CENTENARY REVIEW

The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches

A. B. SMITH[1]*, B. R. CULLIS[1] AND R. THOMPSON[2]

[1] *Wagga Wagga Agricultural Institute, Private Mail Bag, Wagga Wagga, NSW 2650, Australia*
[2] *Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

(*Revised MS received 23 June 2005*)

SUMMARY

The analysis of series of crop variety trials has a long history with the earliest approaches being based on ANOVA methods. Kempton (1984) discussed the inadequacies of this approach, summarized the alternatives available at that time and noted that all of these approaches could be classified as multiplicative models. Recently, mixed model approaches have become popular for the analysis of series of variety trials. There are numerous reasons for their use, including the ease with which incomplete data (not all varieties in all trials) can be handled and the ability to appropriately model within-trial error variation. Currently, the most common mixed model approaches for series of variety trials are mixed model versions of the methods summarized by Kempton (1984). In the present paper a general formulation that encompasses all of these methods is described, then individual methods are considered in detail.

INTRODUCTION

The breeding and evaluation of improved crop varieties is one of the oldest agricultural research pursuits and has had a major impact on world food production, particularly in the last 100 years. As Nobel peace prize laureate Dr Norman Borlaug stated 'During the twentieth century, conventional breeding has produced – and continues to produce – a vast number of varieties and hybrids that have contributed immensely to much higher grain yields, stability of harvests and farm incomes, while also sparing vast tracts of land for nature (wildlife habitats, forests, outdoor recreation)' (speech given at Tuskegee University, April 2001). The challenge now is to maintain the improvement of crops at a rate that will meet a rapidly increasing world population (projected to be 8·3 billion people in 2025). Borlaug makes the important point that in order to achieve this, both conventional breeding and biotechnology methodologies will be needed. He suggests that 'While biotechnology research tools offer much promise, it is also important to recognize that conventional plant

breeding methods are continuing to make significant contributions to improved food production and enhanced nutrition'. It is, therefore, vital that the statistical methods used to design and analyse data from crop cultivar breeding and evaluation programmes are as accurate, efficient and informative as possible. In this paper current analytical methods are reviewed with this objective in mind. The focus is on the analysis of grain yield data although many of the concepts apply to other traits such as quality traits related to end-use product manufacture (for example, bread making).

Typically, the yield data generated from crop breeding and evaluation programmes arises from series of field trials known as multi-environment trials (MET). These trials allow the investigation of varietal yield performance across a range of geographic locations and, possibly, years (seasons). The development of statistical methods for the analysis of MET data has a long history. Early methods were focused on Analysis of Variance (ANOVA) techniques that partitioned total variation into sources due to varieties, environments (location/year combinations), variety by environment (V × E) interaction and within-trial error variation. As Kempton (1984) pointed out, a major drawback with this approach

* To whom all correspondence should be addressed.
Email: alison.smith@dpi.nsw.gov.au

is that it provides no insight into the nature of $V \times E$ interaction. This may hinder varietal selection and recommendation decisions. For example, it may be insufficient merely to obtain an estimate of overall (average) variety performance across environments. Additionally, a measure of varietal stability to environmental change may be required, either for the identification of varieties that are both high yielding and stable (so are suitable for broad usage) or varieties that perform exceptionally well under certain conditions (so may be suitable for use in specific environments).

As a result, a range of more complex and informative models has been proposed for the analysis of MET data. This paper does not attempt to review all methods but defers to Kempton (1984), who summarized the models succinctly by stating that 'methods for describing patterns in $V \times E$ are the method of principal components, and of regression onto an independent environmental variable or onto the marginal means of the $V \times E$ table'. He continued by noting that all of these methods '... describe crop response as a series of multiplicative terms, each term being a product of a variety and an environment effect' so can all be classified under the heading of multiplicative models. The models discussed by Kempton (1984) involve fixed effects only (with the exception of residual error). Recently, the advantages afforded by linear mixed models compared with ordinary linear models have been recognized for the purpose of analysing MET data. These advantages include the ease with which incomplete data (not all varieties in all environments) can be handled, the ability to use more realistic within-trial models for error variation (e.g. incomplete blocks, spatial correlation models) and the ability to assume some sets of effects (e.g. variety and/or environment effects) to be random rather than fixed. Thus, linear mixed models have become popular for the analysis of MET data. They range from simple variance component models that provide information similar to ANOVA through to multiplicative mixed models that aim to explore or better accommodate $V \times E$ interaction. The most commonly used models are essentially mixed model analogues of the afore-mentioned methods described by Kempton (1984). These models provide the focus of the current paper, which is arranged as follows. The next section describes a general formulation for the mixed model analysis of a series of variety trials. Four of the most popular mixed model approaches are considered in detail, namely variance component models, regressions on environmental variables, regressions on environmental means and multiplicative models. Models for within-trial error variation are also discussed. The section 'Estimation and software' provides a general description of estimation procedures for these models. Finally, some concluding remarks are made.

LINEAR MIXED MODEL FOR MET DATA

MET data may be summarized as a two-way table indexed by varieties and environments. The extension to higher order tables (for example, when environments comprise the factorial combination of geographic locations and years) will be discussed later. Consider a series of t trials (synonymous with environments) in which a total of m varieties has been grown (without necessarily all varieties grown in all trials). A 'base-line' model for the yield of the kth replicate of variety i in trial j can be written as

$$y_{ijk} = \eta_{ij} + \varepsilon_{ijk} \tag{1}$$

where η_{ij} is the effect of variety i in environment j and ε_{ijk} is the residual effect for replicate k of this variety by environment combination. In the current paper, statistical models are represented using vector notation so the model in Eqn (1) is re-written as

$$y = M\eta + \varepsilon \tag{2}$$

where y is the $n \times 1$ vector of individual plot yields combined across trials (ordered as plots within trials) and $n = \sum_{j=1}^{t} n_j$ where n_j is the number of plots in the jth trial. The vector $\eta = (\eta_{11}, \eta_{21} \ldots \eta_{m1} \ldots \eta_{1t} \ldots \eta_{mt})'$ is the $mt \times 1$ vector of variety by environment effects (ordered as varieties within environments) and M is an $n \times mt$ design (replication) matrix that assigns variety by environment combinations to the vector of yields. Note that M will contain columns whose elements are all zero if not all varieties appeared in all trials. The $n \times 1$ vector ε is the combined vector of residual effects from all trials (ordered as for the vector of yields).

The most commonly used mixed models for MET data can be formed from the base-line model in Eqn (2) by specifying two submodels, namely a model for the 'table of' $V \times E$ effects (η) and a model for the residual effects (ε). These are described in the following sections.

Models for $V \times E$ effects

As previously noted, current mixed model approaches for the analysis of MET data can be viewed as mixed model analogues of the approaches discussed by Kempton (1984). Thus, each of these models is considered in turn.

Variance component mixed models

In the classical ANOVA approach for the analysis of MET data the model for the effect of variety i in environment j is given by

$$\eta_{ij} = \mu + \alpha_i + \theta_j + \delta_{ij} \tag{3}$$

where μ is an overall mean effect, α_i is the main effect for variety i, θ_j is the main effect for environment j

and δ_{ij} is the interaction effect for variety i and environment j. The model in Eqn (3) can be written in vector notation by considering the complete set of variety by environment effects, that is, the $mt \times 1$ vector $\boldsymbol{\eta}$. Thus

$$\boldsymbol{\eta} = \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha} + (\boldsymbol{I}_t \otimes \mathbf{1}_m)\boldsymbol{\theta} + \boldsymbol{\delta} \qquad (4)$$

where $\boldsymbol{\alpha} = (\alpha_1 \ldots \alpha_m)'$ is the $m \times 1$ vector of variety main effects, $\boldsymbol{\theta} = (\theta_1 \ldots \theta_t)'$ is the $t \times 1$ vector of environment main effects and $\boldsymbol{\delta}$ is the $mt \times 1$ vector of $V \times E$ interaction effects (ordered as for $\boldsymbol{\eta}$). Standard notation for unit vectors and identity matrices is used so that, for example, the vector $\mathbf{1}_t$ denotes the unit vector of length t and the matrix $\boldsymbol{I}_t$ denotes the $t \times t$ identity matrix. The symbol $\otimes$ represents the Kronecker product of two matrices or vectors. A definition of Kronecker products is given in the Appendix, together with a small example to show how they are used to expand Eqn (4).

Historically, all effects in this model were regarded as fixed, but more recently mixed model versions of Eqn (4) have been used (Patterson *et al.* 1977; Patterson & Silvey 1980; Talbot 1984; Patterson & Nabugoomu 1992; Cullis *et al.* 1996*a*,*b*; Frensham *et al.* 1997, 1998; Smith *et al.* 2001*a*,*b*, for example). In all these approaches the $V \times E$ interaction effects, together with at least one set of main effects, is assumed to be random. Each set of random effects is assumed to be independent and to follow a Gaussian distribution with zero mean and constant variance.

The issue of whether the main effects, particularly variety effects, should be regarded as fixed or random is an important one and is not confined to the simple ANOVA type models but arises in the context of all mixed model analyses of MET data. Therefore, the issue is discussed in detail later. For now, the ANOVA type mixed model is considered, using the two scenarios that appear most commonly in the literature, namely (a) random variety and fixed environment main effects and (b) fixed variety and random environment main effects. The extension to both variety and environment main effects as random is trivial.

First, consider case (a), that is, the model in Eqn (4) with the assumption that the vector $\boldsymbol{\theta}$ comprises fixed effects and $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ are random effects with

$$\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\delta} \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \sigma_\alpha^2 \boldsymbol{I}_m & \mathbf{0} \\ \mathbf{0} & \sigma_\delta^2 (\boldsymbol{I}_t \otimes \boldsymbol{I}_m) \end{pmatrix} \right]$$

where σ_α^2 and σ_δ^2 are the variance components for the variety main effects and $V \times E$ interaction respectively. Thus

$$E(\boldsymbol{\eta}) = \mathbf{1}_{mt}\mu + (\boldsymbol{I}_t \otimes \mathbf{1}_m)\boldsymbol{\theta}$$
$$\text{var}(\boldsymbol{\eta}) = (\sigma_\alpha^2 \boldsymbol{J}_t + \sigma_\delta^2 \boldsymbol{I}_t) \otimes \boldsymbol{I}_m \qquad (5)$$

where $\boldsymbol{J}_t$ is the $t \times t$ unit matrix (i.e. with all elements equal to one). Expansion of this variance structure (see Appendix) reveals that all $V \times E$ effects have the same variance ($\sigma_\alpha^2 + \sigma_\delta^2$), $V \times E$ effects for different varieties are uncorrelated and $V \times E$ effects for different pairs of environments (for the same variety) all have the same covariance (σ_α^2) and thence correlation. This variance structure is known as a uniform (or compound symmetric) structure.

If the converse scenario (case (b)) is assumed, in which the vector $\boldsymbol{\alpha}$ comprises fixed effects and $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ are random effects, the following is obtained

$$E(\boldsymbol{\eta}) = \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha}$$
$$\text{var}(\boldsymbol{\eta}) = \boldsymbol{I}_t \otimes (\sigma_\theta^2 \boldsymbol{J}_m + \sigma_\delta^2 \boldsymbol{I}_m) \qquad (6)$$

where σ_θ^2 is the variance component for the environment main effects. Once again this is a uniform variance structure but now $V \times E$ effects for different environments are uncorrelated and $V \times E$ effects for different pairs of varieties (for the same environment) all have the same covariance (σ_θ^2).

With balanced data, the models in Eqns (5) and (6) can be analysed using ANOVA techniques. This analysis provides least squares means for varieties, environments and the two-way table of $V \times E$ means. Variance components may be estimated by equating mean squares in the ANOVA table with their expectations. In the more common situation of unbalanced data, the ANOVA method is unsuitable so an alternative is required. The most popular method for estimation of variance components in unbalanced data is Residual Maximum Likelihood (REML, Patterson & Thompson 1971). Given estimates of the variance components, the fixed effects in Eqns (5) and (6) may be estimated using Empirical Best Linear Unbiased Estimation (E-BLUE) and the random effects predicted using Empirical Best Linear Unbiased Prediction (E-BLUP). (See section 'Estimation and software' for full details on methods of estimation.)

The analyses based on the models in Eqns (5) and (6) are deficient in two key areas. Firstly, they only provide information on the magnitude of $V \times E$ interaction (through the size of the estimate of the variance component σ_δ^2) and do not formally explore patterns of interaction. Secondly, the variance assumptions underpinning the models may be unrealistic. The uniform variance structure may be unsatisfactory due to heterogeneity of both variance and covariance. Many authors (including Patterson & Nabugoomu 1992; Frensham *et al.* 1997; Cullis *et al.* 1998) have recognized the possibility of variance heterogeneity among $V \times E$ interaction effects. In addition, a relaxation of the independence assumption for $V \times E$ interactions may be necessary. These issues are addressed in the more complex mixed models described in the following sections.

Mixed models with regressions on environmental variables

One method for exploring $V \times E$ interaction is to use a regression-based approach that relates varietal performance to environmental information such as rainfall, soil moisture and sowing date. Generally this approach has been considered within the framework of fixed effects models (Knight 1970; Freeman & Perkins 1971; Hardwick & Wood 1972, for example) but more recently mixed model approaches have been used. The focus in the present paper is on the approaches of Piepho et al. (1998) and Theobald et al. (2002). Note also that some authors have used mixed models with varietal covariates in order to investigate $V \times E$ interaction (Cullis et al. 1996 b; Frensham et al. 1998).

The base-line model used by Piepho et al. (1998) for a single environmental covariate is given by

$$\boldsymbol{\eta} = \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha} + (\boldsymbol{x} \otimes \boldsymbol{I}_m)\boldsymbol{\beta} + \boldsymbol{\delta} \qquad (7)$$

where $\boldsymbol{x}$ is the $t \times 1$ vector of covariate values, $\boldsymbol{\beta}$ is the $m \times 1$ vector of variety regression coefficients and the vector $\boldsymbol{\delta}$ now represents residual $V \times E$ interactions (that is, unexplained by the regression). It is assumed that the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ comprise fixed effects and $\boldsymbol{\delta}$ is a vector of random effects. Piepho et al. (1998) assume a separable variance matrix for $\boldsymbol{\delta}$ that allows correlations between varieties but for simplicity it is assumed in the present paper that $\text{var}(\boldsymbol{\delta}) = \sigma_\delta^2 \boldsymbol{I}_t \otimes \boldsymbol{I}_m$. Thus

$$\begin{aligned} \text{E}(\boldsymbol{\eta}) &= \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha} + (\boldsymbol{x} \otimes \boldsymbol{I}_m)\boldsymbol{\beta} \\ \text{var}(\boldsymbol{\eta}) &= \sigma_\delta^2 \boldsymbol{I}_t \otimes \boldsymbol{I}_m \end{aligned} \qquad (8)$$

The base-line model used by Theobald et al. (2002) is similar to that of Piepho et al. (1998) except that Theobald et al. (2002) use a Bayesian approach to estimation, so assume all effects to be random and specify prior probability distributions for them. Thus the interpretation of the model in terms of what constitutes the mean and variance of the $V \times E$ effects is very different from a frequentist approach. Since all other authors referenced in the present paper have used the latter, no further details of the Theobald et al. (2002) model are presented; the reader is referred instead to that paper.

The use of environmental covariates for the analysis of MET data has been criticized because the regressions often only explain a small proportion of $V \times E$ interaction. There may be additional difficulties in terms of data availability and/or selection of variables to include in the regression. The key advantage, however, is that for suitably chosen covariates, the associated $V \times E$ interaction is predictable. This has important consequences for crop variety evaluation programmes in particular, since it enables local-area predictions, that is, predictions of varietal performance that are specific to individual farmer conditions.

Mixed models with regressions on environmental means

A method that has often been used to study varietal response to environment is the regression of varietal yield on the mean yield of all varieties in each environment. Yates & Cochran (1938) introduced the idea and Finlay & Wilkinson (1963) gave a thorough account. Finlay & Wilkinson (1963) state the logic underpinning the approach, namely that 'The mean yield of all varieties at a site and season [provide] a quantitative grading of environments, and ... varieties specifically adapted to good or poor seasons and those showing general adaptability may be identified.' Thus, the use of environment means is a surrogate for a potentially complex regression model involving numerous environmental covariates.

Mixed model analogues of the Finlay-Wilkinson model have been proposed by Gogel et al. (1995) and Nabugoomu et al. (1999). The base-line mixed model in the Gogel et al. (1995) and Nabugoomu et al. (1999) approaches uses

$$\boldsymbol{\eta} = \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha} + (\boldsymbol{I}_t \otimes \boldsymbol{\beta})\boldsymbol{\theta} + \boldsymbol{\delta} \qquad (9)$$

Gogel et al. (1995) and Nabugoomu et al. (1999) assume that the vector $\boldsymbol{\alpha}$ comprises fixed effects and $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ are random effects with

$$\begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\delta} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \sigma_\theta^2 \boldsymbol{I}_t & \mathbf{0} \\ \mathbf{0} & \sigma_\delta^2 \boldsymbol{I}_t \otimes \boldsymbol{I}_m \end{pmatrix} \right]$$

from which is derived

$$\begin{aligned} \text{E}(\boldsymbol{\eta}) &= \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha} \\ \text{var}(\boldsymbol{\eta}) &= \boldsymbol{I}_t \otimes (\sigma_\theta^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \sigma_\delta^2 \boldsymbol{I}_m) \end{aligned} \qquad (10)$$

Note that the model in Eqn (9) can be re-expressed as

$$\boldsymbol{\eta} = \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha} + (\boldsymbol{\theta} \otimes \boldsymbol{I}_m)\boldsymbol{\beta} + \boldsymbol{\delta} \qquad (11)$$

which has the same form as Eqn (7) except that in Eqn (11) the environmental covariate ($\boldsymbol{\theta}$) is unknown so must be estimated from the data.

Finlay & Wilkinson (1963) proposed a graphical display of the results of their analysis, namely a plot of variety regression coefficients against variety mean yields, in order to visualize varietal stability and performance. This display could also be used in association with the mixed model analysis. As with the regression approach using 'external' environmental covariates, the regression on environment mean yield has been criticized for the inability to explain large portions of $V \times E$ interaction. Also, unlike external covariates, trial mean yield must be estimated from the data themselves so is subject to error. However, there is still the key advantage that local-area

predictions may be possible since farmers have knowledge of the likely mean yield (or range of yields) for their environment.

Multiplicative mixed models

Kempton (1984) discusses the method of principal components as a means of summarizing the patterns of varietal response to different environments. In this method the matrix of estimated $V \times E$ interaction effects from the classical ANOVA model of Eqn (4) is subjected to a principal component analysis (PCA). $V \times E$ interaction is thereby decomposed into a number of multiplicative terms. The conjecture is that most of the $V \times E$ interaction can be explained by the first few terms from the PCA and that these have some meaningful interpretation. This method of analysis dates back to Mandel (1971) and was more recently popularized for MET data under the banner of AMMI (Additive Main effects and Multiplicative Interaction, Gauch 1992). There are variants of the method (the so-called Shifted Multiplicative Model, SHMM) in which variety and/or environment main effects are not fitted in the ANOVA, so that the PCA is performed on combined effects rather than the interactions.

Piepho (1997) and Smith *et al.* (2001*b*) employed a mixed model analogue of PCA for the analysis of MET data. These authors differ in that Smith *et al.* (2001*b*) assume random variety and fixed environment effects, whereas Piepho (1997) assumes the converse. Smith *et al.* (2001*b*) do not explicitly fit variety main effects but the extension is straightforward and for comparative purposes they are included here. Thus the extended Smith *et al.* (2001*b*) approach uses

$$\boldsymbol{\eta} = \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha} + (\boldsymbol{I}_t \otimes \mathbf{1}_m)\boldsymbol{\theta} + (\boldsymbol{\Lambda}_e \otimes \boldsymbol{I}_m)\boldsymbol{f}_v + \boldsymbol{\delta}$$

where $\boldsymbol{\Lambda}_e$ is a $t \times k$ matrix of environment loadings, $\boldsymbol{f}_v$ is the associated $mk \times 1$ vector of variety scores and k is the number of components (multiplicative terms) included in the analysis. It is assumed that the vector $\boldsymbol{\theta}$ comprises fixed effects and $\boldsymbol{\alpha}, \boldsymbol{f}_v$ and $\boldsymbol{\delta}$ are random effects with

$$\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{f}_v \\ \boldsymbol{\delta} \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \sigma_\alpha^2 \boldsymbol{I}_m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_k \otimes \boldsymbol{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_e \otimes \boldsymbol{I}_m \end{pmatrix} \right]$$

where $\boldsymbol{\Psi}_e$ is a diagonal $t \times t$ matrix with elements commonly referred to as specific variances, from which is derived

$$E(\boldsymbol{\eta}) = \mathbf{1}_{mt}\mu + (\boldsymbol{I}_t \otimes \mathbf{1}_m)\boldsymbol{\theta}$$

$$\text{var}(\boldsymbol{\eta}) = \sigma_\alpha^2 (\boldsymbol{J}_t \otimes \boldsymbol{I}_m) + (\boldsymbol{\Lambda}_e \boldsymbol{\Lambda}_e' + \boldsymbol{\Psi}_e) \otimes \boldsymbol{I}_m \quad (12)$$

$$= (\boldsymbol{\Lambda}_e^* \boldsymbol{\Lambda}_e^{*\prime} + \boldsymbol{\Psi}_e) \otimes \boldsymbol{I}_m$$

where $\boldsymbol{\Lambda}_e^* = [\sigma_\alpha \mathbf{1}_t \ \boldsymbol{\Lambda}_e]$. The variance structure for the $V \times E$ interaction effects, namely $(\boldsymbol{\Lambda}_e \boldsymbol{\Lambda}_e' + \boldsymbol{\Psi}_e) \otimes \boldsymbol{I}_m$ is known as a Factor Analytic (FA) structure of order k (see Mardia *et al.* 1988, for example). The overall variance structure for $V \times E$ effects in Eqn (12) is also an FA structure but with order $k+1$ and constraints such that the first set of loadings is proportional to the unit vector. A key consequence of the model is that it allows for fairly general $V \times E$ variance and covariance heterogeneity between environments.

Piepho (1997) assumes a model of the form

$$\boldsymbol{\eta} = \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha} + (\boldsymbol{I}_t \otimes \mathbf{1}_m)\boldsymbol{\theta} + (\boldsymbol{I}_t \otimes \boldsymbol{\Lambda}_v)\boldsymbol{f}_e + \boldsymbol{\delta}$$

where $\boldsymbol{\Lambda}_v$ is a $m \times k$ matrix of variety loadings, $\boldsymbol{f}_e$ is the associated $tk \times 1$ vector of environment scores. It is assumed that the vector $\boldsymbol{\alpha}$ comprises fixed effects and $\boldsymbol{\theta}, \boldsymbol{f}_e$ and $\boldsymbol{\delta}$ are random effects with

$$\begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{f}_e \\ \boldsymbol{\delta} \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \sigma_\theta^2 \boldsymbol{I}_t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_t \otimes \boldsymbol{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\delta^2 \boldsymbol{I}_t \otimes \boldsymbol{I}_m \end{pmatrix} \right]$$

from which is derived

$$E(\boldsymbol{\eta}) = \mathbf{1}_{mt}\mu + (\mathbf{1}_t \otimes \boldsymbol{I}_m)\boldsymbol{\alpha}$$

$$\text{var}(\boldsymbol{\eta}) = \sigma_\theta^2 (\boldsymbol{I}_t \otimes \boldsymbol{J}_m) + \boldsymbol{I}_t \otimes (\boldsymbol{\Lambda}_v \boldsymbol{\Lambda}_v' + \sigma_\delta^2 \boldsymbol{I}_m) \quad (13)$$

$$= \boldsymbol{I}_t \otimes (\boldsymbol{\Lambda}_v^* \boldsymbol{\Lambda}_v^{*\prime} + \sigma_\delta^2 \boldsymbol{I}_m)$$

where $\boldsymbol{\Lambda}_v^* = [\sigma_\theta \mathbf{1}_m \ \boldsymbol{\Lambda}_v]$. As in Eqn (12) the structure in Eqn (13) is an FA structure with order $k+1$, but the model now allows for $V \times E$ variance and covariance heterogeneity between varieties rather than environments. Also note that Piepho (1997) assumes a common residual $V \times E$ variance (σ_δ^2) whereas Smith *et al.* (2001*b*) allow for different residual $V \times E$ variances for each trial (through the diagonal matrix $\boldsymbol{\Psi}_e$).

A key feature of the FA model for MET data is the generality of the associated variance structure for $V \times E$ effects, either in the environment or variety dimension (models (12) or (13) respectively). The most general variance model, and therefore the model that will provide the best fit (in a likelihood sense) to the data, is an unstructured matrix. This can be difficult to fit from a computational perspective, particularly for large structures (large m for a matrix in the variety dimension or large t for the environment dimension). The FA model with sufficient multiplicative terms has been found to provide a good and parsimonious approximation to the unstructured form and is generally more computationally robust (see Thompson *et al.* 2003). Smith *et al.* (2001*b*) use the FA model in this context where the analysis was motivated by the quantitative genetics approach to $V \times E$ as explained in Falconer & Mackay (1996).

Falconer & Mackay (1996) state that 'The concept of genetic correlation can be applied to the solution of some problems concerned with the interaction of genotype and environment … a character measured in two different environments is to be regarded not as one character but as two … If the genetic correlation between them is high, then performance in two different environments represents very nearly the same character … If it is low, then the characters are to a great extent different.' Thus Smith *et al.* (2001*b*) use an FA model to approximate an unstructured matrix for the environment dimension of var $(\boldsymbol{\eta})$ (that is, the matrix of genetic variances and covariances between environments).

Estimation of the models (12) and (13) provides estimates (or predictions) of tables of means for varieties, environments and the two-way table of $V \times E$ means. Additionally, estimates (or predictions) of loadings and scores can be obtained, thence displayed graphically using tools such as bi-plots in order to better understand $V \times E$ interaction. The FA model has some similarities with the regression approach considered earlier. For example, the Smith *et al.* (2001*b*) approach can be viewed as a regression of varietal yield on environmental covariates (the loadings) but the covariates are estimated from the data rather than measured externally. This has the advantage that the regressions usually account for a large proportion of $V \times E$ interaction. However, the environmental covariates are data-dependent so cannot be used for local-area predictions.

Variety effects: fixed or random?

With the widespread adoption of mixed model analyses for MET data there has been a dichotomy of thought as to the classification of variety effects as fixed or random. This is evident from the examples presented in the previous sections. The present authors believe the choice depends on the aim of the analysis and consideration of the properties of the two types of estimation procedures, namely empirical best linear unbiased prediction (E-BLUP) for random effects and empirical best linear unbiased estimation (E-BLUE) for fixed effects (see Section 'Estimation and software').

If the aim of the analysis is selection (that is, to identify the best varieties of those under consideration) then the rankings of the estimated variety effects are required to be as close as possible to the rankings of the true variety effects. In more exact terms, a set of estimates of variety effects is required that best predict the true effects. By definition, this implies the use of BLUP so that variety effects should be regarded as random. The optimality properties of BLUP are based on the assumption that the variance parameters in the model are known. In general, this is not the case and the parameters are estimated from the data. The only question that remains, therefore, is whether the estimates of the variance parameters are sufficiently precise to ensure that the optimality of BLUP is maintained with E-BLUP.

If the aim of the analysis is to determine the difference between specific pairs of varieties, then the use of BLUP as an estimation method is inappropriate since the BLUP of a specific difference is biased. Thus, in this case variety effects should be regarded as fixed.

The key issue, therefore, is a clear definition of the aim of the analysis. In order to pursue this, common practice is followed with differentiating between breeding and evaluation programmes, although the distinction is sometimes hazy. Breeding programmes are concerned with the early stages of varietal evaluation (Finney (1980) refers to this as the 'cradle to kindergarten' phase) in which large numbers (often greater than 1000) of new breeding lines are grown in small numbers (usually less than 3) of field trials. The 'best' lines are selected to continue to the next stage of testing, in which fewer lines are evaluated in more locations. The process culminates in the testing of a small number (usually less than 40) of elite breeding lines, together with commercial standard varieties, in a large number of trials that span a wide range of geographic locations and several growing seasons. On the basis of these trials, a new breeding line may be recommended for commercial use and thence make the transition to a commercial variety. These trials (Finney (1980): 'kindergarten to grave') are usually the domain of crop variety evaluation programmes (CVEP). In many countries, CVEP are funded by government institutions and/or farmer groups, and their mission is to conduct independent evaluation of potential new varieties. In the UK, for example, the evaluation system is co-ordinated and run under a statutory authority. Breeders and seed merchants enter new varieties for testing in so-called National List (NL) trials and thence, if selected, the varieties proceed to Recommended List (RL) trials from which farmer recommendations are made (see Patterson & Silvey (1980) for further details). In Australia, CVEP are run by state departments of agriculture, but a nationally based system similar to that in the UK is about to be launched.

It is clear that the aim of the analysis of breeding data is selection so that the use of random variety effects is appropriate. Some statisticians advocate the use of random effects in this setting because they regard that the varieties themselves are a random sample from a population. After some unspecified number of stages of selection, this ceases to be a reasonable assumption so that at this point variety effects are regarded as fixed. The present authors do not adhere to this line of reasoning.

Most of the literature on methods for analysing MET data appears to be focused on evaluation data (or at least the example data-sets used are of this nature). It is in this setting that the fixed versus random variety effects issue is most heatedly debated. We believe that the aim of analysis of data from evaluation programmes such as those in Australia and the UK is still selection but it is now the farmer making the selection decisions rather than the plant breeder. The farmer wishes to know which varieties are best for his/her environment. These views are shared by Patterson & Silvey (1980) in their landmark paper describing the analysis of data from the UK evaluation system in which it was stated that 'The main objective of a series of NL or RL trials is to identify, with minimum selection error, the best varieties for cultivation and use.' Thus, once again the rankings of estimated variety effects (possibly within specific environments) are required to correlate well with the true rankings. In contrast, a seed company may wish to know the difference between their potential new variety and other commercial varieties, an aim that would require the use of fixed variety effects. The present authors believe, however, that the analysis of evaluation data is conducted 'for the common good', that is, to allow farmers to identify and thus adopt the best varieties for their environment. The assumption of random variety effects for both breeding and evaluation data is therefore made.

Of course, with balanced data and orthogonal analyses, the rankings of varieties would be the same in both the fixed and random variety settings. Even so, the present authors still prefer the use of random variety effects since the resultant predictions of genetic gain are more realistic than those based on fixed variety effects. The latter are generally over-optimistic due to selection bias (Patterson & Silvey 1980). An additional key advantage with the use of random variety effects is that it allows a valid analysis of data combined across stages of selection (often corresponding to a sequence of years). The analysis of such data is crucial for plant breeders since it provides more reliable estimates of variety main effects (being based on all relevant data, not merely the data for the current year) and since years are synonymous with seasons the analysis provides information on variety by season interactions. Henderson *et al.* (1959) discuss the estimation of fixed effects in the presence of selection. In our context, their suggestions imply that a likelihood approach with random variety effects reduces bias in the estimation of (fixed) year effects. Thompson (1973) also proposes a likelihood approach for data arising from selection and shows that this allows unbiased estimation of genetic variance parameters both before and after selection. The prediction of genetic effects (in our case variety effects) under selection is discussed by Thompson (1979).

It is interesting to note that the variance structures for V × E effects for all models presented earlier have two basic forms, namely

$$\text{var}\,(\boldsymbol{\eta}) = \boldsymbol{G}_e \otimes \boldsymbol{I}_m \tag{14}$$

or

$$\text{var}\,(\boldsymbol{\eta}) = \boldsymbol{I}_t \otimes \boldsymbol{G}_v \tag{15}$$

where $\boldsymbol{G}_e$ and $\boldsymbol{G}_v$ are positive definite symmetric matrices of dimension $t \times t$ and $m \times m$ respectively. The model in Eqn (14) implies that V × E effects are correlated between environments whereas in Eqn (15) they are correlated between varieties. The 'side' on which effects are correlated is a direct consequence of whether variety or environment effects are assumed random. With random variety and fixed environment effects the model in Eqn (14) is obtained and the converse leads to the model in Eqn (15). The choice between the models in Eqns (14) and (15) may be driven by various influences including parsimony (clearly if t is much larger than m then Eqn (14) would be preferred), biology (for example, as discussed earlier, the theory of quantitative genetics would lead to the use of Eqn (14)) and goodness of fit. The variance structure for V × E effects need not be restricted to the 'one-sided' forms in Eqns (14) and (15). A more general form is

$$\text{var}\,(\boldsymbol{\eta}) = \boldsymbol{G}_e \otimes \boldsymbol{G}_v$$

Both matrices $\boldsymbol{G}_e$ and $\boldsymbol{G}_v$ may contain unknown parameters (to be estimated). An important case, however, is where $\boldsymbol{G}_v$ is a known matrix reflecting the pedigrees or marker genotypes of the varieties.

Models for residual effects

The vector of residual effects, $\boldsymbol{\varepsilon}$, in Eqn (2) consists of sub-vectors, that is, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1 \ldots \boldsymbol{\varepsilon}'_t)'$ where $\boldsymbol{\varepsilon}_j$ is the $n_j \times 1$ vector of residual effects for the jth trial, $j = 1 \ldots t$. The two key issues for specification of models for $\boldsymbol{\varepsilon}$ are the form of analysis that would be used for an individual trial and the need to allow for heterogeneity of the associated variance parameters across trials. The literature on methods for individual variety trial analysis is quite diverse but the methods can be broadly classified as either randomization or model based. In the former, the model for residual effects is determined purely from the experimental design, whereas in the latter it is either assumed or selected with the objective of providing a good fit to the data. In order to explore this further, the general statistical model for the $n_j \times 1$ vector $\boldsymbol{y}_j$ of yields for the jth trial is specified. This can be expressed as

$$\boldsymbol{y}_j = \boldsymbol{M}_j \boldsymbol{\eta}_j + \boldsymbol{\varepsilon}_j \tag{16}$$

where $\boldsymbol{\eta}_j$ represents the variety effects for the jth trial (that is, the sub-vector of $\boldsymbol{\eta}$ corresponding to

that trial) and M_j is the associated design matrix. A further model for the residual effects may be required:

$$\boldsymbol{\varepsilon}_j = X_{p_j}\boldsymbol{\tau}_{p_j} + Z_{p_j}\boldsymbol{u}_{p_j} + \boldsymbol{e}_j \qquad (17)$$

where $\boldsymbol{\tau}_{p_j}$ and $\boldsymbol{u}_{p_j}$ are vectors of fixed and random effects, respectively, with associated design matrices X_{p_j} (assumed to have full column rank) and Z_{p_j} and $\boldsymbol{e}_j$ are vectors of plot error effects for the jth trial. The vectors $\boldsymbol{\tau}_{p_j}$ and $\boldsymbol{u}_{p_j}$ represent effects peripheral to the variety effects in the sense that they are not the main focus of the analysis. Typically, they are associated with the experimental design and/or are included to model field trend (see below). The vectors $\boldsymbol{u}_{p_j}$ and $\boldsymbol{e}_j$ are assumed to follow a Gaussian distribution with zero means and variance matrix given by

$$\text{var}\begin{pmatrix} \mathbf{u}_{p_j} \\ \boldsymbol{e}_j \end{pmatrix} = \begin{bmatrix} \boldsymbol{G}_{p_j} & 0 \\ 0 & \boldsymbol{R}_j \end{bmatrix}$$

Finally, by substituting Eqn (17) into Eqn (16) the full mixed model for the jth trial can be written as

$$\boldsymbol{y}_j = M_j\boldsymbol{\eta}_j + X_{p_j}\boldsymbol{\tau}_{p_j} + Z_{p_j}\boldsymbol{u}_{p_j} + \boldsymbol{e}_j \qquad (18)$$

Variety trials are usually laid out as rectangular arrays that can be indexed by field rows and columns. This scenario is assumed in the following, with the data ordered correspondingly as rows within columns. The assumption of a lattice arrangement of plots can be relaxed but the ordering of data according to the field plot layout is important for the model-based approaches outlined below.

A randomization-based analysis may be conducted using the model in Eqn (18) with sub-vectors of $\boldsymbol{u}_{p_j}$ corresponding to terms in the block structure of the experiment (see Nelder 1965, for a complete account). For example, if the experiment were designed as a randomized complete block (RCB) experiment with n_{r_j} replicates (complete blocks) then $\boldsymbol{u}_{p_j}$ would have length n_{r_j} (comprising an effect for each replicate) and the effects would be assumed independent with constant variance $\sigma_{r_j}^2$, say. Thus $\boldsymbol{G}_{p_j} = \sigma_{r_j}^2 I_{n_{r_j}}$. The vector of plot error effects would then comprise independent effects with constant variance σ_j^2, thence $R_j = \sigma_j^2 I_{n_j}$. In an incomplete block (IB) design with n_{r_j} replicates and n_{b_j} (incomplete) blocks per replicate there would be two sub-vectors in $\boldsymbol{u}_{p_j}$, the first corresponding to the replicate effects and the second to block within replicate effects. Independence is assumed both within and between these sub-vectors and the effects have associated variance components of $\sigma_{r_j}^2$ and $\sigma_{b_j}^2$ for replicates and blocks within replicates, respectively. Thus $\boldsymbol{G}_{p_j} = \text{diag}(\sigma_{r_j}^2 \boldsymbol{I}_{n_{r_j}}, \sigma_{b_j}^2 \boldsymbol{I}_{n_{b_j}})$. As in the RCB design this would give $\boldsymbol{R}_j = \sigma_j^2 \boldsymbol{I}_{n_j}$. Note that Nelder (1954) discusses the need to allow variance components associated with blocking factors to be negative in order for the mixed model to

provide a proper surrogate for the randomization analysis.

Model-based approaches for the analysis of field trials focus on the need to control spatial variation. As implied by the terminology, this variation is linked to the location of plots in the field and may be due, for example, to fluctuations in soil fertility. Numerous authors have proposed analytical methods to remove the effects of such trend from the estimation of varietal contrasts. The earliest method was that of Papadakis (1937), in which neighbouring plot yields were used as covariates in the analysis. The next major contribution to the area was that of Wilkinson et al. (1983), who suggested that spatial field trend could be expressed as the sum of two components, namely a smooth trend and an independent error term. The assumed trend was removed by (second) differencing the data. Other authors have used the method of differencing adjacent plot yields as a means of removing trend (Green et al. 1985; Besag & Kempton 1986, for example).

Gleeson & Cullis (1987), Martin (1990) and Cullis & Gleeson (1991) proposed approaches that model trend directly using time series models (with differencing still having a role as a means to achieve stationarity). A key aspect of Martin (1990) and Cullis & Gleeson (1991) is the use of separable correlation models to accommodate trend in two dimensions (field rows and columns). Zimmerman & Harville (1991) also propose a direct modelling approach but use models based on the theory of random fields; spatial variation was viewed as comprising two sources, namely large-scale variation that is modelled through the mean, and small-scale variation that is modelled through a correlation structure.

Gilmour et al. (1997) extended the approach of Cullis & Gleeson (1991) by partitioning spatial variation into two types of smooth trend (local and global) and extraneous variation. Local trend reflects, for example, small-scale soil depth and fertility fluctuations. Global trend reflects non-stationary trend across the field. Extraneous variation is often linked to the management of the trial. In the Gilmour et al. (1997) approach, global trend and extraneous variation are accommodated in the model by including appropriate effects in $\boldsymbol{\tau}_{p_j}$ and/or $\boldsymbol{u}_{p_j}$. Local stationary trend is accommodated using a correlation structure for $\boldsymbol{R}_j$. Thus there are similarities with the Zimmerman & Harville (1991) approach. Gilmour et al. (1997) suggest that a separable autoregressive correlation process of order one usually provides a reasonable fit to the data. They stress the importance of using diagnostic tools to check model adequacy.

Most of the current spatial approaches for the analysis of field trials are of the form advocated by Zimmerman & Harville (1991) and Gilmour et al. (1997), that is, they involve a direct modelling of

local spatial trend using a covariance model. In terms of the model in Eqn (18) this is specified through the form of R_j.

Model-based spatial analyses for variety trials now have wide usage. They can lead to substantial gains in terms of response to selection compared with the randomization based approach, particularly when the experimental design involves large block sizes. The major criticism of model-based approaches is that estimates of treatment effects and their standard errors rely solely on the chosen model, whereas the randomization-based analysis is validated by recourse to randomization theory. In the experience of the current authors with conducting the annual analyses of variety trials from most Australian public breeding and evaluation programmes, the gains of a spatial approach outweigh this potential disadvantage. To safeguard against this to some extent, an approach is used that merges the randomization and spatial approaches. The randomization-based model is used as the baseline (maintaining the associated random terms irrespective of their significance) and this is built on to model remaining spatial variation. For the latter, the approach of Gilmour *et al.* (1997) is used. Thus a spatial model is not regarded as a replacement for the randomization-based model but rather as an enhancement to better accommodate field trend.

The benefits of spatial methods flow through to the analysis of MET data, where the importance of allowing for heterogeneity of residual variance parameters (that is, associated with u_{p_j} and e_j) between trials is also noted (see Smith *et al.* 2001*b*, for example). Many authors assume simple block models (RCB, for example) with both common block variance components for all trials and a common plot error variance for all trials. With Australian data, such models rarely provide a good fit and can induce misleading sources of V × E interaction. In terms of spatial modelling in the MET analysis, Cullis *et al.* (1998) and Smith *et al.* (2001*b*), who estimate a separate spatial covariance structure for each trial, are followed.

The full model

In terms of the overall analysis for MET data the model in Eqn (2) can now be written as

$$y|\eta = M\eta + X_p\tau_p + Z_p u_p + e$$
$$\eta = X_\eta\tau_\eta + Z_\eta u_\eta \tag{19}$$

where $\eta = (\eta'_1 \ldots \eta'_t)'$ is, as before, the vector of V × E effects with associated design matrix $M = \text{diag}(M_j)$. The vectors $\tau_p = (\tau_{p'_1} \ldots \tau_{p'_t})'$ and $u_p = (u_{p'_1} \ldots u_{p'_t})'$ are the combined vectors of trial-specific peripheral effects (fixed and random, respectively) and $e = (e'_1 \ldots e'_t)'$ is the combined vector of plot error effects

from all trials. The design matrices for τ_p and u_p are given by $X_p = \text{diag}(X_{p_j})$ and $Z_p = \text{diag}(Z_{p_j})$. The vectors τ_η and u_η (with design matrices X_η and Z_η) are the fixed and random effects associated with the model for V × E effects. The random effects u_η are assumed to follow a Gaussian distribution with zero mean and variance matrix G_η. For example, in the variance component V × E model of Eqn (5) the vector of fixed effects is given by $\tau_\eta = (\mu, \theta')'$ with design matrix $X_\eta = [\mathbf{1}_{mt} \ I_t \otimes \mathbf{1}_m]$ and the vector of random effects is given by $u_\eta = (\alpha', \delta')'$ with design matrix $Z_\eta = [\mathbf{1}_t \otimes I_m \ I_t \otimes I_m]$ and variance matrix $G_\eta = \text{diag}(\sigma_\alpha^2 I_m, \sigma_\delta^2 I_t \otimes I_m)$.

The random effects in Eqn (19) are assumed to follow a Gaussian distribution with zero mean and variance matrix

$$\text{var}\begin{pmatrix} u_\eta \\ u_p \\ e \end{pmatrix} = \begin{bmatrix} G_\eta & 0 & 0 \\ 0 & G_p & 0 \\ 0 & 0 & R \end{bmatrix}$$

Independence of the sub-vectors in u_p and e is assumed so that $G_p = \text{diag}(G_{p_j})$ and $R = \text{diag}(R_j)$. As a justification for this, recall that trials may comprise factorial combinations of geographic locations and years. It is clearly valid to assume independence of block and plot error effects between geographic locations, but the temporal aspect requires clarification. In terms of annual crops, the independence assumption is reasonable since a new trial is sown each year and it is unlikely that an identical set of field plots would be used in successive years. The situation for perennial crops (pastures, for example) is very different, since the associated data comprise repeated measurements made on the same plots. The focus in the present paper is on data from annual crops so block diagonal forms for G_p and R are assumed. The extension to allow correlations between trials (for repeated measures MET data, for example) is possible.

ESTIMATION AND SOFTWARE

Most of the literature concerning mixed model analyses for MET data contain frequentist approaches to estimation in which the variance parameters of the model are estimated using residual maximum likelihood (REML, Patterson & Thompson 1971) and the fixed and random effects are estimated using best linear unbiased estimation (BLUE) and best linear unbiased prediction (BLUP) respectively. There is an exception, namely, Theobald *et al.* (2002) who use Bayesian estimation methods. In the following, a brief summary of the frequentist approach is presented. For a more detailed account of estimation in general linear mixed models the reader is referred elsewhere (see Cullis *et al.* 2004, for example).

The model in Eqn (19) is re-written to the usual form for a general linear mixed model, namely

$$y = X\tau + Zu + e \qquad (20)$$

where $\tau = (\tau_\eta', \; \tau_p')'$ is the vector of fixed effects with associated design matrix $X = [X_\eta \; X_p]$ (assumed full rank) and $u = (u_\eta', \; u_p')'$ is the vector of random effects with associated design matrix $Z = [Z_\eta \; Z_p]$. It is assumed that the joint distribution of u and e is Gaussian with zero mean and variance matrix

$$\mathrm{var}\begin{pmatrix} u \\ e \end{pmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \qquad (21)$$

where $G = \mathrm{diag}\;(G_\eta, \; G_p)$ and $G = G(\gamma)$, $R = R(\phi)$ and $\gamma = (\gamma_\eta', \; \gamma_p')'$ is the vector of variance parameters associated with u (partitioned into $V \times E$ and peripheral parameters) and ϕ is the vector of variance parameters associated with e. The distribution of y is then Gaussian with mean $X\tau$ and variance matrix var $(y) = H = ZGZ' + R$.

The fixed and random effects in Eqn (20) are estimated using best linear unbiased estimation and prediction, respectively. This leads to the estimates

$$\hat{\tau} = (X'H^{-1}X)^{-1}X'H^{-1}y$$

$$\tilde{u} = GZ'H^{-1}(y - X\hat{\tau})$$

Generally, the parameters in H are unknown so in practice are replaced by their REML estimates (see below). The resultant estimates of the fixed and random effects are then termed empirical BLUEs (E-BLUEs) and empirical BLUPs (E-BLUPs).

The most popular method for estimation of the variance parameters in a linear mixed model is REML. This involves maximization of the residual log-likelihood that can be written as

$$\ell_R = -\tfrac{1}{2}\{\log|H| + \log|X'H^{-1}X| + y'Py\} \qquad (22)$$

where $P = H^{-1} - H^{-1}X\,(X'H^{-1}X)^{-1}\,X'H^{-1}$.

In general, maximization of this likelihood with respect to the vector of variance parameters $\kappa = (\gamma', \; \phi')'$ requires an iterative scheme. In the original REML paper, Patterson & Thompson (1971) used a Fisher Scoring (FS) algorithm that requires calculation of the expected information matrix for κ. This is very computer-intensive and may be untenable for large data sets or complex variance models. As a result, methods have been devised that are less computer-intensive and employ sparse matrix methods. These include derivative free methods (Smith & Graser 1986), first-order schemes such as the Expectation-Maximization (EM) scheme (Dempster et al. 1977) and the computationally efficient second-order scheme known as the Average Information (AI) algorithm (Gilmour et al. 1995). A thorough account and comparison of these and related iterative schemes is given in Cullis et al. (2004). The present authors use the AI algorithm as implemented in the commercial software ASReml (Gilmour et al. 2002), GENSTAT (1998) and samm (Butler et al. 2003) (a suite of functions written for S-language (Becker et al. 1988) environments). What follows is a summary of the algorithm.

In derivative-based methods (including the AI algorithm) the residual log-likelihood is maximized using the score equations:

$$U_R(\kappa_i) = -\tfrac{1}{2}\{\mathrm{tr}(P\dot{H}_i) - y'P\dot{H}_iPy\}$$

where $\dot{H}_i = \partial H/\partial_{\kappa_i}$. The REML estimate of κ is obtained as the solution to $U_R(\kappa) = 0$. In general, this must be solved iteratively. Given an estimate $\kappa = \kappa^{(m)}$, an update can be obtained as

$$\kappa^{(m+1)} = \kappa^{(m)} + [I^{(m)}]^{-1}U_R(\kappa^{(m)})$$

where $I^{(m)}$ is an information matrix for κ evaluated at $\kappa^{(m)}$. The type of information matrix used varies between schemes. The FS algorithm uses the expected information matrix, the Newton Raphson algorithm uses the observed information matrix and the AI algorithm uses the so-called average information matrix that is given by

$$I = \tfrac{1}{2}Q'PQ$$

where the columns of Q are working variables corresponding to κ and are given by

$$q_{\kappa_i} = \dot{H}_iPy$$

It should be noted that, for models in which the variance structure is linear in the parameters (for example in variance component models), elements in the average information matrix are exact averages of the corresponding elements in the observed and expected information matrices.

All of the models described in the current paper can be easily fitted using the AI algorithm. The key information required for each model are the score equations and working variables. The calculation of the score and working variables for standard variance component parameters is detailed in Gilmour et al. (1995). Calculations for spatial variance parameters can be found in Cullis et al. (1998) and Smith et al. (2001b). In the following, specific issues are considered for estimation of the $V \times E$ variance parameters for all models presented earlier.

In the variance component model of Eqn (5) the vector γ_η is given by $\gamma_\eta = (\sigma_\alpha^2, \; \sigma_\delta^2)'$. In the model of Eqn (6), $\gamma_\eta = (\sigma_\theta^2, \; \sigma_\delta^2)'$. Thus, in both models the parameters are standard variance components, the scores and working variables for which are given in Gilmour et al. (1995).

The regression model of Eqn (8) contains a single variance component so that $\gamma_\eta = \sigma_\delta^2$. In the regression on the mean model of Eqn (10) we have

$\gamma_{\eta} = (\sigma_{\theta}^2, \, \beta, \, \sigma_{\delta}^2)'$. First, it should be noted that the model is over-parameterized. Gogel *et al.* (1995) overcome this by reparameterizing the model in terms of $v_i = \beta_i \sigma_{\theta}$ ($i = 1 \dots m$) with $\sigma_{\theta} = \bar{v}$ whereby $\beta_i = v_i/\bar{v}$.

The multiplicative models described earlier are the most complex of the models presented. The vector γ_{η} for the model of Eqn (12) comprises σ_{α}^2, the individual elements of the loading matrix Λ_e and the diagonal elements (specific variances) of Ψ_e. The vector γ_{η} for the model of Eqn (13) is similar, comprising σ_{θ}^2, the individual elements of the loading matrix Λ_v and σ_{δ}^2. Smith *et al.* (2001*b*) give details of the scores and working variables for loadings and specific variances and note the need for constraints on the elements of the loading matrix when $k > 1$ in order to ensure uniqueness. The use of heterogeneous $V \times E$ variance in Eqn (12) compared with a common variance in Eqn (13) can lead to models with variance structures of less than full rank. This may occur when estimates of specific variances tend to zero. In this situation standard algorithms (including AI) for REML estimation are no longer applicable. Thompson *et al.* (2003) present a sparse implementation of the AI algorithm for fitting Factor Analytic and Reduced Rank (RR) variance models. This has the advantages over the estimation approach described in Smith *et al.* (2001*b*) in that convergence for FA models is faster and estimation of RR models is possible. It is noted that, in practice in the analysis of Australian MET data, the occurrence of zero specific variances is quite common.

There are now several statistical packages (including ASReml, GENSTAT, S-language packages and SAS; Littel *et al.* 1996) that allow REML estimation of a range of mixed models. The present authors have found the packages ASReml and GENSTAT and the samm functions (through S-language environments) to be the most suitable for the analysis of MET data, both in terms of the generality of models that can be fitted and the ease with which predictions and inference about varietal effects can be made. All models in the current paper are easily fitted and summarized using these software (code is available from the authors on request). An additional advantage with ASReml is the size of problem that can be handled. MET data-sets often involve large numbers of data points (often greater than 10 000) and require mixed model analyses with large numbers of random effects (often greater than 60 000). The present authors have found this to be possible only using the AI algorithm as implemented in ASReml.

Two-stage analyses

So far in the current paper, attention has been restricted to mixed model analyses of individual plot data. This is referred to as the 'one-stage' approach in which the models for residual effects are estimated simultaneously with models for $V \times E$ effects. Many authors (Patterson & Silvey 1980; Talbot 1984; Patterson & Nabugoomu 1992) use a two-stage approach in which variety means are first obtained from the analyses of individual trials and are then combined to form the data for an overall mixed model analysis. The two-stage approach is an approximation to the more efficient one-stage approach. Historically the two-stage approach was required due to the difficulty in obtaining individual plot data. Often this was not stored electronically but this should no longer be an issue. There may, however, be computational issues with the one-stage approach when complex $V \times E$ models are fitted to large data-sets. For example, ASReml has been used to fit the Smith *et al.* (2001*b*) model to individual plot data from 250 trials that involved a total of 400 varieties. At present this represents an upper bound to the size of problems that can be tackled, but work is in progress to extend this. In general, use of the more efficient one-stage approach is recommended. If the two-stage approach is to be adopted readers are referred to Smith *et al.* (2001*a*) who describe techniques, in particular the use of appropriate weights, aimed at more closely approximating the one-stage analysis.

CONCLUDING REMARKS

In the current paper the most popular mixed model approaches for the analysis of MET data have been described. They range from ANOVA type models (that is, models based on simple variance component assumptions for the random effects) through to models with more complex forms for modelling $V \times E$ interaction and error variation. Despite the clear benefits of the general mixed model approach, adoption within plant breeding and crop variety evaluation programmes has been very slow. In particular, the use of the more complex (and informative) models and the assumption of random rather than fixed variety effects is not widespread. This is in stark contrast to animal breeding programmes, in which REML and BLUP have been used for many years as the basis for selection and estimation of breeding values and genetic parameters. The reasons for the difference between disciplines are unclear but may have historical foundations. Plant breeding data are derived from field trials that were originally analysed (as far back as the 1930s) using an ANOVA framework where treatment (variety) effects were regarded as fixed and block effects as random. The approach was extended to MET data by regarding environments as blocks. This doctrine remained unchallenged until relatively recently when statisticians began to advocate the use of more general mixed models for MET data. It has therefore required a major culture change for plant breeding programmes to adopt the

more complex models and only a small number have done so. The challenge therefore remains to improve adoption worldwide.

An historical argument against the use of mixed models for plant breeding data was the lack of suitable software. As discussed in the section 'Estimation and software', this is no longer an issue as the tools to fit complex mixed models to large MET data sets are now available.

A further challenge is to encourage the use of random rather than fixed variety effects. This is not an easy task, particularly as this is still a controversial topic among statisticians. As discussed earlier, the present authors believe that variety effects should be assumed to be random since this minimizes selection errors when identifying the best varieties, it provides more realistic predictions of genetic gain and allows a valid analysis of data combined across stages of selection.

REFERENCES

BECKER, R. A., CHAMBERS, J. M. & WILKS, A. R. (1988). *The New S Language*. Pacific Grove, CA: Wadsworth & Brooks/Cole.

BESAG, J. & KEMPTON, R. A. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics* **42**, 231–251.

BUTLER, D. G., CULLIS, B. R., GILMOUR, A. R. & GOGEL, B. J. (2003). *Samm Reference Manual*. Training series, No QE02001. Brisbane, QLD: QLD Department of Primary Industries and Fisheries.

CULLIS, B. R. & GLEESON, A. C. (1991). Spatial analysis of field experiments – an extension to two dimensions. *Biometrics* **47**, 1449–1460.

CULLIS, B. R., GOGEL, B. J., VERBYLA, A. P. & THOMPSON, R. (1998). Spatial analysis of multi-environment early generation trials. *Biometrics* **54**, 1–18.

CULLIS, B. R., SMITH, A. & THOMPSON, R. (2004). Perspectives of ANOVA, REML and a general linear mixed model. In Methods and Models in Statistics, symposium held to honour the 80th birthday of Professor John Nelder FRS (Eds N. M. Adams, M. J. Crowder, D. J. Hand & D. A. Stephens), pp. 53–94. London: Imperial College.

CULLIS, B. R., THOMSON, F. M., FISHER, J. A., GILMOUR, A. R. & THOMPSON, R. (1996 *a*). The analysis of the NSW wheat variety database. I. Modelling trial error variance. *Theoretical and Applied Genetics* **92**, 21–27.

CULLIS, B. R., THOMSON, F. M., FISHER, J. A., GILMOUR, A. R. & THOMPSON, R. (1996 *b*). The analysis of the NSW wheat variety database. II. Variance component estimation. *Theoretical and Applied Genetics* **92**, 28–39.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* **39**, 1–38.

FALCONER, D. S. & MACKAY, T. F. C. (1996). *Introduction to Quantitative Genetics*. 4th edition. London: Longman Scientific and Technical.

FINLAY, K. W. & WILKINSON, G. N. (1963). The analysis of adaptation in a plant breeding programme. *Australian Journal of Agricultural Research* **14**, 742–754.

FINNEY, D. J. (1980). Contribution to the discussion of: Statutory recommended list trials of crop varieties in the United Kingdom. *Journal of the Royal Statistical Society Series A* **143**, 245–246.

FREEMAN, G. H. & PERKINS, J. M. (1971). Environmental and genotype-environmental components of variability. VIII. Relationships between genotypes grown in different environments and measures of these environments. *Heredity* **27**, 15–23.

FRENSHAM, A. B., CULLIS, B. R. & VERBYLA, A. P. (1997). Genotype by environment variance heterogeneity in a two-stage analysis. *Biometrics* **53**, 1373–1383.

FRENSHAM, A. B., BARR, A. R., CULLIS, B. R. & PELHAM, S. D. (1998). A mixed model analysis of 10 years of oat evaluation data: use of agronomic information to explain genotype by environment interaction. *Euphytica* **99**, 43–56.

GAUCH, H. G., JR. (1992). *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs*. Amsterdam: Elsevier.

GENSTAT 5 COMMITTEE (1998). *Genstat 5 Release 4.1 Reference Manual Supplement*. Oxford, UK: Numerical Algorithms Group.

GILMOUR, A. R., THOMPSON, R. & CULLIS, B. R. (1995). AI, an efficient algorithm for REML estimation in linear mixed models. *Biometrics* **51**, 1440–1450.

GILMOUR, A. R., CULLIS, B. R. & VERBYLA, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics* **2**, 269–293.

GILMOUR, A. R., CULLIS, B. R., GOGEL, B., WELHAM, S. J. & THOMPSON, R. (2002). *ASReml User Guide*. Release 1.0. Hemel Hempstead, UK: VSN International Ltd.

GLEESON, A. C. & CULLIS, B. R. (1987). Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics* **43**, 277–288.

GOGEL, B. J., CULLIS, B. R. & VERBYLA, A. P. (1995). REML estimation of multiplicative effects in multi-environment variety trials. *Biometrics* **51**, 744–749.

GREEN, P. J., JENNISON, C. & SEHEULT, A. H. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B* **47**, 299–315.

HARDWICK, R. C. & WOOD, J. T. (1972). Regression methods for studying genotype-environment interactions. *Heredity* **28**, 209–222.

HENDERSON, C. R., KEMPTHORNE, O., SEARLE, S. R. & VON KROSIGK, C. M. (1959). The estimation of environmental

and genetic trends from records subject to culling. *Biometrics* **15**, 192–218.

KEMPTON, R. A. (1984). The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science, Cambridge* **103**, 123–135.

KNIGHT, R. (1970). The measurement and interpretation of genotype-environment interactions. *Euphytica* **19**, 225–235.

LITTELL, R., MILLIKEN, G., STROUP, W. & WOLFINGER, R. (1996). *SAS System for Mixed Models.* Cary, NC: SAS Institute Inc.

MANDEL, J. (1971). A new analysis of variance for non-additive data. *Technometrics* **13**, 1–18.

MARDIA, K. V., KENT, J. T. & BIBBY, J. M. (1988). *Multivariate Analysis.* London: Academic Press.

MARTIN, R. J. (1990). The use of time-series models and methods in the analysis of agricultural field trials. *Communications in Statistics: Theory and Methods* **19**, 55–81.

NABUGOOMU, F., KEMPTON, R. A. & TALBOT, M. (1999). Analysis of series of trials where varieties differ in sensitivity to locations. *Journal of Agricultural, Biological and Environmental Statistics* **4**, 310–325.

NELDER, J. A. (1954). The interpretation of negative components of variance. *Biometrika* **41**, 544–548.

NELDER, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proceedings of the Royal Society, A* **283**, 147–162.

PAPADAKIS, J. S. (1937). Methode statistique pour des experiences sur champ. *Bulletin Scientifique, Institut d'Amelioration des Plantes a Thessaloniki* (Grece), **23**, 12–28.

PATTERSON, H. & NABUGOOMU, F. (1992). REML and the analysis of series of crop variety trials. In *Proceedings from the 16th International Biometric Conference*, pp. 77–93. Hamilton, New Zealand.

PATTERSON, H. D. & SILVEY, V. (1980). Statutory and recommended list trials of crop varieties in the United Kingdom. *Journal of Royal Statistical Society, A* **143**, 219–252.

PATTERSON, H. D., SILVEY, V., TALBOT, M. & WEATHERUP, S. T. C. (1977). Variability of yields of cereal varieties in U.K. trials. *Journal of Agricultural Science, Cambridge* **89**, 239–245.

PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545–554.

PIEPHO, H.-P. (1997). Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics* **53**, 761–767.

PIEPHO, H.-P., DENIS, J.-B. & VAN EEUWIJK, F. A. (1998). Predicting cultivar differences using covariates. *Journal of Agricultural, Biological and Environmental Statistics* **3**, 151–162.

SMITH, A., CULLIS, B., LUCKETT, D., HOLLAMBY, G. & THOMPSON, R. (2002). Exploring variety-environment data using random effects AMMI models with adjustments for spatial field trend. Part II: Applications. In *Quantitative Genetics, Genomics and Plant Breeding* (Ed. M. Kang), pp. 337–352. UK: CABI Publishing.

SMITH, A. B., CULLIS, B. R. & GILMOUR, A. (2001a). The analysis of crop variety evaluation data in Australia. *Australian and New Zealand Journal of Statistics* **43**, 129–145.

SMITH, A. B., CULLIS, B. R. & THOMPSON, R. (2001b). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* **57**, 1138–1147.

SMITH, S. P. & GRASER, K.-U. (1986). Estimating variance components in a class of mixed models by restricted maximum likelihood. *Journal of Dairy Science* **69**, 1156–1165.

TALBOT, M. (1984). Yield variability of crop varieties in the UK. *Journal of Agricultural Science, Cambridge* **102**, 315–321.

THEOBALD, C., TALBOT, M. & NABUGOOMU, F. (2002). A Bayesian approach to regional and local-area prediction from crop variety trials. *Journal of Agricultural, Biological and Environmental Statistics* **7**, 403–419.

THOMPSON, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics* **29**, 527–550.

THOMPSON, R. (1979). Sire evaluation. *Biometrics* **35**, 339–353.

THOMPSON, R., CULLIS, B. R., SMITH, A. B. & GILMOUR, A. R. (2003). A sparse implementation of the Average Information algorithm for factor analytic and reduced rank variance models. *Australian and New Zealand Journal of Statistics* **45**, 445–460.

WILKINSON, G. N., ECKERT, S. R., HANCOCK, T. W. & MAYO, O. (1983). Nearest neighbour (NN) analysis of field experiments (with discussion). *Journal of the Royal Statistical Society, Series B* **45**, 151–211.

YATES, F. & COCHRAN, W. G. (1938). The analysis of groups of experiments. *Journal of Agricultural Science, Cambridge* **28**, 556–580.

ZIMMERMAN, D. L. & HARVILLE, D. A. (1991). A random field approach to the analysis of field plot experiments and other spatial experiments. *Biometrics* **47**, 223–239.

APPENDIX: KRONECKER PRODUCTS

Definition

Let $A = \{a_{ij}\}$ be an $m \times n$ matrix and $B = \{b_{kl}\}$ be a $p \times q$ matrix. Then the Kronecker product of A and B, denoted $A \otimes B$, is given by the $mp \times nq$ matrix

$$
\begin{bmatrix}
a_{11}B & a_{12}B & \cdots & a_{1n}B \\
a_{21}B & a_{22}B & \cdots & a_{2n}B \\
. & . & & . \\
. & . & & . \\
. & . & & . \\
a_{m1}B & a_{m2}B & \cdots & a_{mn}B
\end{bmatrix}
$$

Expansion of variance matrix for $V \times E$ effects

All the models considered in this paper have a variance matrix for the $V \times E$ effects of the form

$$\mathrm{var}\,(\boldsymbol{\eta}) = A \otimes B$$

where A is a $t \times t$ variance matrix for the environment dimension and B is an $m \times m$ variance matrix for the variety dimension. Consider $m = 4$ varieties and $t = 2$ environments. Then

$$
\mathrm{var}
\begin{pmatrix}
\begin{bmatrix}
\eta_{11} \\ \eta_{21} \\ \eta_{31} \\ \eta_{41} \\ \eta_{12} \\ \eta_{22} \\ \eta_{32} \\ \eta_{42}
\end{bmatrix}
\end{pmatrix}
=
\begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix}
\otimes
\begin{bmatrix}
b_{11} & b_{12} & b_{13} & b_{14} \\
b_{21} & b_{22} & b_{23} & b_{24} \\
b_{31} & b_{32} & b_{33} & b_{34} \\
b_{41} & b_{42} & b_{43} & b_{44}
\end{bmatrix}
$$

This means that the variance of the $V \times E$ effect for variety i and environment j is given by

$$\mathrm{var}\,(\eta_{ij}) = a_{jj}b_{ii}$$

and the covariance between effects for variety i, environment j and variety k, environment l is

$$\mathrm{cov}\,(\eta_{ij}, \eta_{kl}) = a_{jl}b_{ik}$$

Expansion of model formulae

Consider the specific example of Eqn (4) and assume $m = 4$ varieties and $t = 2$ environments. Then

$$
\begin{bmatrix}
\eta_{11} \\ \eta_{21} \\ \eta_{31} \\ \eta_{41} \\ \eta_{12} \\ \eta_{22} \\ \eta_{32} \\ \eta_{42}
\end{bmatrix}
=
\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu
+
\left(
\begin{bmatrix} 1 \\ 1 \end{bmatrix}
\otimes
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\right)
\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}
+
\left(
\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\otimes
\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}
\right)
\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}
+
\begin{bmatrix}
\delta_{11} \\ \delta_{21} \\ \delta_{31} \\ \delta_{41} \\ \delta_{12} \\ \delta_{22} \\ \delta_{32} \\ \delta_{42}
\end{bmatrix}
$$

$$
=
\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu
+
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}
+
\begin{bmatrix}
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1
\end{bmatrix}
\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}
+
\begin{bmatrix}
\delta_{11} \\ \delta_{21} \\ \delta_{31} \\ \delta_{41} \\ \delta_{12} \\ \delta_{22} \\ \delta_{32} \\ \delta_{42}
\end{bmatrix}
$$