Aust. N. Z. J. Stat. **43**(2), 2001, 129–145

THE ANALYSIS OF CROP VARIETY EVALUATION DATA IN AUSTRALIA

ALISON SMITH[1]*, BRIAN CULLIS[1] AND ARTHUR GILMOUR[2]

Wagga Wagga Agricultural Institute and Orange Agricultural Institute

Summary

The major aim of crop variety evaluation is to predict the future performance of varieties. This paper presents the routine statistical analysis of data from late-stage testing of crop varieties in Australia. It uses a two-stage approach for analysis. The data from individual trials from the current year are analysed using spatial techniques. The resultant table of variety-by-trial means is combined with tables from previous years to form the data for an overall mixed model analysis. Weights allow for the data being estimates with varying accuracy. In view of the predictive aim of the analysis, variety effects and interactions are regarded as random effects. Appropriate inferential tools have been developed to assist with interpretation of the results. Analyses must be conducted in a timely manner so that variety predictions can be published and disseminated to growers immediately after harvest each year. Factors which facilitate this include easy access to historic data and the use of specialist mixed model software.

Key words: crop variety evaluation; mixed model; residual maximum likelihood; BLUP.

1. Introduction

Decisions about the commercial value of new crop varieties are based on data from a series of comparative trials, known as multi-environment trials (METs), conducted in a range of geographic locations and over several years. In Australia, state-based crop variety evaluation programs (CVEPs), most of which are run by State Departments of Agriculture, carry out variety evaluation. Although many of the varieties tested are unique to individual programs there is some overlap. Therefore the results from individual programs are of national interest. The analysis, reporting and interpretation of results is substantially the responsibility of the host organization. Most CVEPs report varietal yields expressed as a percentage of a standard commercial variety. Historically, these figures were based on pairs of means from those trials in which the variety and standard were grown together. A t-test was used to assess the significance of the difference. Despite the wealth of literature on the analysis of MET data, including a substantial Australian contribution (see e.g. Finlay & Wilkinson, 1963; Brennan & Byth, 1979; Pederson & Rathjen, 1981; Kroonenberg & Basford, 1989; Cooper & de Lacy, 1994; Cullis *et al.*, 1996a,b) the simplistic pairwise comparison approach was maintained for the routine analysis of CVEP data well into the 1990s. In the mid 1990s several programs moved to adopt more modern approaches to analysis. The key was the use of mixed model

Received October 1999; revised June 2000; accepted July 2000.

* Author to whom correspondence should be addressed.

[1] Wagga Wagga Agricultural Institute, Wagga Wagga, NSW 2650. e-mail: alison.smith@agric.nsw.gov.au

[2] Orange Agricultural Institute, Orange, NSW 2340.

Acknowledgments. The authors thank Colleen Hunt, Patrick Lim, Katia Stefanova, Dave Butler, Alison Kelly, Fiona Thomson and Ben Braysher for their assistance with the national statistical support project. They thank Rob Wheeler, Dean Diepeveen and Richard Gammie for support and advice, Robin Thompson for stimulating exchanges, and a referee for insightful comments that have improved the manuscript. The authors acknowledge the vision of the late Mike Carroll for supporting a co-ordinated national system for the evaluation of crop varieties in Australia, and the financial support of the Grains Research and Development Corporation.

technology which allowed the analysis of incomplete tables of variety-by-trial data. Although this is a well accepted technique among statisticians it caused some confusion among plant breeders and crop evaluators. The most basic concern was the fact that the data used for variety comparisons were not restricted to those trials in which the varieties were grown together. Other issues requiring clarification included assumptions about variety effects, namely whether they should be regarded as fixed or random effects, and the definition of the dataset, namely whether it should include all varieties in the testing system or merely the small subset whose results were to be reported to growers. These issues highlighted the need to establish protocols for data analysis and the reporting and interpretation of results. Only then could there be a co-ordinated, rigorous approach to the analysis of crop variety evaluation data in Australia.

Public plant breeding and variety evaluation programs in Australia receive significant financial support from the Grains Research and Development Corporation (GRDC). The GRDC has been concerned about inadequate co-ordination between CVEPs for some time. They initiated a sequence of reviews (Lazenby, 1986; Clements, Rosielle & Hilton, 1992; Lazenby *et al.*, 1994) which led to the establishment of an Australian Crop Accreditation System (ACAS). As outlined in Leslie *et al.* (1997) the objective of the system is to '. . . provide sound information on the performance of grain crop varieties, so that users of grains can make rational selections from the varieties available to them'. A key component of ACAS was the development of protocols for the evaluation of crop varieties. As part of this initiative, in 1996 the GRDC funded a project to co-ordinate statistical practices in CVEPs in Australia. This has improved the consistency of trial design, methods of analysis and reporting of results. Patterson & Silvey (1980) presented a comprehensive account of the statistical methods used in the United Kingdom variety evaluation system. Our paper has a similar objective, namely to describe current methods and suggest some future directions for the analysis of crop variety evaluation data in Australia.

The paper is arranged as follows. Section 2 presents an overview of the structure of CVEPs in Australia, outlining their objectives, system dynamics and trial design protocol. Section 3 describes the statistical analysis of data from CVEPs, and includes the current analytical methods for MET datasets and a discussion of future directions. The analysis of the MET dataset for wheat from the South Australian CVEP illustrates the methods. We briefly outline the analysis of individual trials and the derivation of the weights required for the analysis of the MET datasets, and conclude by discussing some key aspects of the approach.

2. Australian crop variety evaluation programs

2.1. Objectives

The major objective of CVEPs is to provide reliable information about the performance of varieties relative to existing 'standard' commercial varieties. The aim is to accurately identify the varieties that can provide significant commercial gains, either on average across a set of target environments or in particular environments. Variety performance is measured in terms of a number of traits, including yield, quality characteristics and disease resistance. In this paper we focus on the analysis of yield data.

2.2. Program dynamics

The CVEPs are concerned with the widespread testing of potential new varieties. Each potential variety in a program is termed an 'entry' in that program. Most entries originate from

breeding programs, both public and private, and have progressed through numerous stages of selection.

Many CVEPs conduct two sequential stages of testing, known as S3 and S4. Entries accepted by CVEPs for testing are usually first grown in S3 trials which are conducted at a small number of locations (less than 10) across the set of target environments. They comprise two or three replicates of a large number of entries (up to 150). Entries are tested in S3 trials for a single year only, the best being promoted to the final (S4) stage of testing.

The main focus of this paper is S4 testing which involves the evaluation of a small number of elite entries (usually between 20 and 40) at a large number of geographic locations and possibly over several years. The S4 testing programs encompassed in the GRDC statistical support project (Table 1) include all major public CVEPs in Australia. Note that for some crops, testing is divided according to maturity classification. The New South Wales (NSW) CVEP conducts trials for early-sown and main-season-sown barley and wheat. The Queensland (Qld) CVEP divides the testing of wheat into trials for quick, intermediate and slow maturing varieties. A complete set of three trials is usually conducted at each location.

The number of trials varies according to the commercial significance of the crop and the magnitude of the growing area, ranging from an average of six or seven trials per year for the testing of lentils in South Australia (SA) and NSW to more than 70 per year for the testing of wheat in Western Australia (WA), NSW and Qld and barley in NSW (Table 1, aggregated over maturity series where relevant). The number of trials per year and the entry retention rates (to be discussed later) in Table 1 were based on the past six years of testing for most CVEPs. Due to funding pressures the numbers of trials per year have been declining and this trend is likely to continue.

Trial sites are chosen to be representative of target environments. Site selection is often stratified on the basis of pre-defined geographic or rainfall regions (see Table 1). This reflects the belief that variety-by-region interactions are important so that results should be presented on a regional basis. Furthermore, some varieties may be targeted for specific regions so may be grown in only a subset of the trials.

Trials are located both on experimental research stations and on growers' properties. In SA, Vic and Qld, trial sites are almost invariant from one year to the next whereas in other programs the sites may be moved up to 50 km. This has implications for data analysis.

Due to the cost of variety evaluation there is a limit on the total number of entries in a trial. As such, new entries can only be included if some previously tested entries are discarded. Thus in any one year the trials comprise a mixture of 'standard' commercial varieties (which remain in the testing scheme for many years and provide the basis for comparison), entries which were tested in the previous one or two years and entries being tested for the first time. The length of retention of entries varies between CVEPs. Most entries tested in WA, Qld and Vic programs are only tested for a single year (Table 1). Cullis *et al.* (2000) show that this has an adverse affect on the response to selection. Entries are not generally tested for more than three years before a decision is made regarding commercial release.

2.3. Individual trial design

Historically, trials have been designed using randomized complete block (RCB) designs, resolvable incomplete block (IB) designs and neighbour balanced designs. Currently, most programs use designs from the software package SpaDes (Coombes, 1999). These designs in-

TABLE 1

Description of a range of Australian CVEP*s*

Crop	Trials per year	Regions	% entries tested 1 year	2 years	≥ 3 years
New South Wales					
Wheat (early sown)	27	4	27	19	44
Wheat (main season sown)	50	4	46	13	41
Barley (early sown)	21	6	39	30	31
Barley (main season sown)	56	6	43	28	29
Lupins	14	2	62	19	19
Peas	11	2	22	27	51
South Australia					
Wheat	24	6	45	21	34
Barley	19	6	28	30	43
Oats	14	6	52	20	29
Lentils	6	4	7	21	71
Peas	10	4	34	12	54
Canola	14	6	64	23	13
Lupins	14	5	18	26	56
Victoria					
Wheat	29	7	57	17	26
Barley	19	4	64	21	15
Lentils	7	2	48	21	30
Peas	13	5	41	17	42
Western Australia					
Wheat	71	19	52	20	28
Barley	48	20	69	15	17
Oats	65	18	51	25	23
Lupins	52	21	68	15	18
Peas	35	21	52	24	23
Queensland					
Wheat (quick maturing)	25	–[A]	74	8	18
Wheat (intermediate maturing)	25	–[A]	73	17	10
Wheat (slow maturing)	25	–[A]	59	21	20
Barley	19	–[A]	63	16	21

[A] Queensland CVEPs do not present results on a regional basis

corporate partial neighbour balance, edge and corner plot balance (Martin & Eccleston, 1991), are binary in rows and columns, and where possible allow for replication in two directions. The CVEPs in Qld use alpha-latinized row–column designs (John & Williams, 1998). Most trials are laid out in the field as a single rectangular array of plots, indexed generically by rows and columns. Varieties are usually equally replicated with three or four replicates.

3. Analysis of CVEP data

Analyses of CVEP data are conducted to:

I. make decisions about discarding entries from the testing system (to make way for new entries) and about recommendations of entries for commercial release, and
II. provide farmers with reliable predictions about varieties currently available to them (including recently released varieties and older standard varieties).

We refer to these two types of information as Type I and Type II. Until recently, a single analysis has been used to obtain both types of information. The analysis was based on a MET dataset that included sufficient years to provide a representative sample of seasons (usually

between 5 and 10 years). This is appropriate for obtaining Type II information. We now recommend that the data used for obtaining Type I information be restricted to trials that are directly relevant to the cohort of entries under consideration. This is likely to involve all S3 and S4 trials from the current year and previous one or two years.

The MET datasets are highly unbalanced because entries are constantly moving through the system (Table 1). This necessitates the use of mixed model methods (see e.g. Patterson & Silvey, 1980; Patterson *et al.*, 1977). Cullis *et al.* (1998) present an efficient approach for the analysis of MET data which involves the modelling of individual plot data. Their mixed model accommodates error variance heterogeneity between trials and spatial heterogeneity within trials. Smith *et al.* (1998) extend this approach to allow for multiplicative modelling of variety-by-environment interaction, leading to a realistic form for the genetic variance structure. This analysis provides relevant estimates of overall variety performance and detailed and interpretable information regarding variety-by-environment interaction.

The Smith *et al.* (1998) approach is widely used for the analysis of early generation plant breeding trials in Australia. Due to the importance of retention and recommendation decisions we recommend this approach for Type I CVEP data, but have not, as yet, adopted it for Type II data. There are two major reasons for this. First, there are computational issues associated with fitting complex mixed models to MET datasets with large numbers (greater than 100) of trials, but work is in progress and we expect a resolution in the near future. Second, appropriate yield data were not always available prior to commencement of the GRDC project. In general, individual plot data have not been stored electronically. Variety mean yields from the analysis of individual trials have been more readily available. Efforts are now being made to ensure that all relevant data are stored.

These problems have necessitated the use of a two-stage approach in which variety means obtained from the analysis of individual trials are combined to form the data for an overall mixed model analysis. This approach has been used for the analysis of MET data (see Patterson & Silvey, 1980; Talbot, 1984; Patterson & Nabugoomu, 1992; Cullis *et al.*, 1996a,b). This paper presents several key enhancements (see Sections 3.3 and 3.4) aimed at more closely approximating the efficient one-stage analysis of Cullis *et al.* (1998).

3.1. Combined analysis of MET data

3.1.1. Description of data

The dataset for each crop comprises mean yields from all relevant S4 trials from a number of years before and including the current year. All entries tested in this time span are included in the dataset unless they were grown in fewer than a minimum number of trials deemed appropriate for the crop. In terms of the time span we are mindful of the trade-off between obtaining a representative sample of seasons, which requires a lengthy time span, and maintaining a reasonable level of connectivity and a representative sample of varieties, both of which require a shorter time span. These issues have not been completely resolved. In the United Kingdom reporting is based on about five years of data (Silvey, 1978).

All trials are included in the dataset unless known to be unrepresentative. For example, if the trial mean yield is less than a pre-determined threshold the trial is excluded. Coefficients of variation (CV) from the analysis of individual trials are not used as an exclusion criterion. Cullis *et al.* (1996a) showed that the CV is an inappropriate measure for comparing the reliability of trials since it does not stabilize the relationship between mean and variance. Instead

it over-corrects, making low-yielding trials appear relatively more variable. Heterogeneity of error variance can be properly accommodated by using an appropriate weighting scheme in the combined analysis (see Section 3.4).

3.1.2. Mixed model analysis

The combined analysis of mean yields is based on a mixed model that accounts for all sources of variation: variation associated with variety (V) effects, variety-by-environment (V.E) interaction and within-trial plot error. The V.E interaction is partitioned in a manner appropriate to the dataset. This is partly dependent on whether there are regions and whether trial locations are invariant across years (this is the case in SA, Vic and Qld programs). Variety covariates (and their interactions with environment) are included in the analysis where necessary. For example, in the testing of oats in SA, three different plant types (varieties with and without a semi-dwarf gene and hulless varieties) are grown together in the same trials although recommendations are only relevant within each type. There are large differences in mean yield between the types and large type-by-environment interactions (Frensham *et al.*, 1997a), so a 'type' covariate and associated interactions are included in the analysis.

Table 2 shows the terms included in the mixed model (and their classification as fixed or random) for an example dataset, namely SA wheat. Location effects (and their interactions) are partitioned into regions and locations within regions, the latter represented by the terms 'region.loc', etc.

The mixed model represented in Table 2 is comparable to that employed by Patterson & Silvey (1980), the major difference being that we regard variety effects as random. We choose to do this because it provides more reliable estimates of variety performance and addresses the problem of selection bias. As Patterson & Silvey (1980) themselves point out:

> Selection bias arises because a variety is more likely to be recommended if its trial mean yield exceeds its true mean yield. The bias can be roughly estimated by regressing true means of the varieties on their means in the trials (Finney, 1964). Regression coefficients are calculated indirectly from varieties and varieties by environments components of variance.

This regression adjustment is essentially what transpires in best linear unbiased prediction (BLUP) which is the standard method for estimation of random effects (see Section 3.1.3). Thus the indirect adjustment proposed, though not used, by Patterson & Silvey (1980) can be incorporated in the analysis simply by regarding variety effects as random. In the example they give, the regression adjustment shows that '... differences between trial means for newly recommended varieties, are, on the average, about 27% too large'. This demonstrates the deficiency in the traditional fixed variety-effects approach in terms of obtaining reliable predictions of future yield performance.

The data, $\boldsymbol{y}^{(n\times 1)}$, for the combined analysis, are the vector of variety mean yields from the analyses of individual trials. If $\boldsymbol{y}_j{}^{(n_j\times 1)}$ is the vector of mean yields from the jth trial and there are m trials, then $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)$ and the total number of data points is $n = \sum_{j=1}^{m} n_j$. The linear mixed model for the analysis of these data is given by

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{e} = \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{\xi} + \boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{\tau}^{(t\times 1)}$ and $\boldsymbol{u}^{(b\times 1)}$ are vectors of fixed and random effects respectively with associated design matrices $\boldsymbol{X}^{(n\times t)}$ and $\boldsymbol{Z}^{(n\times b)}$ (the former assumed to be of full column rank). The residuals, $\boldsymbol{e}^{(n\times 1)} = \boldsymbol{\xi} + \boldsymbol{\eta}$ are the sums of two components. The term $\boldsymbol{\xi}$ represents that component of V.E interaction which completely indexes the data. This depends on the partitioning of V.E interaction. The term $\boldsymbol{\eta}$ is included to account for the data being estimates and

TABLE 2

Terms in the mixed model for SA wheat data

Term	Decomposition	Type	Effects
Variety			
	variety	R	$\boldsymbol{u}_1$
Environment			
	year [A]	F	
	region [A]	F	
	region.loc [A]	F	
	year.region [A]	F	
	year.region.loc [A]	F	
Variety.Environment			
	variety.year	R	$\boldsymbol{u}_{12}$
	variety.region	R	$\boldsymbol{u}_{13}$
	variety.region.loc	R	$\boldsymbol{u}_{134}$
	variety.year.region	R	$\boldsymbol{u}_{123}$
	variety.year.region.loc	R	$\boldsymbol{\xi}$
Error			
	error	R	η

[A] For computational efficiency these effects are not fitted but are replaced by a single factor that indexes environments. This produces equivalent estimates of variance components.

therefore subject to uncertainty. It reflects within-trial plot error variation. The inclusion of this component is crucial; otherwise the residual variance can be misinterpreted as purely V.E interaction variance.

For the SA wheat data the mixed model is given by

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{\xi} + \boldsymbol{\eta} \\ &= \boldsymbol{X\tau} + \boldsymbol{Z}_1\boldsymbol{u}_1 + \boldsymbol{Z}_{12}\boldsymbol{u}_{12} + \boldsymbol{Z}_{13}\boldsymbol{u}_{13} + \boldsymbol{Z}_{134}\boldsymbol{u}_{134} + \boldsymbol{Z}_{123}\boldsymbol{u}_{123} + \boldsymbol{\xi} + \boldsymbol{\eta},\end{aligned}$$

where the random effects $\boldsymbol{u}_i$ are sub-vectors of $\boldsymbol{u}$ and are as defined in Table 2. The term $\boldsymbol{\xi}$ represents 'variety.year.region.loc' interaction.

It is assumed that the joint distribution of $(\boldsymbol{u}, \boldsymbol{\xi}, \boldsymbol{\eta})$ is normal with zero mean and variance matrix

$$\sigma^2 \begin{bmatrix} \boldsymbol{G} & 0 & 0 \\ 0 & \boldsymbol{\Omega} & 0 \\ 0 & 0 & \boldsymbol{\Sigma} \end{bmatrix}.$$

The matrix $\sigma^2\boldsymbol{\Sigma}$ is the conditional variance matrix of $\boldsymbol{y}$ given $(\boldsymbol{u}, \boldsymbol{\xi})$. It represents the uncertainty due to the data being estimates. The matrix is block diagonal with blocks given by $\boldsymbol{\Sigma}_j$. In practice we replace σ^2 and $\boldsymbol{\Sigma}_j$ by estimates $\tilde{\sigma}^2$ and $\tilde{\boldsymbol{\Sigma}}_j$ obtained from the analyses of individual trials (see Section 3.3).

The matrices $\boldsymbol{G}$ and $\boldsymbol{\Omega}$ are functions of vectors of unknown parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$. They may be completely general, although in the standard two-stage MET analysis they are usually diagonal with the parameters being variance component ratios. For example, one block of $\boldsymbol{G}$ is given by $\gamma_1\boldsymbol{I}$ where $\tilde{\sigma}^2\gamma_1$ is the variance component for variety effects. It is also usual to assume independence and constant variance for all V.E interaction terms. Frensham, Cullis & Verbyla (1997b) present a method for accommodating heterogeneity of residual V.E variance in a two-stage MET (i.e. heterogeneity in $\boldsymbol{\Omega}$) but we have not adopted this in the routine analysis of CVEP data.

The distribution of the data $\boldsymbol{y}$ is thus normal with mean $\boldsymbol{X\tau}$ and variance matrix

$$\mathrm{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{H} = \sigma^2(\boldsymbol{ZGZ}^\top + \boldsymbol{R}), \qquad \text{where } \boldsymbol{R} = \boldsymbol{\Omega} + \tilde{\boldsymbol{\Sigma}}.$$

3.1.3. Estimation

Estimation of the fixed and random effects in the mixed model (1) proceeds using the mixed model equations:

$$\begin{bmatrix} \boldsymbol{X}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X} & \boldsymbol{X}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z} & \boldsymbol{X}^\top\tilde{\boldsymbol{\Sigma}}^{-1} \\ \boldsymbol{Z}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X} & \boldsymbol{Z}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} & \boldsymbol{Z}^\top\tilde{\boldsymbol{\Sigma}}^{-1} \\ \tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X} & \tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z} & \tilde{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Omega}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}} \\ \tilde{\boldsymbol{u}} \\ \tilde{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y} \\ \tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y} \end{bmatrix}. \tag{2}$$

The solution of (2) requires values for the variance parameters associated with $\boldsymbol{G}$ and $\boldsymbol{\Omega}$; in practice, residual maximum likelihood (REML) (Patterson & Thompson, 1971) estimates are substituted. The matrix $\tilde{\boldsymbol{\Sigma}}$ is known.

This leads to best linear unbiased estimates (BLUEs) of the fixed effects:

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

and BLUPs of the random effects:

$$\tilde{\boldsymbol{u}} = \boldsymbol{GZ}^\top\boldsymbol{Py} \quad \text{and} \quad \tilde{\boldsymbol{\xi}} = \boldsymbol{\Omega Py}, \quad \text{where } \boldsymbol{P} = \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{H}^{-1}.$$

The most efficient estimation of effects and variance parameters would be achieved by including the full (known) matrix $\tilde{\boldsymbol{\Sigma}}$ in the mixed model equations. However, it is computationally difficult both to store and then to use the full constituent matrices, $\tilde{\boldsymbol{\Sigma}}_j$, from each trial. The standard linear models approach for accommodating known error variance heterogeneity is to use a vector of weights. We have adopted this approach, so we replace $\tilde{\boldsymbol{\Sigma}}^{-1}$ by a diagonal matrix, $\boldsymbol{\Pi}$, of weights. If orthogonal analyses such as RCB are used for the analysis of individual trials then $\tilde{\boldsymbol{\Sigma}}^{-1}$ is truly diagonal, whereas if the more efficient but non-orthogonal spatial analyses (or IB) are used, the diagonal form is an approximation to the complete structure. A practical and reasonable approximation can be achieved with an appropriate choice of weights (see Section 3.4).

3.1.4. Computing

The mixed model analysis of CVEP data is computationally intensive. The datasets often involve more than 5000 records and the analysis may require the solution of more than 40 000 equations. We are unaware of any standard commercially available software that can handle this type of analysis. Only since the development of the specialist mixed model program ASREML (Gilmour *et al.*, 1999) has it been possible to conduct these analyses. The ASREML program employs sparse matrix methods and uses the Average Information algorithm (Gilmour, Thompson & Cullis, 1995) for REML estimation of variance parameters. Both these features contribute to the efficiency of ASREML. One of the reasons for developing the Average Information algorithm was to estimate variance components in a large unbalanced MET dataset (Gilmour *et al.*, 1995). The authors showed that the algorithm was much faster than existing algorithms for this type of problem.

As discussed in Section 3.1.2 the mixed model for CVEP data requires inclusion of a known component of residual variance. The ASREML program has the facility to do this (the relevant code is available on request). Note that Patterson & Nabugoomu (1992) attempted this using GENSTAT (Genstat 5 Committee, 1998) but the method they presented for REML estimation was incorrect. We note, however, that the core routines of ASREML are now accessible through the REML directive of GENSTAT.

TABLE 3

Summary of trial information for SA wheat data

	Min	Max	Mean	Median
Mean yield (t/ha)	0.24	5.82	2.28	2.09
Error variance $(t/ha)^2$	0.0012	0.1436	0.0257	0.0169
CV (%)	2.6	28.6	7.3	6.4

TABLE 4

SA wheat data: numbers of varieties common across years (diagonal entries are numbers for individual years)

Year	1992	1993	1994	1995	1996	1997	1998
1992	32						
1993	21	31					
1994	18	20	37				
1995	13	15	23	36			
1996	11	11	19	24	38		
1997	11	12	14	15	21	34	
1998	10	10	12	13	15	23	36

3.1.5. Reporting and interpretation of results

Combined analyses of CVEP data are conducted annually. Publications produced by the individual organizations report the results. There is now some consistency in the method of reporting, the basic information comprising predicted variety means from the mixed model analysis. Variety predictions for the complete set of target environments are based on the BLUPs of the variety main effects, while predictions at a regional level are based on the sum of variety main effect and variety-by-region interaction BLUPs.

Means are supplemented with the number of trials in which the varieties were tested. This provides a crude measure of reliability for the prediction. Issues concerning confidence statements and inference for random effects must be carefully considered. Formerly, the information reported by Australian CVEPs was analogous to fixed effect variety means with a least significant difference (LSD) for comparison with one or more standard commercial varieties. This parallels the reporting in the UK system (Silvey, 1978). Varieties with a mean yield significantly greater than certain standards were considered for commercial release.

Inferential tools such as LSDs are inappropriate in the random effects setting. An alternative is to consider selection theory. For example, the commercial release decisions can be phrased in terms of the probability that a variety is truly superior if its BLUP of yield is greater than that of a standard. These probabilities are based on the correlation between the true variety effects and those predicted from the mixed model analysis (see Cullis *et al.*, 2000). Published CVEP results now include the variety means supplemented with the probability of superiority over a pre-determined commercial standard (see Section 3.2).

3.2. Example of a combined analysis

In this section we consider the analysis of data from the SA wheat program from 1992–1998. The dataset consisted of 6028 mean yields from 174 trials. The data summary (Table 3) shows a large range in trial mean yield which is typical in Australian MET data. There was a 120-fold range in trial error variance, reinforcing the need to accommodate heterogeneity.

TABLE 5

SA wheat data: variance components

Source of variation	Component (t/ha)^2	% of total
Variety		
variety	0.01348	17.4
Variety.Environment		
variety.year	0.00601	7.8
variety.region	0.00374	4.8
variety.region.loc	0.00183	2.4
variety.year.region	0.00706	9.1
variety.year.region.loc	0.01908	24.7
Error		
error	0.02618	33.8

TABLE 6

SA wheat data: predicted means for six regions and overall state means

	Mean yield (t/ha)							Number of trials						
Entry/Region[A]	MN	LEP	MM	UEP	YP	SE	State	MN	LEP	MM	UEP	YP	SE	State
Worrakatta	3.36	3.29	2.23	1.53	3.55	3.62	2.92	15	12	24	32	12	7	102
Krichauff	3.28	3.25	2.17	1.50	3.41	3.52	2.85	15	12	24	32	12	7	102
Excalibur	3.29	3.19	2.17	1.49	3.38	3.54	2.84	27	21	41	55	21	9	174
Rac655	3.29	3.10	2.19	1.48	3.33	3.48	2.81	19	15	30	40	15	8	127
Brookton	3.27	3.18	2.11	1.43	3.30	3.45	2.79	7	6	12	16	6	4	51
VI184	3.16	3.10	2.14	1.45	3.40	3.48	2.78	7	6	12	16	6	5	52
WI96080	3.20	3.15	2.12	1.42	3.31	3.51	2.78	7	6	12	16	6	5	52
Frame	3.25	3.10	2.11	1.38	3.36	3.49	2.78	23	18	35	47	18	8	149
Janz	3.30	3.07	2.10	1.38	3.30	3.47	2.77	27	21	41	55	21	9	174
WI96091	3.15	3.12	2.06	1.41	3.27	3.52	2.75	7	6	12	16	6	5	52
Silverstar	3.23	3.11	2.04	1.39	3.27	3.44	2.75	15	12	24	32	12	7	102
WI96112	3.17	3.09	2.08	1.41	3.31	3.41	2.74	7	6	12	16	6	5	52
Carnamah	3.18	3.08	2.05	1.40	3.25	3.36	2.72	7	6	12	16	6	5	52
WI96114	3.15	3.07	2.03	1.38	3.23	3.42	2.71	7	6	12	16	6	5	52
Spear	3.11	3.06	2.06	1.35	3.25	3.42	2.71	27	21	41	55	21	9	174
Tatiara	3.11	3.05	2.02	1.34	3.25	3.44	2.70	27	21	41	55	21	9	174
Diamondbird	3.03	2.94	1.96	1.30	3.18	3.36	2.63	11	9	18	24	9	6	77
Barunga	3.09	2.95	1.95	1.31	3.13	3.31	2.63	27	21	41	55	21	9	174
Machete	3.07	2.93	1.97	1.29	3.07	3.29	2.61	28	21	41	55	21	9	175
Schomburgk	2.92	2.96	1.90	1.26	3.11	3.29	2.58	15	12	23	31	12	6	99
Goldmark	2.99	2.88	1.94	1.28	3.05	3.26	2.57	11	9	18	24	9	6	77
Halberd	2.92	2.87	2.01	1.28	3.09	3.21	2.57	27	21	41	55	21	9	174
Molineux	2.97	2.84	1.86	1.25	3.03	3.23	2.54	27	21	41	55	21	9	174

[A] Region codes: MN = Mid North, LEP = Lower Eyre Peninsula, MM = Murray Mallee, UEP = Upper Eyre Peninsula, YP = Yorke Peninsula, SE = South East

A total of 104 varieties was included in the analysis. In the SA CVEP all entries are grown in all trials in any one year. Thus the only source of imbalance arises between years. The degree of commonality between individual years is shown in Table 4. Only 65% of entries are common in successive years. Once again this reflects the relatively high rejection rate after a single year of testing (also see Table 1). Only the standard entries are common to trials separated by more than three years (approximately 30% of the entries in any trial).

The estimated variance components for the SA wheat data (Table 5) reveal a pattern that is common to most Australian CVEP data, namely that the total variation in the data is

TABLE 7

SA wheat data: probability of superiority of entries over Spear

Entry	Mean yield (t/ha)	Probability
Worrakatta	2.92	0.999
Krichauff	2.85	0.986
Excalibur	2.84	0.986
RAC655	2.81	0.948
Brookton	2.79	0.855
VI184	2.78	0.854
WI96080	2.78	0.850
Frame	2.78	0.880
Janz	2.77	0.850
WI96091	2.75	0.729
Silverstar	2.75	0.731
WI96112	2.74	0.678
Carnamah	2.72	0.563
WI96114	2.71	0.532
Spear	2.71	
Tatiara	2.70	0.451
Diamondbird	2.63	0.131
Barunga	2.63	0.081
Machete	2.61	0.043
Schomburgk	2.58	0.022
Goldmark	2.57	0.022
Halberd	2.57	0.009
Molineux	2.54	0.002

dominated by within-trial plot error (33.8%) and the highest order V.E interactions, in this case variety.year.region.loc (24.7%). Interactions linked to seasonal variation (variety.year, variety.year.region and variety.year.region.loc) are much larger than 'static' interaction (variety.region and variety.region.loc). (See Cullis *et al.*, 2000 for a comprehensive discussion of sources of variation in Australian CVEPs.)

An extract of the information reported by the South Australian CVEP (Wheeler & McMurray, 1999) from this analysis is presented in Table 6. We have selected the subset of entries that have been tested in both 1997 and 1998 because this covers the commercial standards and those entries being considered for commercial release. Table 4 shows that there are 23 such entries. They have been ranked in order of their overall predicted means. Note that the standard used as the basis for comparison, Spear, is ranked 15 in this set.

The probability that entries are superior to Spear is presented in Table 7. These probabilities were derived from the conditional distribution of $\boldsymbol{u}$ given $\tilde{\boldsymbol{u}}$ (Cullis *et al.*, 2000). The probability for the top four ranked varieties exceeds 0.90.

3.3. Analysis of individual trials

Various methods have been used for the analysis of individual yield testing trials, including RCB, IB and spatial analysis. Currently, all programs listed in Table 1 use the spatial approach of Gilmour, Cullis & Verbyla (1997). This facilitates appropriate choices for plot error variance structures.

Consider the analysis of the jth trial in a MET dataset. In the mixed model for these data (given below) the data vector, model terms and variance matrices should have a suffix of j. We have omitted this for simplicity. Thus we denote the vector of data (ordered as rows

within columns) for the jth trial by $\boldsymbol{y}_*^{(N\times 1)}$. The mixed model is given by:

$$\boldsymbol{y}_* = \boldsymbol{A}\boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{\zeta} + \boldsymbol{\epsilon} = \boldsymbol{A}\boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{e}_*,$$

where $\boldsymbol{\alpha}^{(t\times 1)}$ and $\boldsymbol{\beta}^{(b\times 1)}$ represent fixed and random effects respectively and have associated design matrices $\boldsymbol{A}^{(N\times t)}$ and $\boldsymbol{B}^{(N\times b)}$, the former assumed to be of full column rank. The residuals $\boldsymbol{e}_*^{(N\times 1)}$ are comprised of a vector of random local spatial trend effects, $\boldsymbol{\zeta}^{(N\times 1)}$ and a white noise component (also called 'measurement error') $\boldsymbol{\epsilon}^{(N\times 1)}$. The latter may be omitted.

It is assumed that the joint distribution of $(\boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon})$ is normal with zero mean and variance matrix

$$\sigma_s^2 \begin{bmatrix} \boldsymbol{G}_* & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r & 0 \\ 0 & 0 & \psi \boldsymbol{I} \end{bmatrix}.$$

The distribution of the data $\boldsymbol{y}_*$ is thus normal with mean $\boldsymbol{A}\boldsymbol{\alpha}$ and variance matrix

$$\mathrm{V}(\boldsymbol{y}_*) = \sigma_s^2 \boldsymbol{H}_* = \sigma_s^2 (\boldsymbol{B}\boldsymbol{G}_*\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + \psi \boldsymbol{I}).$$

The (two-dimensional) spatial trend process, $\boldsymbol{\zeta}$, is assumed to have a separable variance structure. The matrices $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Sigma}_r$ are the correlation matrices for the column and row trend respectively and often correspond to autoregressive processes of order 1. The trend process is assumed to be stationary. Investigation of the adequacy of this correlation model and the detection of extraneous variation (for example, linked to agronomic practices such as serpentine harvesting) are made possible through the use of diagnostic tools such as the sample variogram (see Gilmour *et al.*, 1997).

Covariate information is used wherever possible. An important source of error is variation in plot length within a trial. Plot lengths should be recorded and used as a covariate, instead of adjusting yields before analysis.

The matrix $\boldsymbol{G}_*$ is a function of a vector of unknown parameters $\boldsymbol{\gamma}_*$. There may be several random terms and the effects are usually assumed to be independent both between and within terms. Thus $\boldsymbol{G}_*$ is diagonal with the parameters being variance component ratios. For example, if variety effects are assumed to be random then one block of $\boldsymbol{G}_*$ is given by $\gamma_v \boldsymbol{I}$ where $\sigma_s^2 \gamma_v$ is the genetic variance for the trial.

In the spatial mixed model, variety effects can be regarded as fixed or random. For the analysis of individual late stage yield testing trials variety effects have traditionally been taken as fixed (see e.g. Patterson *et al.*, 1977; Patterson & Silvey, 1980) but we analyse them as random. There are two main reasons for this. First, it avoids the inconsistency which arises when variety effects are regarded as fixed for individual trials but random for the combined analysis (see Section 3.1.2). More importantly, however, the assumption of random variety effects means that the information used to estimate spatial parameters is comparable to the one-stage spatial MET analysis. In fact, the starting point for spatial modelling in the one-stage analysis is a model in which the genetic effects in different trials are independent (Gilmour *et al.*, 1998). This is analogous to conducting a separate analysis for each trial with the assumption of random variety effects.

Analysing individual trials with random variety effects leads to BLUPs of effects which are contracted versions of the BLUEs from the fixed effects case. The degree of shrinkage depends largely on the ratio of genetic to total variance, which varies from trial to trial. Variety BLUPs from separate analyses are thus scaled differently and are not comparable across trials.

Thus the BLUPs cannot be used as data in the second stage analysis — we must 'unshrink' them first. This can be achieved using a two-stage process. First, each trial is analysed with the assumption of random variety effects. An appropriate spatial model is determined and variance parameters are estimated. This model is then re-fitted with the assumption of fixed variety effects and fixing the variance matrix $\sigma_s^2 \boldsymbol{H}_*$ at the value estimated in the first analysis. (Note that $\boldsymbol{H}_*$ excludes that part which corresponded to the random variety effects.) The resultant BLUEs of variety effects are based on more efficient estimates of spatial parameters (compared to those estimated under the assumption of fixed variety effects) and, unlike the BLUPs, they are comparable across trials. They are used to form the vector of variety mean yields, denoted $\boldsymbol{y}_j^{(n_j \times 1)}$, for the combined analysis. That is,

$$\boldsymbol{y}_j = \boldsymbol{T}\hat{\boldsymbol{\alpha}}$$

where $\boldsymbol{T}^{(n_j \times t)}$ is an appropriately chosen matrix. The asymptotic variance matrix for the estimated variety mean yields is then given by

$$\tilde{\boldsymbol{\Sigma}}_j = \mathrm{V}(\boldsymbol{y}_j) = \tilde{\sigma}_s^2 \boldsymbol{T}(\boldsymbol{A}^\top \tilde{\boldsymbol{H}}_*^{-1} \boldsymbol{A})^{-1} \boldsymbol{T}^\top. \tag{3}$$

This matrix has an important role in the calculation of weights for the combined analysis (see Section 3.4).

3.4. Calculation of weights for combined analysis

As discussed in Section 3.1.3 the known weight matrix $\tilde{\boldsymbol{\Sigma}}^{-1}$ is approximated by the diagonal matrix $\boldsymbol{\Pi} = \mathrm{diag}\,(\pi_{ij})$, where π_{ij} denotes the weight for variety i in trial j. Thus

$$\mathrm{V}(\boldsymbol{\eta}) = \sigma^2 \boldsymbol{\Sigma} \approx \tilde{\sigma}^2 \boldsymbol{\Pi}^{-1}.$$

If an orthogonal analysis had been used for the analysis of individual trials the variance matrix in (3) would be given by $\tilde{\boldsymbol{\Sigma}}_j = \mathrm{diag}\,(\tilde{\sigma}_j^2 / r_{ij})$ where $\tilde{\sigma}_j^2$ is the error mean square for trial j and r_{ij} is the number of non-missing plot yields for variety i in trial j. Thus

$$\mathrm{var}\,(\eta_{ij}) \approx \frac{\tilde{\sigma}_j^2}{r_{ij}} = \tilde{\sigma}^2 \frac{\tilde{\sigma}_j^2}{\tilde{\sigma}^2 r_{ij}} \;\Rightarrow\; \pi_{ij} = \frac{r_{ij}}{\tilde{\sigma}_j^2 / \tilde{\sigma}^2}\,. \tag{4}$$

The weights are simply replication divided by a relative error variance for each trial.

With the use of non-orthogonal analyses $\tilde{\boldsymbol{\Sigma}}_j$ is not diagonal so the formula in (4) no longer holds. We could generalize to $\pi_{ij} = \tilde{\sigma}^2 \tilde{\sigma}_{jii}^{-1}$ where $\tilde{\sigma}_{jii}$ is the diagonal element of $\tilde{\boldsymbol{\Sigma}}_j$ corresponding to variety i. This is the standard linear models approach for accommodating known error variance heterogeneity, namely to use weights based on the inverse of the variances of the associated data points.

We can see from the mixed model equations (2), however, that what is needed is the best vector approximation to $\tilde{\boldsymbol{\Sigma}}^{-1}$ and not $\tilde{\boldsymbol{\Sigma}}$. In a least squares sense the best vector approximation to $\tilde{\boldsymbol{\Sigma}}^{-1}$ is given by the diagonal of the matrix. Thus the weights are formed as

$$\pi_{ij} = \tilde{\sigma}^2 \tilde{\sigma}_j^{ii}\,, \tag{5}$$

where $\tilde{\sigma}_j^{ii}$ is the ith diagonal element of $\tilde{\boldsymbol{\Sigma}}_j^{-1}$. Note that the standard weights based on $\tilde{\sigma}_{jii}^{-1}$ are too small unless the covariances between estimated variety means within a trial are negligible (see Frensham, Cullis & Verbyla, 1997c). For RCB analyses the two sets of weights

TABLE 8

Overview of models fitted to the 1998 Pinnaroo wheat trial (random variety effects)

Model	Error variance model: Global/Extraneous [A]	Error variance model: Local [B]	Parameters	REML log-likelihood	REML log-likelihood: Pr(*D*) [C]
1		AR1 × AR1	3	230.3	
2	ran(col)	AR1 × AR1	4	236.5	0.000
3	ran(col) + ran(row)	AR1 × AR1	5	242.5	0.000
4	ran(col) + ran(row)	AR1 × ID	4	242.1	0.38

[A] ran(col) and ran(row) represent (random) factors based on the column and row indices and are included in $\boldsymbol{B}$

[B] correlation models for separable (row × column) spatial process: AR1 = autoregressive of order 1; ID = identity matrix (independence)

[C] Significance of REML likelihood ratio tests comparing successive models

are identical since all covariance terms are zero, but for spatial (or IB) analyses there may be large discrepancies.

The mixed model analysis requires specification of the known portion of residual variance, that is, $\tilde{\sigma}^2$. This is also required to calculate the weights. We obtain $\tilde{\sigma}^2$ as the pooled error variance across trials. In contrast to an RCB analysis, there is no explicit error variance in a spatial (or IB) analysis. However, an estimated effective error variance, $\tilde{\sigma}_j^2$, can be calculated as

$$\tilde{\sigma}_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{r_{ij}}{\tilde{\sigma}_j^{ii}} .$$

Note that in the case of an equally replicated RCB analysis this formula gives the standard estimate of error variance, namely the error mean square.

The pooled error variance is then calculated as $\tilde{\sigma}^2 = \sum_{j=1}^{m} \tilde{\sigma}_j^2 / m$. The weights in (5) can then be expressed as a function of scale and replication as in (4), namely

$$\pi_{ij} = \frac{r_{ij}^*}{\tilde{\sigma}_j^2 / \tilde{\sigma}^2} , \qquad \text{where } r_{ij}^* = \tilde{\sigma}_j^{ii} \tilde{\sigma}_j^2 .$$

For an orthogonal analysis r_{ij}^* is the actual replication for each variety, and for a non-orthogonal analysis it reflects 'effective' replication. This form shows that the weights accommodate both heterogeneity of error variance across trials and unequal replication within a trial.

The information that must be saved from the analysis of individual trials therefore consists of the vector $\boldsymbol{y}_j$ of estimated variety mean yields, the diagonal elements $\tilde{\sigma}_j^{ii}$ of the inverse of the associated variance matrix, and the number r_{ij} of non-missing plot yields for each variety.

3.4.1. Example

In this section we demonstrate the calculation of weights for a trial from the SA wheat MET dataset investigated in Section 3.2. The trial was located at Pinnaroo in 1998 and consisted of four replicates of 39 entries arranged in the field as four columns by 39 rows. The design was constructed using SpaDes (Coombes, 1999). The trial mean yield was 1.51 t/ha and there were two missing values.

The initial analysis was conducted with random variety effects (see Section 3.3). The spatial modelling approach of Gilmour *et al.* (1997) led to the sequence of models displayed in Table 8.

TABLE 9

Information saved from analysis of wheat trial

Entry	Mean yield (t/ha)	Non-missing yields	Scaled weights
Barunga	1.45	3	711
Brookton	1.76	4	998
Carnamah	1.76	4	903
Diamondbird	1.44	4	946
Excalibur	1.75	4	948
Frame	1.44	4	948
⋮	⋮	⋮	⋮
WI97107	1.32	4	965
WI97115	1.73	4	1001
WI97119	1.54	4	913
Westonia	1.82	3	711
Worrakatta	1.54	4	944

TABLE 10

Sub-section of asymptotic variance matrix of mean yields for the wheat trial ($\tilde{\mathbf{\Sigma}}_j \times 10\,000$)

Entry	Barunga	Brookton	Carnamah	Diamondbird	WI97119	Westonia	Worrakatta
Barunga	23.74	7.90	8.00	8.19	9.18	8.57	8.65
Brookton	7.90	19.24	8.77	8.98	8.08	8.52	8.99
Carnamah	8.00	8.77	20.17	8.93	8.06	7.43	8.94
Diamondbird	8.19	8.98	8.93	19.5	7.76	8.10	8.16
WI97119	9.18	8.08	8.06	7.76	20.14	7.35	8.18
Westonia	8.57	8.52	7.43	8.10	7.35	24.17	7.96
Worrakatta	8.65	8.99	8.94	8.16	8.18	7.96	19.93

The final spatial model (model 4) was re-fitted with variety effects taken as fixed and with the variance parameters (components for the random row and column terms, autoregressive parameter for rows, and the error variance) held constant at the values obtained from model 4. A subset of the resultant BLUEs of variety means is shown in Table 9. The variances of the means (Table 10) are similar for all varieties except those which had missing data (Barunga and Westonia). The covariance terms were non-negligible, due to the existence of spatial variation (both extraneous and local).

Table 9 contains all the information that needs to be stored for use in the combined analysis. The scaled weights are $\tilde{\sigma}_j^{ii} (= \pi_{ij}/\tilde{\sigma}^2)$. These are similar for all varieties except Barunga and Westonia, highlighting that the weights in the combined analysis not only accommodate variance heterogeneity between trials but also spatial variation and unequal replication within trials. Since the covariance terms in $\tilde{\mathbf{\Sigma}}_j$ are substantial (Table 10) weights based on $\tilde{\sigma}_{j_{ii}}^{-1}$ would be too small. The full set of $\tilde{\sigma}_j^{ii}$ ranges from 711 to 1054 whereas $\tilde{\sigma}_{j_{ii}}^{-1}$ range from 414 to 534.

4. Discussion

In this paper we have presented protocols for the routine analysis and reporting of Australian CVEP information. This provides the opportunity for informed discussion, alternative proposals and refinements. We believe that the adoption of these protocols has substantially improved the science of crop variety evaluation in Australia. As a consequence the key objec-

tive of ACAS has been met, namely the provision of reliable information about the performance of crop varieties.

One possibly contentious issue is the choice between fixed and random effects in the analysis. We have considered this in great depth and lament the lack of direction in the statistical literature. The standard text-book notion of effects being random if they have been sampled from a population and fixed if attention is confined only to those effects in the model (see Searle, 1971, for example) is unhelpful and can lead to a circular argument. In our opinion the choice depends on the aim of the analysis. In terms of variety effects our aim is to *predict* future performance. This is best achieved by assuming the effects to be random. Initially, plant breeders and evaluators were sceptical about the use of BLUPs. They now accept the method because the predictions have been more realistic. It is no longer true that yield gains observed by farmers are substantially lower than those predicted by CVEPs. We do not wish to predict environment effects. The effects could be assumed random in order to recover information on varieties, but the variance component for environments is usually so large that very little information is recovered. The magnitude of the component also means there is very little shrinkage of environment effects, with the result that BLUPs and BLUEs are almost identical. We therefore assume that environment effects are fixed.

A key area for further research is the definition of regions. The current boundaries have a historical basis and have not been investigated scientifically. If varietal information is to be reported at a regional level the variance component for variety-by-region interaction should be large. In an examination of variance components in 22 Australian CVEP datasets Cullis *et al.* (2000) show that this is rarely the case, with the variety-by-region variance components ranging from 0% to 8% of total variance (average of 3%). There is a clear need to use historical data to examine regional boundaries. Techniques such as pattern analysis and AMMI (Gauch, 1992) could be used for this purpose. The multiplicative mixed model of Smith *et al.* (1998) provides a random effects analogue of these approaches, thereby allowing incomplete data to be examined more efficiently.

References

BRENNAN, P.S. & BYTH, D.E. (1979). Genotype × environment interactions in wheat yields and selection for widely adapted wheat genotypes. *J. Agric. Sci., Cambridge* **30**, 221–232.

CLEMENTS, R.J., ROSIELLE, A.A. & HILTON, R.D. (1992). *National Review of Crop Improvement in the Australian Grains Industry: a Report to the Board of the Grains Research and Development Corporation.* Canberra, ACT: Grains Research and Development Corporation.

COOMBES, N.E. (1999). *SpaDes, a Spatial Design Generator.* Biometric Bulletin. NSW Agriculture, Wagga Wagga, NSW.

COOPER, M. & DE LACY, I.H. (1994). Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoret. Appl. Genetics* **88**, 561–572.

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.B., HUNT, C.H. & GILMOUR, A.R. (2000). An examination of the efficiency of Australian crop variety evaluation programmes. *J. Agric. Sci., Cambridge* **135**, 213–222.

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

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

FINLAY, K.W. & WILKINSON, G.N. (1963). The analysis of adaptation in a plant breeding programme. *Austral. J. Agric. Res.* **14**, 742–754.

FINNEY, D.J. (1964). The replication of variety trials. *Biometrics* **20**, 1–15.

FRENSHAM, A.B., BARR, A.R., CULLIS, B.R. & PELHAM, S.D. (1997a). The analysis of 10 years of South Australian oat evaluation data. *Euphytica* **99**, 43–56.

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

FRENSHAM, A.B., CULLIS, B.R. & VERBYLA, A.P. (1997c). *Efficient Estimation in the Two-stage Analysis of Multi-environment Trials.* Research Report Series. Department of Statistics 97/7. University of Adelaide.

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. Rothamsted Experimental Station, United Kingdom: Lawes Agricultural Trust.

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. *J. Agric., Biological and Environmental Statistics* **2**, 269–273.

GILMOUR, A.R., CULLIS, B.R., FRENSHAM, A.B. & THOMPSON, R. (1998). (Co)Variance structures for linear models in the analysis of plant improvement data. In *COMPSTAT98 Proceedings in Computational Statistics.* pp. 51–64. Heidelberg: Physica-Verlag.

GILMOUR, A.R., CULLIS, B.R., WELHAM, S.J. & THOMPSON, R. (1999). *ASREML Reference Manual.* Biometric Bulletin, No 3. NSW Agriculture, Orange, NSW.

JOHN, J.A. & WILLIAMS, E.R. (1998). t-Latinized designs. *Aust. N. Z. J. Stat.* **40**, 111–118.

KROONENBERG, P.M. & BASFORD, K.E. (1989). An investigation of multi-attribute genotype response across environments using three-mode principal component analysis. *Euphytica* **44**, 109–123.

LAZENBY, A. (1986). *Australia's Plant Breeding Needs: A Report to the Minister for Primary Industries.* Canberra, ACT: AGPS.

LAZENBY, A., BARTHOLOMAEUS, M., BOUCHER, B., BOYD, W.R., CAMPBELL, A., CRACKNELL, R., EAGLES, H., LEE, J., LUKEY, G. & MARSHALL, B. (1994). *Trials and Errors: A Review of Variety Testing and Release Procedures in the Australian Grains Industry.* Canberra, ACT: Grains Research and Development Corporation.

LESLIE, J.K., ARNEY, J., CRACKNELL, R., GILL, W., LUKEY, G., McNEE, D.A.K., PLOWMAN, D. & SMITH, M. (1997). *National Strategy for the Evaluation of Grain Crop Varieties in Australia: Report of the Steering Committee on Evaluation and Accreditation of Grain Crop Varieties in Australia.* Canberra, ACT: Grains Research and Development Corporation.

MARTIN, R.J. & ECCLESTON, J.A. (1991). Efficient block designs for correlated observations. *Austral. J. Statist.* **33**, 299–311.

PATTERSON, H.D. & SILVEY, V. (1980). Statutory and recommended list trials of crop varieties in the United Kingdom. *J. Roy. Statist. Soc. Ser. A* **143**, 219–252.

PATTERSON, H.D. & THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **31**, 100–109.

PATTERSON, H.D., SILVEY, V., TALBOT, M. & WEATHERUP, S.T.C. (1977). Variability of yields of cereal varieties in UK trials. *J. Agric. Sci., Cambridge* **89**, 238–245.

PATTERSON, H.D. & NABUGOOMU, F. (1992). REML and the analysis of series of crop variety trials. In *Proceedings from the 25th International Biometric Conference, Hamilton, New Zealand.* pp. 77–93.

PEDERSON, D.G. & RATHJEN, A.J. (1981). Choosing trial sites to maximize selection response for grain yield in spring wheat. *Austral. J. Agric. Res.* **32**, 411–424.

SEARLE, S.R. (1971). *Linear Models.* New York: John Wiley and Sons, Inc.

SILVEY, V. (1978). Methods of analysing NIAB variety trial data over many sites and several seasons. *J. National Institute of Agric. Botany* **14**, 385–400.

SMITH, A.B., CULLIS, B.R., GILMOUR, A.R. & THOMPSON, R. (1998). Multiplicative models for interaction in spatial mixed model analyses of Multi-environment trial data. *Proc. Internat. Biometrics Conf.*, Cape Town, South Africa. Invited Papers, pp. 91–101.

TALBOT, M. (1984). Yield variability of crop varieties in the UK. *J. Agric. Sci., Cambridge* **102**, 315–321.

WHEELER, R. & McMURRAY, L. (1999). *Post Harvest Report 1998/99.* Adelaide, SA: South Australian Field Crop Evaluation Program.