different in concept, even if the actual estimates are different.

For data collected over a period of years, it is recommended that breeders obtain estimates of the genotype and genotype by environment interaction components of variance by the two formulations and from ANOVA and REML methods. Then, the results for genetic correlations and heritabilities can be computed for all estimates and compared with the actual values achieved in the program. Such summarizations and applications will verify the validity of any particular procedure for the breeding program in question. Differences from actual can be obtained and compared for all the procedures.

Overall, we recommend the CP formulation because of meaning of its parameters and corresponding variance components. Users will be more confident in prediction in the
managed environments (M), but not overconfident in prediction in the target environments (T). Importantly, the genetic gain (predicted response to selection in the target environments from the managed environments) is the same under both formulations.

References


Table 1: Expected mean squares for the $r$ replicate $a \times b$ factorial experiment with factor $A$ fixed and factor $B$ random under the unconstrained-parameters (UP) and constrained-parameters (CP) formulations.

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>EMS for UP formulation</th>
<th>EMS for CP formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$ (env)</td>
<td>$MSA$</td>
<td>$\sigma^2 + r \sigma^2_{ab} + br s^2(\alpha_j)$</td>
<td>$\sigma^2 + r \sigma^2_{bD} + br s^2(\tau_j)$</td>
</tr>
<tr>
<td>$B$ (gen)</td>
<td>$MSB$</td>
<td>$\sigma^2 + r \sigma^2_{ab} + ar \sigma^2_B$</td>
<td>$\sigma^2 + ar \sigma^2_D$</td>
</tr>
<tr>
<td>$AB$ (gen×env)</td>
<td>$MSAB$</td>
<td>$\sigma^2 + r \sigma^2_{ab}$</td>
<td>$\sigma^2 + r \sigma^2_{bD}$</td>
</tr>
<tr>
<td>Error</td>
<td>$MSE$</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

For $s^2(\alpha_j) = \Sigma(\alpha_i - \bar{\alpha})^2 / (a-1)$, $\bar{\alpha} = \Sigma \alpha_i / a$ and $s^2(\tau_j) = \Sigma \tau_i^2 / (a-1)$.
Table 2: Mean squares from the analysis of the grain yield (t ha\(^{-1}\)) of 15 genotypes grown in randomized complete block designs within each of 18 managed environments.

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) (env)</td>
<td>67.226</td>
</tr>
<tr>
<td>(B) (gen)</td>
<td>3.052</td>
</tr>
<tr>
<td>(AB) (gen\times env)</td>
<td>0.318</td>
</tr>
<tr>
<td>Error</td>
<td>0.099</td>
</tr>
</tbody>
</table>
Table 3: Genetic parameter estimates from the analysis of the grain yield (t ha\(^{-1}\)) of 15 genotypes grown in randomized complete block designs within each of 18 managed environments under the unconstrained-parameters (UP) and constrained-parameters (CP) formulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>UP formulation</th>
<th>CP formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma^2_g)</td>
<td>0.076</td>
<td>0.082</td>
</tr>
<tr>
<td>(\sigma^2_p)</td>
<td>0.085</td>
<td>0.085</td>
</tr>
<tr>
<td>(h^2_M)</td>
<td>0.896</td>
<td>0.968</td>
</tr>
<tr>
<td>(r_{g(M,T)})</td>
<td>0.78</td>
<td>0.72</td>
</tr>
<tr>
<td>(r_{p(M,T)})</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>(\Delta G_{(T</td>
<td>M)})</td>
<td>0.003</td>
</tr>
</tbody>
</table>