

Kansas State University Libraries

**New Prairie Press**

---

Conference on Applied Statistics in Agriculture

2011 - 23rd Annual Conference Proceedings

---

## COMPARISON OF LINEAR MIXED MODELS FOR MULTIPLE ENVIRONMENT PLANT BREEDING TRIALS

Carl A. Walker

Fabiano Pita

Kimberly Garland Campbell

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Walker, Carl A.; Pita, Fabiano; and Campbell, Kimberly Garland (2011). "COMPARISON OF LINEAR MIXED MODELS FOR MULTIPLE ENVIRONMENT PLANT BREEDING TRIALS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1054>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

# COMPARISON OF LINEAR MIXED MODELS FOR MULTIPLE ENVIRONMENT PLANT BREEDING TRIALS

Carl A. Walker<sup>1</sup>, Fabiano Pita<sup>2</sup>, Kimberly Garland Campbell<sup>1,3</sup>

<sup>1</sup> Dept. of Crop and Soil Sciences, Washington State University; <sup>2</sup> Quantitative Genetics Group, Dow AgroSciences; <sup>3</sup> USDA-ARS, Wheat Genetics, Quality, Physiology, and Disease Research Unit

## Abstract

Evaluations of multiple environment trials (MET) often reveal substantial genotype by environment interactions, and the effects of genotypes within environments are often estimated using cell means, i.e. the simple mean of the observations of each genotype in each environment. However, these estimates are inaccurate, especially for unreplicated or partially replicated trials, so alternative methods of analysis are necessary. One possible approach utilizes information, often from pedigree data, about relationships among the tested genotypes through the use of a genetic relationship matrix (GRM). Predictive accuracy may also be improved by the use of factor analytic (FA) structures for environmental covariances. In this study, data were simulated to resemble results from a range of MET. These simulated data sets covered a range of scenarios with varying numbers of environments and genotypes, environmental relationship patterns, field trial designs, and magnitudes of experimental error. The simulated data were used to evaluate 20 mixed models, ten of which included GRMs and ten which did not. The models included ten structures for environmental covariances including structures with no environmental correlation, structures with constant correlation among environments, and six FA structures. These models were compared to each other and to cell means and Additive Main effects and Multiplicative Interaction (AMMI) methods in terms of successful convergence and predictive accuracy. For most of the scenarios, models which included a GRM and a compound symmetric, constant variance structure produced the most accurate estimates. Models with GRM and FA structures were more accurate only when used to evaluate scenarios simulated with Toeplitz patterns of relationships and more than 25 genotypes or five environments. Unfortunately, the improved accuracy with the FA structures in these scenarios came at the cost of reduced convergence rates, so FA structures may not be reliable enough for some uses.

## 1. Introduction

Evaluations of genotypes in varied environmental conditions are referred to as multiple environment trials (MET), and are used in advanced stages of plant breeding programs to identify genotypes with superior performance across environments and within specific environments or sets of environments. Yield data from MET often show genotype by environment interactions (G×E), and often are analyzed using a two-way analysis of variance (ANOVA) model where genotype, environment, and their interaction are treated as fixed effects:

$$y_{ijk} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ijk}$$

where  $y_{ijk}$  is the yield (or other response variable) of the  $k^{\text{th}}$  replicate of the  $i^{\text{th}}$  genotype in the  $j^{\text{th}}$  environment,  $\mu$  is the overall mean,  $g_i$  is the fixed effect of the  $i^{\text{th}}$  genotype,  $e_j$  is the fixed effect of the  $j^{\text{th}}$  environment,  $(ge)_{ij}$  is the interaction between the  $i^{\text{th}}$  genotype and the  $j^{\text{th}}$  environment, and  $\varepsilon_{ijk}$  is the experimental error associated with the  $ijk^{\text{th}}$  observation;  $i = 1 \dots N_g$ ,  $j = \dots N_e$ ,  $k =$

1...N<sub>r</sub>. The estimates of genotype within environment effects are the means across replicates of each genotype in each environment (i.e. cell means). The major disadvantage of this approach is that these estimates are usually based on very little data (dependent on the number of replicates) and so are less predictively accurate than some alternative estimators. An additional disadvantage of this approach is that it cannot be used to estimate G×E effects when genotypes are not replicated within environments, since the effect of G×E and experimental error would be confounded. That is, experimental error cannot be separated from the specific effect of each genotype and environment combination.

Various estimators have been shown to be more accurate for MET than cell means. These include the Additive Main effects Multiplicative Interaction (AMMI) models (Gauch and Zobel, 1988; Gauch, 1988) and sites regression (SREG; Cornelius and Crossa, 1999) model families, which are sometimes referred to as linear-bilinear models. These two fixed-effect model families include sums of multiplicative terms, resulting from singular value decomposition, replacing  $(ge)_{ij}$ , in the case of AMMI, or  $g_i + (ge)_{ij}$  for SREG. The AMMI and SREG models have been shown to be relatively equivalent in terms of predictive accuracy (Cornelius and Crossa, 1999). Like the analysis of G×E in a fixed-effects ANOVA, the standard implementation of these models cannot be used when data from any genotype and environment combination is missing. However, the expectation-maximization algorithm has been used to impute missing data with the AMMI model (Gauch and Zobel, 1990).

Instead of treating genotype effects within environments as fixed, these effects can be considered random values leading to linear mixed models, which have important inherent benefits over fixed-effects models. Using mixed models, non-constant error variance structures, including within-field spatial correlation, can be easily incorporated in the same model as genotype and environment effects. Additionally, mixed models easily handle missing data and, with some specific models, even unreplicated data or prediction of genotype effects in environments they were not tested in.

Mixed model analyses have a long history in animal breeding (Henderson, 1973), and recent research has demonstrated new approaches to make them very effective in plant breeding. If a mixed linear model is used, genotypes are selected based on empirical best linear unbiased predictors (BLUPs) calculated using the estimated variance parameters. A very basic mixed model would assume a random effect of genotypes within environments that has a variance-covariance matrix of  $\sigma^2\mathbf{I}$ , where  $\sigma^2$  is a constant variance parameter and  $\mathbf{I}$  is an identity matrix. In most breeding programs, plant or animal, at least a portion of the genotypes assessed in a trial are related and therefore would be expected to show some correlation in their effects. Pedigree information can be used through a Genetic Relationship Matrix (GRM; or additive relationship matrix, or numerator relationship matrix) to take advantage of these relationships and improve predictive accuracy (Henderson, 1973). The GRM is usually symbolized as  $\mathbf{A}$ , and  $\mathbf{A} = 2[\mathbf{f}_{ii}]$ , where  $f_{ii}$  is the coefficient of parentage or coancestry between genotypes  $i$  and  $i'$  (Mrode and Thompson, 2005). When a GRM is used in a linear mixed model, genotype performance can be predicted in environments they were not replicated in. The GRM allows the model to use information from related genotypes to predict the unreplicated genotype, because known covariances are modeled between pairs of related genotypes.

Another modification that may improve the predictive accuracy of mixed models is to increase the complexity of the variance-covariance matrix of the random G×E effect beyond  $\sigma^2\mathbf{I}_g$  (Piepho, 1994). The matrix can be described as the product of two other matrices, such that  $\mathbf{G}_{ge} = \mathbf{G}_e \otimes \mathbf{I}_g$ , where  $\mathbf{I}_g$  is an identity matrix with dimensions equal to the number of genotypes, and

structures of varying complexity can be used to model  $G_e$  (Smith et al., 2001). One option for  $G_e$  is the factor analytic (FA) structure which increases in complexity with the number of factors used. When using a FA structure researchers must choose how many factors to include. More factors allow for greater flexibility, but may reduce model parsimony. Multiple researchers have demonstrated that a factor analytic structure can be combined with pedigree information to improve model fit, as measured by information criteria (Crossa et al., 2006; Oakey et al., 2007; Kelly et al., 2009; Beeck et al., 2010). These researchers have analyzed a limited number of real MET data sets; a simulation study could determine if the FA model with a GRM is the most effective model for a much wider range of MET.

The objective of this work was to simulate and analyze MET data sets to determine which linear models would be most effective in breeding programs by consistently providing the most accurate estimates and determine how the ideal model changes as a result of different MET conditions.

## 2. Methods

Simulations were conducted to generate data sets which resemble MET across a range of conditions. The simulations included randomly generated true effects of genotypes within environments and the phenotype of each observation, resulting from the addition of a random experimental error to the true genotype effect. These simulated data sets covered a range of scenarios with varying numbers of environments and genotypes, environmental relationship patterns, field trial designs, and magnitudes of experimental error.

True genotype effects within each environment were simulated as random samples from multivariate normal distributions with means of 0 and covariance matrices ( $\Sigma_{GE}$ ) that differed among scenarios. The  $\Sigma_{GE}$  were the Kronecker (or direct) product of a matrix of correlations between environments ( $\Sigma_E$ ) and a matrix of correlations between genotypes ( $\Sigma_G$ ). The  $\Sigma_E$  were of two sizes: 5 by 5 or 10 by 10, corresponding to scenarios with 5 or 10 environments, respectively. The degree of correlation in each  $\Sigma_E$  followed two patterns: compound symmetry, with constant correlations of 0.7, or Toeplitz structures with bands of constant correlation decreasing, from 0.9 to 0.3 or from 0.85 to -0.3, with distance from the diagonal for the 5 or 10 environment scenarios, respectively. These specific covariance values are by no means the only covariance values that could occur in a MET, but were chosen to be near values observed in MET. The  $\Sigma_E$  were generated with homogeneous or heterogeneous variances within environments. The variances ranged from 0.5 to 1.5 for the least and greatest variances, respectively. This three to one ratio is often used as a rule-of-thumb cutoff for considering variances to be heterogeneous, but greater and less heterogeneity are also possible. Two options for  $\Sigma_G$  were considered, corresponding to scenarios of 25 or 50 genotypes. A GRM was estimated from the pedigree in a Dow AgroSciences early generation study of North American Stiff Stalk maize inbred lines, and the two options for  $\Sigma_G$  were both submatrices of this GRM, specifically the submatrices corresponding to the first 25 and 50 genotypes in the study. For each combination of these options, i.e. each  $\Sigma_{GE}$ , true genotype effects were sampled 100 times.

For each set of genotype within environment effects, simulations were generated for three trial designs (RCBD – randomized complete block designs, MAD – modified augmented designs, and unreplicated designs) and three experimental error variances. Since spatial field effects were not considered, the only effect of the experimental design was to determine the number of replicates of each genotype in each environment. Therefore, other designs commonly

used in MET will still have equal or unequal replication, regardless of blocking structure, and so would not add much beyond the designs tested here. For the RCBD scenarios, every genotype was replicated three times. Whereas in the MAD scenarios, genotypes were not replicated except for primary and secondary “checks” that were replicated five and two times, respectively, for every 23 non-check genotypes. In the unreplicated design, each genotype appeared once in each environment. Every observation had a unique phenotype equal to the true effect of the genotype in an environment plus a random experimental error selected from a normal distribution with mean of 0 and one of three error variances ( $\sigma_e^2 = 0.1, 0.5, \text{ or } 2.0$ ). These error variances corresponded to repeatabilities (variance of the genotype within environment effects divided by the variance of the phenotypes for a given simulation) of about 0.9, 0.6, and 0.25, respectively.

A total of 20 related linear mixed models were compared for their ability to predict the simulated genotype effects within environments based on the simulated phenotypic data. Models were fit using the program ASReml-R, release 3.0 (Butler et al., 2009), which is a package for the R programming language (R Development Core Team, 2010). The models were all of the form:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}\mathbf{f} + \boldsymbol{\varepsilon} ,$$

where  $\mathbf{y}$  is the vector of observed phenotypes,  $\boldsymbol{\mu}$  is a vector of the grand mean,  $\mathbf{f}$  is the vector of genotype within environment effects, and  $\boldsymbol{\varepsilon}$  is the vector of experimental error terms. The joint distribution of  $\mathbf{f}$  and  $\boldsymbol{\varepsilon}$  is given by:

$$\begin{bmatrix} \mathbf{f} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim \text{MVN} \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma_R^2 \mathbf{I} \end{bmatrix} \right) ,$$

where  $\sigma_R^2$  is a constant error variance and  $\mathbf{G}$  is a covariance structure that varies for each of the 20 models and is separable:

$$\mathbf{G} = \mathbf{G}_E \otimes \mathbf{B},$$

where both  $\mathbf{G}_E$  and  $\mathbf{B}$  were varied, resulting in 20 models:

$$\mathbf{B} = \mathbf{I} \text{ or } \mathbf{A},$$

where  $\mathbf{I}$  is an identity matrix and  $\mathbf{A}$  is a GRM. This study evaluated the ideal situation, when the GRM used perfectly reflects the actual relationships among the genotypes; therefore,  $\mathbf{A}$  was set equal to the  $\Sigma_G$  used in the simulation of each scenario.

Ten structures were used to model  $\mathbf{G}_E$  and these are shown below for a five environment example. The simplest structure was independence (no covariance) and identical variances:

$$\mathbf{G}_E = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

A generalization of this is the diagonal structure, where environments are still independent, but each can have different variance:

$$\mathbf{G}_E = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix}$$

A constant covariance can be added, yielding compound symmetric structures with uniform or heterogeneous variances:

$$\mathbf{G}_E = \begin{bmatrix} \sigma_1 + \sigma^2 & \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 + \sigma^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 + \sigma^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 + \sigma^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 + \sigma^2 \end{bmatrix}$$

$$\mathbf{G}_E = \begin{bmatrix} \sigma_1 + \sigma_1^2 & \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 + \sigma_2^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 + \sigma_3^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 + \sigma_4^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 + \sigma_5^2 \end{bmatrix}$$

Six FA structures were compared. Structures were fit with one to three factors and uniform or heterogeneous specific variances:

$$\mathbf{G}_E = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \\ \lambda_{51} \end{bmatrix}, \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \end{bmatrix}, \text{ or } \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \\ \lambda_{41} & \lambda_{42} & \lambda_{43} \\ \lambda_{51} & \lambda_{52} & \lambda_{53} \end{bmatrix} \quad \mathbf{\Psi} = \psi(\mathbf{I}_5) \text{ or } \begin{bmatrix} \psi_1 & 0 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 & 0 \\ 0 & 0 & 0 & \psi_4 & 0 \\ 0 & 0 & 0 & 0 & \psi_5 \end{bmatrix}$$

In an effort to improve convergence rates, models were fit sequentially in the order of the  $\mathbf{G}_E$  structures described above. Parameter estimates from simpler models were used as the starting values of the next more complex structure for which the simple structure was a specific case. If a model did not converge, the next more complex structure was not attempted. The percentage of simulations of a scenario for which a model converged was defined as the convergence rate.

In addition to the mixed linear models, estimates of genotype effects within environments were derived from cell means (the mean of the replicates of a genotype in each environment) and Additive Main effects Multiplicative Interaction (AMMI) models. The AMMI models are fixed effects linear models with main effects for genotype and environment. The effects of genotype by environment interaction have been replaced with an approximation of the matrix using a reduced set of the principle components (Gauch, 1988). Only RCBD scenarios were analyzed using the AMMI models. The AMMI models were fit with all possible numbers of principle components. The most accurate, as judged by the correlations described below, of the AMMI models for each simulation were compared to the mixed linear models.

To evaluate each model's predictive accuracy, Pearson correlations between the estimated effects of genotypes within environments and the true effects generated in the simulation were calculated for each simulation. The mean of these correlations across the 100 simulations for each scenario allows conclusions to be drawn to identify the best model for each scenario. While not formally evaluated, mean correlations appeared to stabilize after the first 50 simulations.

### 3. Results and Discussion

Due to the large number of scenarios evaluated, it is ineffectual to present all convergence rates and prediction accuracy estimates. Instead, salient data are presented to provide evidence for conclusions. Complete data are available from the author upon request. Not all models converged for all data sets, and convergence rates generally decreased as model complexity or MET size increased. Accuracy results are not reported for models that could not be used for certain scenarios or converged for less than 10 percent of the simulations.

As a baseline scenario, consider the MET scenario with 5 environments with a compound symmetric  $\Sigma_E$  that has homogeneous within environment variances, 25 genotypes, a randomized complete block design, and an error variance of 2.0 (repeatability  $\approx .25$ ; Table 1). The mean (over all 100 simulations) correlation of the cell means with the true genotype within environment effects was a modest 0.69. On average, the best AMMI model improved predictive accuracy (0.78). Only the best of the mixed models matched or exceeded this value. This might lead one to conclude that the AMMI is nearly the best method for the analysis of MET. However, it is important to note that the most accurate AMMI model was chosen for each simulation and in actual practice this is not always possible, since it is not known and must be guessed by various methods. Additionally, a major disadvantage of the standard AMMI model compared here is that cannot be used for experimental designs where some genotypes are not replicated.

Accuracy of the mixed linear models was usually greater than the cell means. The mean accuracy for the mixed model without a GRM and with an identical structure for  $\mathbf{G}_E$  was equal to that of the cell means. The equal accuracy, as measured by Pearson correlations, occurs because the empirical BLUPs of the mixed model differ from the cell means only by a constant shrinkage factor. The shrinkage factor varies over genotypes within environments for all of the other mixed models considered. When variances were allowed to vary across environments (diagonal structure), mean predictive accuracy decreased, suggesting that the additional parameters only served to model noise in the phenotypes. Inclusion of a constant correlation term by using the compound symmetric structures improved accuracy of the predictions, closer to that of the AMMI model. This improvement was expected, since the data were simulated with a constant correlation of 0.7 across environments. Again, allowing for heterogeneous variances did not improve predictive accuracy. The use of the more complex FA structures provided no additional improvement in accuracy over the compound symmetric structure. Allowing for heterogeneous specific variances did not improve predictive accuracy. Adding additional factors beyond the first did not noticeably change the mean accuracy. The lack of superiority of the FA structures is not surprising, since this scenario was simulated with a constant correlation among environments, just like the compound symmetric structure used for  $\mathbf{G}_E$  in the analysis. The use of a GRM when modeling simulations of this scenario consistently improved predictive accuracy across, no matter the structure used for  $\mathbf{G}_E$ . The benefit was greatest when the compound symmetric structure with constant variance was used with it. This combination resulted in a

predictive accuracy greater than the best AMMI model. The changes in accuracy caused by different  $G_E$  were similar whether or not a GRM was included. Therefore, the best model included a GRM and a compound symmetric, uniform within environment covariance structure (GRM CSU). The preceding results applied to the baseline scenario, and in the following we will describe how these results change in other scenarios and if the GRM CSU model is always the most accurate.

As error variance decreased from 2.0 to 0.5 to 0.1 (repeatability  $\approx$  0.25, 0.6, and .9, respectively), the accuracy of all models increased dramatically, but unevenly (Table 2). The benefits of using more complex models for  $G_E$  or using a GRM decreased with decreasing error variance. At the lowest level of experimental error, using the best mixed linear model provided estimates of effectively no greater accuracy than the cell means. These results were nearly uniformly consistent across all other scenario variations. The effects of other scenario variations were generally consistent across error variances and were most dramatic when  $\sigma_e^2 = 2.0$ . Therefore, only results from scenarios with error variances of 2.0 will be presented for clarity. Spatial field effects are often a major concern when evaluating MET (Smith et al., 2001). However we did not simulate them here to focus on determining the most appropriate models for random effects of genotypes within environments. If real MET data has substantial spatial effects that have not been accounted for, estimates may be biased or experimental error may be increased. Even if spatial effects are accounted for in the model, the accuracy of the model is likely to decrease due to errors in estimating the spatial effects. The magnitude of this decrease in accuracy will need to be determined in future studies.

Experimental design had a dramatic effect on model accuracies and altered the effects of other scenario variations (Table 3). Not all models could be used with all experimental designs. As described above, the traditional AMMI model can only be applied to estimate genotypes that are replicated in each environment, limiting it to RCBD. All other models are suitable for RCBD and MAD, but only models that incorporate a GRM can be used with unreplicated designs. With all other variables held constant, estimates from RCBD scenarios were consistently and substantially more accurate than those from MAD scenarios, which were themselves slightly more accurate than those from unreplicated scenarios. The accuracy of each model varies across the designs, but the ranking of the models is generally consistent. Of note is that the GRM CSU model is consistently the most accurate model, but the benefit of using it instead of the cell means model is greater for the MAD and unreplicated designs. Effects of other scenario variations were not very consistent across the experimental designs and MET experimental design flexibility is often limited, so further results will be presented for all three. The lower accuracies for MAD and unreplicated designs are likely simply due to the reduced number of observations for most genotype by environment combinations. This increases the value of the way the GRM CSU model uses information from related genotypes and environments to adjust the cell means.

When a scenario with 50 genotypes was compared to the baseline scenario with 25, predictive accuracy was found to increase for all models (Table 4). This was generally a minor improvement, but was greater for many of the models with heterogeneous variances or FA structures. The increased number of genotypes brought these models closer to parity with the CSU models. It is well known that variance parameters are better estimated when the number of treatment levels is high. This is because each level is an observation from an unknown distribution of effects, and so more levels increases the sample size used to estimate the parameters of the distribution. Here the greater number of genotypes provides more information



for the estimation of parameters in  $\mathbf{G}_E$ , increasing accuracy for the more complex models. The lesser improvement for RCBD scenarios suggests diminishing accuracy returns when each genotype is replicated. That is, both increasing numbers of treatments or increasing replication improve accuracy, but not in an additive fashion.

Increasing the number of environments in the MET again increased the accuracy of almost all models, but only very slightly for models that did not include correlations in  $\mathbf{G}_E$  (Table 5). The benefits of using a structure that allows correlation between environments became more dramatic with the greater number of environments, but the difference between the CSU and one factor models decreased. Despite this, the GRM CSU model was the most accurate choice for either number of environments.

To this point, all the scenarios considered have been simulated using a compound symmetric constant variance pattern for  $\Sigma_E$ , so it is entirely expected that models which include the compound symmetric constant variance structure for  $\mathbf{G}_E$  would be the most accurate. These scenarios were included to reflect the situations that can occur in real MET when environments are similarly related. What is more interesting is whether this changes for scenarios where the MET were simulated with more complex patterns of relationships.

In real MET data sets, environments often exhibit differing amounts of variability in genotype performance; that is, in certain environments the differences among genotypes are more dramatic than in other environments. To simulate such scenarios,  $\Sigma_E$  were chosen that had a 3:1 ratio between the most varied and least varied environments. When compared to the scenarios with constant variances, accuracies from a scenario with heterogeneous variances were very similar (Table 6). Models including heterogeneous variances or FA structures were only slightly more competitive with the more parsimonious models. Most importantly, the GRM CSU model was still the most accurate, if only just slightly for RCBD and MAD scenarios. We see the same trends when comparing constant and heterogeneous variance scenarios with 50 genotypes or 10 environments (data not shown). These results suggest that even when environments differ in their variance (at least at a 3:1 ratio), it is more important for a model to be parsimonious rather than flexible in this way. This can be explained in that a more flexible model requires more parameters to be estimated with the same amount of information, reducing the information per parameter and reducing the accuracy of the estimation of each parameter. Here we observed that this reduction in accuracy of estimating individual parameters eliminated any gain in overall accuracy provided by better matching the heterogeneous nature of the data variance.

Accuracy means for each model were very similar between the 25 genotype, five environment, compound symmetric scenarios and matching scenarios simulated with a Toeplitz pattern (Table 7). Factor analytic structures were more competitive when used to analyze Toeplitz scenarios, but the GRM CSU model was still the most accurate. This suggests that there may have been insufficient information to accurately estimate the increased number of parameters included by the FA structures. That is, the well estimated single correlation parameter more accurately reflected the different correlations among pairs of environments than did the poorly estimated factors. When assessing the effect of relationship pattern for scenarios with 50 genotypes or 10 environments, FA structures were even more competitive (Tables 8 & 9). For some Toeplitz scenarios, FA models had superior accuracy compared to the GRM CSU model. This may be due to the wider range of simulated covariance values (0.85 to -0.3 versus 0.9 to 0.3), in the case of 10 environments. When more genotypes were simulated these provided more information to better estimate the additional parameters in the FA structures.

Interestingly, the use of a GRM was less beneficial for these scenarios. It is important to note that the GRM FA models converged for fewer simulations than the GRM CSU models.

These results suggest that when estimating the effects of genotypes within environments from MET, using a GRM produces more accurate estimates, as does using a CSU or FA structure, depending on the pattern of relationships among environments. This study assumed the ideal situation, when the GRM matrix used perfectly matches the true genetic relationships. In reality the GRM matrix is usually estimated based on pedigree or marker data and so will be subject to error. Such errors would reduce the predictive accuracy of models that incorporate a GRM, possibly to a degree that would eliminate the advantage observed in these simulations. Further study is necessary to determine the likelihood of such a situation. Other authors have demonstrated superior model fit, as measured by information criteria, using GRM with FA structures rather than compound symmetric structures (Crossa et al., 2006; Oakey et al., 2007). This suggests that these authors analyzed data sets with complex relationship patterns similar to the Toeplitz simulations in this study. In real MET, very simple relationship patterns, i.e. near compound symmetry, are unlikely, so the results for the Toeplitz pattern simulations may be more applicable. In the Toeplitz simulations, FA structures were only more accurate than the CSU structure for some of the scenarios with more than 25 genotypes or 5 environments. Even when the FA structures were more accurate, they were only a small improvement over the CSU structure. At this point it becomes important to note the differing convergence rates between models with these two structures. For the scenarios simulated with 10 environments in a Toeplitz pattern, 25 genotypes, and an error variance of 2.0, convergence rates for models using FA structures were much lower than when a CSU structure was included (Table 10). Model convergence might be improved through trial and error techniques, such as introducing different starting values. In real world high throughput applications, the ease of fitting a model can be much more important than small differences in accuracy, so a model which almost always converges and has competitive accuracy, i.e. the GRM CSU model, may be the best choice.

## 4. Summary

This study used simulations to help determine how various mixed models perform when used to analyze a range of MET. For most of the scenarios, models which included a GRM and a compound symmetric, constant variance structure produced the most accurate estimates. Models with GRM and FA structures were more accurate only when used to evaluate scenarios simulated with Toeplitz patterns of relationships and more than 25 genotypes or five environments. The choice of compound symmetric versus FA structures may depend on whether accuracy or reliability is more important. Analyses of additional simulations with more genotypes and/or more environments could determine if these conclusions are consistent for larger MET.

Table 1. Baseline Scenario. Pearson correlations between estimates/predictions of effects of genotypes within environments and the true simulated effects of genotypes within environments for a baseline scenario of 25 genotypes, a randomized complete block design, an error variance of 2.0, and 5 environments simulated with a compound symmetric relationship and a common within environment variance.

Cell Means	AMMI	ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U
0.69	0.78	0.69	0.66	0.74	0.73	0.73	0.72	0.73	0.72	0.73
GRM ID	GRM Diag	GRM CSU	GRM CSH	GRM FA1U	GRM FA1H	GRM FA2U	GRM FA2H	GRM FA3U		
0.72	0.69	0.8	0.78	0.77	0.77	0.74	0.75	0.77		

Table 2. Effect of Error Variance. Pearson correlations between estimates/predictions of effects of genotypes within environments and the true simulated effects of genotypes within environments for scenarios with three levels of error variance and 25 genotypes, a randomized complete block design, and 5 environments simulated with a compound symmetric relationship and a common within environment variance.

Design	ErrorVar	Cell										
		Means	AMMI	ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U
RCBD	0.1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
RCBD	0.5	0.89	0.89	0.89	0.89	0.9	0.9	0.9	0.9	0.9	0.9	0.9
RCBD	2	0.69	0.78	0.69	0.66	0.74	0.73	0.73	0.72	0.73	0.72	0.73
Design	ErrorVar	GRM ID	GRM Diag	GRM CSU	GRM CSH	GRM FA1U	GRM FA1H	GRM FA2U	GRM FA2H	GRM FA3U		
RCBD	0.1	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.97		
RCBD	0.5	0.89	0.89	0.92	0.91	0.91	0.91	0.91	0.91	0.9		
RCBD	2	0.72	0.69	0.8	0.78	0.77	0.77	0.74	0.75	0.77		

Table 3. Effect of Experimental Design. Pearson correlations between estimates/predictions of effects of genotypes within environments and the true simulated effects of genotypes within environments for scenarios with three experimental designs and 25 genotypes, an error variance of 2.0, and 5 environments simulated with a compound symmetric relationship and a common within environment variance.

Design	Cell Means	AMMI	ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U
RCBD	0.69	0.78	0.69	0.66	0.74	0.73	0.73	0.72	0.73	0.72	0.73
MAD	0.5		0.5	0.44	0.57	0.51	0.52	0.54	0.54	0.54	0.54
Unrep	0.48										

  

Design	GRM ID	GRM Diag	GRM CSU	GRM CSH	GRM FA1U	GRM FA1H	GRM FA2U	GRM FA2H	GRM FA3U
RCBD	0.72	0.69	0.8	0.78	0.77	0.77	0.74	0.75	0.77
MAD	0.51	0.47	0.66	0.51	0.6	0.61	0.59	0.6	0.62
Unrep	0.48	0.43	0.64	0.49	0.59	0.58	0.57	0.57	0.56

Table 4. Effect of genotype quantity. Pearson correlations between estimates/predictions of effects of genotypes within environments and the true simulated effects of genotypes within environments for scenarios with 25 or 50 genotypes, three experimental designs, an error variance of 2.0, and 5 environments simulated with a compound symmetric relationship and a common within environment variance.

# of Gen	Design	Cell Means	AMMI	ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U
25	RCBD	0.69	0.78	0.69	0.66	0.74	0.73	0.73	0.72	0.73	0.72	0.73
25	MAD	0.5		0.5	0.44	0.57	0.51	0.52	0.54	0.54	0.54	0.54
25	Unrep	0.48										
50	RCBD	0.71	0.79	0.71	0.69	0.78	0.77	0.77	0.77	0.77	0.77	0.76
50	MAD	0.5		0.51	0.45	0.61	0.58	0.58	0.58	0.58	0.58	0.57
50	Unrep	0.5										

  

# of Gen	Design	GRM ID	GRM Diag	GRM CSU	GRM CSH	GRM FA1U	GRM FA1H	GRM FA2U	GRM FA2H	GRM FA3U
25	RCBD	0.72	0.69	0.8	0.78	0.77	0.77	0.74	0.75	0.77
25	MAD	0.51	0.47	0.66	0.51	0.6	0.61	0.59	0.6	0.62
25	Unrep	0.48	0.43	0.64	0.49	0.59	0.58	0.57	0.57	0.56
50	RCBD	0.74	0.73	0.82	0.81	0.81	0.8	0.79	0.8	0.8
50	MAD	0.55	0.51	0.7	0.66	0.65	0.65	0.64	0.64	0.65
50	Unrep	0.5	0.46	0.68	0.58	0.63	0.64	0.63	0.64	0.62

Table 5. Effect of environment quantity. Pearson correlations between estimates/predictions of effects of genotypes within environments and the true simulated effects of genotypes within environments for scenarios with three experimental designs, 25 genotypes, an error variance of 2.0, and 5 or 10 environments simulated with a compound symmetric relationship and a common within environment variance.

# of Env	Design	Cell Means	AMMI	ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U
5	RCBD	0.69	0.78	0.69	0.66	0.74	0.73	0.73	0.72	0.73	0.72	0.73
5	MAD	0.5		0.5	0.44	0.57	0.51	0.52	0.54	0.54	0.54	0.54
5	Unrep	0.48										
10	RCBD	0.7	0.82	0.7	0.68	0.79	0.78	0.79	0.78	0.79	0.78	
10	MAD	0.51		0.52	0.44	0.67	0.61	0.65	0.61		0.56	
10	Unrep	0.5										

  

# of Env	Design	GRM ID	GRM Diag	GRM CSU	GRM CSH	GRM FA1U	GRM FA1H	GRM FA2U	GRM FA2H	GRM FA3U
5	RCBD	0.72	0.69	0.8	0.78	0.77	0.77	0.74	0.75	0.77
5	MAD	0.51	0.47	0.66	0.51	0.6	0.61	0.59	0.6	0.62
5	Unrep	0.48	0.43	0.64	0.49	0.59	0.58	0.57	0.57	0.56
10	RCBD	0.73	0.7	0.84	0.82	0.82	0.81			
10	MAD	0.52	0.47	0.74	0.65	0.7	0.69			
10	Unrep	0.48	0.44	0.72	0.59	0.7	0.68			

Table 6. Effect of heterogeneous environmental variance. Pearson correlations between estimates/predictions of effects of genotypes within environments and the true simulated effects of genotypes within environments for scenarios with three experimental designs, 25 genotypes, an error variance of 2.0, and 5 environments simulated with a compound symmetric relationship and constant or heterogeneous within environment variances.

Constant Variance	Design	Cell Means	AMMI	ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U
Yes	RCBD	0.69	0.78	0.69	0.66	0.74	0.73	0.73	0.72	0.73	0.72	0.73
Yes	MAD	0.5		0.5	0.44	0.57	0.51	0.52	0.54	0.54	0.54	0.54
Yes	Unrep	0.48										
No	RCBD	0.7	0.77	0.7	0.68	0.75	0.75	0.75	0.75	0.75	0.75	0.74
No	MAD	0.5		0.51	0.46	0.56	0.53	0.54	0.54	0.56	0.55	0.58
No	Unrep	0.48										

  

Constant Variance	Design	GRM ID	GRM Diag	GRM CSU	GRM CSH	GRM FA1U	GRM FA1H	GRM FA2U	GRM FA2H	GRM FA3U
Yes	RCBD	0.72	0.69	0.8	0.78	0.77	0.77	0.74	0.75	0.77
Yes	MAD	0.51	0.47	0.66	0.51	0.6	0.61	0.59	0.6	0.62
Yes	Unrep	0.48	0.43	0.64	0.49	0.59	0.58	0.57	0.57	0.56
No	RCBD	0.73	0.71	0.8	0.79	0.78	0.79	0.78	0.78	0.79
No	MAD	0.51	0.49	0.65	0.57	0.64	0.62	0.64	0.62	0.63
No	Unrep	0.47	0.44	0.64	0.5	0.58	0.57	0.55	0.52	0.52

Table 7. Effect of relationship pattern. Pearson correlations between estimates/predictions of effects of genotypes within environments and the true simulated effects of genotypes within environments for scenarios with three experimental designs, 25 genotypes, an error variance of 2.0, and 5 environments simulated with a compound symmetric relationship or a Toeplitz pattern and constant within environment variances.

Pattern	Design	Cell										
		Means	AMMI	ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U
CS	RCBD	0.69	0.78	0.69	0.66	0.74	0.73	0.73	0.72	0.73	0.72	0.73
CS	MAD	0.5		0.5	0.44	0.57	0.51	0.52	0.54	0.54	0.54	0.54
CS	Unrep	0.48										
Toep	RCBD	0.69	0.79	0.69	0.66	0.76	0.75	0.75	0.75	0.76	0.76	0.73
Toep	MAD	0.5		0.5	0.44	0.57	0.54	0.55	0.55	0.57	0.57	0.55
Toep	Unrep	0.51										

  

Pattern	Design	GRM										
		ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U		
CS	RCBD	0.72	0.69	0.8	0.78	0.77	0.77	0.74	0.75	0.77		
CS	MAD	0.51	0.47	0.66	0.51	0.6	0.61	0.59	0.6	0.62		
CS	Unrep	0.48	0.43	0.64	0.49	0.59	0.58	0.57	0.57	0.56		
Toep	RCBD	0.71	0.68	0.8	0.78	0.77	0.77	0.78	0.78	0.76		
Toep	MAD	0.49	0.45	0.65	0.55	0.61	0.6	0.6	0.59	0.58		
Toep	Unrep	0.45	0.43	0.66	0.53	0.61	0.62	0.62	0.62	0.62		



Table 8. Effect of relationship pattern: 10 environments. Pearson correlations between estimates/predictions of effects of genotypes within environments and the true simulated effects of genotypes within environments for scenarios with three experimental designs, 25 genotypes, an error variance of 2.0, and 10 environments simulated with a compound symmetric relationship or a Toeplitz pattern and constant within environment variances.

Pattern	Design	Cell Means	AMMI	ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U
CS	RCBD	0.7	0.82	0.7	0.68	0.79	0.78	0.79	0.78	0.79	0.78	
CS	MAD	0.51		0.52	0.44	0.67	0.61	0.65	0.61		0.56	
CS	Unrep	0.5										
Toep	RCBD	0.73	0.81	0.73	0.71	0.75	0.75	0.79	0.78	0.82	0.82	
Toep	MAD	0.53		0.54	0.48	0.56	0.57	0.65	0.64	0.67	0.67	
Toep	Unrep	0.52										

  

Pattern	Design	GRM ID	GRM Diag	GRM CSU	GRM CSH	GRM FA1U	GRM FA1H	GRM FA2U	GRM FA2H	GRM FA3U
CS	RCBD	0.73	0.7	0.84	0.82	0.82	0.81			
CS	MAD	0.52	0.47	0.74	0.65	0.7	0.69			
CS	Unrep	0.48	0.44	0.72	0.59	0.7	0.68			
Toep	RCBD	0.77	0.74	0.8	0.78	0.79	0.78	0.83	0.82	
Toep	MAD	0.59	0.55	0.67	0.66	0.69	0.69			
Toep	Unrep	0.56	0.52	0.65	0.58	0.65	0.63			

Table 9. Effect of relationship pattern: 50 genotypes. Pearson correlations between estimates/predictions of effects of genotypes within environments and the true simulated effects of genotypes within environments for scenarios with three experimental designs, 50 genotypes, an error variance of 2.0, and 5 environments simulated with a compound symmetric relationship or a Toeplitz pattern and constant within environment variances.

# of Gen	Design	Cell Means	AMMI	ID	Diag	CSU	CSH	FA1U	FA1H	FA2U	FA2H	FA3U
CS	RCBD	0.71	0.79	0.71	0.69	0.78	0.77	0.77	0.77	0.77	0.77	0.76
CS	MAD	0.5		0.51	0.45	0.61	0.58	0.58	0.58	0.58	0.58	0.57
CS	Unrep	0.5										
Toep	RCBD	0.71	0.81	0.71	0.69	0.79	0.78	0.78	0.79	0.81	0.81	0.8
Toep	MAD	0.51		0.51	0.45	0.63	0.59	0.59	0.6	0.63	0.62	0.59
Toep	Unrep	0.49										

  

# of Gen	Design	GRM ID	GRM Diag	GRM CSU	GRM CSH	GRM FA1U	GRM FA1H	GRM FA2U	GRM FA2H	GRM FA3U
CS	RCBD	0.74	0.73	0.82	0.81	0.81	0.8	0.79	0.8	0.8
CS	MAD	0.55	0.51	0.7	0.66	0.65	0.65	0.64	0.64	
CS	Unrep	0.5	0.46	0.68	0.58	0.63	0.64	0.63	0.64	0.62
Toep	RCBD	0.74	0.73	0.82	0.81	0.81	0.82	0.84	0.84	0.82
Toep	MAD	0.53	0.5	0.7	0.66	0.66	0.67	0.67	0.66	0.65
Toep	Unrep	0.47	0.46	0.67	0.59	0.65	0.64	0.65	0.65	0.64

Table 10. Number of simulations for which each model converged, out of 100, for scenarios with three experimental designs, 25 genotypes, an error variance of 2.0, and 10 environments simulated with a Toeplitz pattern and constant within environment variances.

Design	ID	Diag	CorV	CorH	FA1U	FA1H	FA2U	FA2H	FA3U	FA3H
RCBD	100	100	100	93	61	79	33	46	6	7
MAD	100	100	100	54	28	46	8	16	1	5

  

Design	GRM ID	GRM Diag	GRM CorV	GRM CorH	GRM FA1U	GRM FA1H	GRM FA2U	GRM FA2H	GRM FA3U	GRM FA3H
RCBD	100	100	100	84	41	69	16	33	1	5
MAD	100	100	100	34	23	36	0	8	0	0
Unrep	100	100	99	32	15	34	1	8	0	4

## 5. Acknowledgements

We would like to thank the members of the Dow AgroSciences Quantitative Genetics group for their assistance, especially Kelly Robins who provided some initial programs and data. We also would like to thank the anonymous reviewers for their suggestions. Dow AgroSciences, the Washington Grain Commission, and USDA project 5348-21000-023-00 helped fund this research.

## 6. References

- Beeck, C.P., W.A. Cowling, A.B. Smith, and B.R. Cullis. 2010. Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. *Genome*. 53: 992-1001.
- Butler, D.G., B.R. Cullis, A.R. Gilmour, and B.J. Gogel. 2009. Mixed models for S language environments ASReml-R reference manual.
- Cornelius, P.L., and J. Crossa. 1999. Prediction assessment of shrinkage estimators of multiplicative models for multi-environment cultivar trials. *Crop Science*. 39(4): 998-1009.
- Crossa, J., J. Burgueño, P.L. Cornelius, G. McLaren, R. Trethowan, and A. Krishnamachari. 2006. Modeling genotype  $\times$  environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. *Crop Science*. 46(4): 1722-1733.
- Gauch, H.G. 1988. Model selection and validation for yield trials with interaction. *Biometrics*. 44(3): 705-715.
- Gauch, H.G., and R.W. Zobel. 1988. Predictive and postdictive success of statistical analyses of yield trials. *Theoret. Appl. Genetics*. 76(1): 1-10.
- Gauch, H.G., and R.W. Zobel. 1990. Imputing missing yield trial data. *Theoret. Appl. Genetics*. 79(6): 753-761.
- Henderson, C.R. 1973. Sire evaluation and genetic trends. *J. Anim Sci.* 1973(Symposium): 10-41.
- Kelly, A., B.R. Cullis, A.R. Gilmour, J.A. Eccleston, and R. Thompson. 2009. Estimation in a multiplicative mixed model involving a genetic relationship matrix. *Genetics Selection Evolution*. 41: 33-42.
- Mrode, R.A., and R. Thompson. 2005. Linear models for the prediction of animal breeding values. 2nd ed. CABI, Cambridge, MA.
- Oakey, H., A.P. Verbyla, B.R. Cullis, X. Wei, and W.S. Pitchford. 2007. Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoretical and Applied Genetics*. 114: 1319-1332.

Piepho, H.-P. 1994. Best Linear Unbiased Prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. *Theoret. Appl. Genetics*. 89(5).

R Development Core Team. 2010. R: A language and environment for statistical computing. Available at <http://www.R-project.org>.

Smith, A., B. Cullis, and R. Thompson. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*. 57(4): 1138-1147.