

C.G. Qiao · K.E. Basford · I.H. DeLacy · M. Cooper

Evaluation of experimental designs and spatial analyses in wheat breeding trials

Received: 17 December 1998 / Accepted: 29 July 1999

Abstract Thirty-three wheat breeding trials were conducted from 1994 to 1996 in the Northern Grains Region (QLD and Northern NSW) of Australia to evaluate the influence of experimental designs and spatial analyses on the estimation of genotype effects for yield and their impact on selection decisions. The relative efficiency of the alternative designs and analyses was best measured by the average standard error of difference between line means. Both more effective designs and spatial analyses significantly improved the efficiency relative to the randomised complete block model, with the preferred model (which combined the design information and spatial trends) giving an average relative efficiency of 138% over all 33 trials. When the Czekanowski similarity coefficient was used, none of the studied models were in full agreement with the randomised complete block model in the selection of the top lines. The agreement was influenced by selection proportions. Hence, the use of these methodologies can impact on the selection decisions in plant breeding.

Key words Wheat · Breeding trials · Experimental designs · Spatial analysis · Selection

Introduction

A feature common to most plant breeding experiments is the presence of systematic heterogeneity among the experimental units. Typically, the nature of this heterogeneity is such that there is appreciable correlation among neighbouring units (Zimmerman and Harville 1991). This makes the widely used randomised complete block (RCB) design unsuitable for such trials. Yates (1939) and

Cochran and Cox (1957) first introduced incomplete block (IB) and lattice square designs, which allowed differences between rows and columns within each replicate to be eliminated from variety comparisons. Williams and John (1989) developed a method for constructing row and column designs with contiguous replicates, while Nguyen and Williams (1993) developed computer algorithms, which made it possible for designs with any number of rows and columns and genotypes to be formed. Patterson and Hunter (1983) and Yau (1997) showed these designs to be more efficient than the RCB design in field crop variety trials. However, these regular arrangements of blocks do not always provide efficient control of spatial variation in field trials (Grondona *et al.* 1996). Robinson *et al.* (1988) used a row-column model to analyse trials and presented an alternative analytical methodology, known as postblocking, in an attempt to use additional information other than design information to account for the spatial variation in field trials. One- and two-dimensional spatial analysis (trend analysis, or neighbour analysis, or nearest neighbour analysis) was developed to accommodate the spatial correlations among neighbouring experimental units in variety trials and has been shown to be more efficient than the conventional analyses (Gleeson and Cullis 1987; Cullis and Gleeson 1989, 1991). Based on a cross-validation approach, Grondona *et al.* (1996) showed that models accounting for spatial correlation in two dimensions were the most efficient analysis of variety trials among the models they examined. Gilmour *et al.* (1997) concluded that the general superiority of the two-dimensional spatial model over the IB model justifies its use as an initial model for spatial analysis. Brownie and Gumpertz (1997) presented further evidence for improved efficiency of spatial models over RCB and IB models. However, the impact of these methodologies on selection response in plant breeding has not yet been examined. Moreover, in assessing these new methods several criteria have been frequently used, such as average variance of difference between line means (AV), effective error mean square (EEMS), average standard error of the difference

Communicated by G. Wenzel

C.G. Qiao (✉) · K.E. Basford · I.H. DeLacy · M. Cooper
School of Land and Food, The University of Queensland,
Brisbane, QLD 4072, Australia
e-mail: c.qiao@mailbox.uq.edu.au,
Fax: +61 7 33651177

between line means (SED) and error mean square (EMS) for the estimation of relative efficiency. We use yield data from wheat multi-environment trials (METs) in the Northern Grains Region (QLD and Northern NSW) of Australia for the evaluation of these methodologies.

The study presented here aimed at (1) comparing different measures of relative efficiency to find the more appropriate one in plant variety trial analysis; (2) evaluating different experimental designs and spatial methods in the estimation of genotype effects for wheat breeding trials in the Northern Grains Region of Australia; and (3) examining the effects of these methodologies on the selection of superior lines.

Material and methods

Experimental methods

Thirty-three wheat trials were conducted from 1994 to 1996 in the Northern Grains Region of Australia. The entries consisted of 72

breeding lines from the Plant Breeding Institute of Sydney University and the Queensland Wheat Research Institute and a group of 19 common lines called probe genotypes (Cooper and Fox 1996). The latter were used in these trials for characterising the target population of environments (TPEs) in this region. Triple lattice designs were used for all the trials, with 9 laid out in a one column by many rows array (called long-face trials) and the rest laid out in a row-column array (called row-column trials), 6 of which were latinised (Table 1). Hence, incomplete block, row-column and latinised row-column models were valid for these three types of trials, respectively.

Analytical methods

Model specifications

The grain yield (t/ha) data were analysed using the software ASREML (Gilmour *et al.* 1998) such that the following models were applied for each of the 33 trials wherever possible:

- 1) randomised complete block (RCB);
- 2) incomplete block (IB);
- 3) row-column (RC);
- 4) latinised row-column (LRC);

Table 1 Experimental details and the preferred model^a for each of the 33 wheat breeding trials

Trial	Year	Location	Design ^b	Layout	Preferred model (PF) ^c
1	1994	Biloela, QLD	LF	1 × 149	l(row)+s(row)+AR1
2	1994	Jimbour, QLD	LF	1 × 149	rep+blk/rep
3	1994	Meandarra, QLD	LF	1 × 149	l(row)+blk/rep+AR1
4	1994	Oakeleigh, QLD	LF	1 × 149	l(row)+blk/rep+AR1
5	1994	The Gum, QLD	LF	1 × 149	blk/rep+AR1
6	1995	Billi Billa, QLD	LF	1 × 149	l(row)+s(row)+rep+AR1
7	1995	Biloela, QLD	LF	1 × 149	l(row)+s(row)+blk/rep+AR1
8	1995	Bungunya, QLD	LF	1 × 149	l(row)+s(row)+blk/rep+AR1
9	1995	Tumaville, QLD	LF	1 × 149	l(row)+s(row)+blk/rep+AR1
10	1994	Fernlees, QLD	RC	7 × 21	fcoll+frow+AR1×AR1
11	1994	Gatton, QLD	RC	5 × 15	fcoll+frow+row/rep+AR1×AR1
12	1995	Gatton, QLD	RC	5 × 15	frow+coll+col/rep+AR1×AR1
13	1996	Billi Billa, QLD	LRC	5 × 15	fcoll+frow+rep+AR1×AR1
14	1996	Bungunya, QLD	LRC	5 × 15	fcoll+frow+l(row)+l(col)+AR1×AR1
15	1996	Fernlees, QLD	LRC	5 × 15	fcoll+frow+col/rep+AR1×AR1
16	1996	Gatton, QLD	LRC	5 × 15	frow+coll+AR1×AR1
17	1996	Jimbour, QLD	LRC	5 × 15	fcoll+frow+rep+AR1×AR1
18	1996	Moonie, QLD	LRC	5 × 15	fcoll+frow+AR1×AR1
19	1994	Moree, NSW	RC	5 × 15	l(row)+l(col)+s(row)+s(col)+row/rep+AR1×AR1
20	1994	Myall Vale, NSW	RC	5 × 15	fcoll+frow+AR1×AR1
21	1994	Narrabri, NSW	RC	5 × 15	fcoll+frow+AR1×AR1
22	1994	Narrabri Late, NSW	RC	5 × 15	fcoll+frow+AR1×AR1
23	1994	North Star, NSW	RC	5 × 15	fcoll+frow+AR1×AR1
24	1994	Warren, NSW	RC	5 × 15	fcoll+frow+blk/rep+AR1×AR1
25	1995	Gilgandra, NSW	RC	6 × 18	col+col/rep+row/rep+AR1×AR1
26	1995	Myall Vale, NSW	RC	6 × 18	row+col+AR1×AR1
27	1995	North Star, NSW	RC	6 × 18	fcoll+frow+AR1×AR1
28	1996	Gilgandra, NSW	RC	18 × 6	fcoll+frow+rep+row/rep+AR1×AR1
29	1996	Moree, NSW	RC	18 × 6	fcoll+row+rep+AR1×AR1
30	1996	Myall Vale, NSW	RC	18 × 6	fcoll+frow+ AR1×AR1
31	1996	Narrabri, NSW	RC	18 × 6	fcoll+frow+rep+AR1×AR1
32	1996	North Star, NSW	RC	18 × 6	fcoll+row/rep+AR1×AR1
33	1996	Spring Ridge, NSW	RC	18 × 6	fcoll+rep+AR1×AR1

^a The grain yield (dependent variable), grand mean, genotype effect and residual terms are omitted for each of the models.

^b LF, Long-face design; RC, row-column design; LRC, latinised row-column design

^c row, random row effect; row/rep, row within replicates effect; col/rep, column within replicates effect; blk/rep, block within replicates effect; col, random column effect; frow, fixed row effect; rep, replicate effect; fcol, fixed column effect; l(col), fixed linear

column effect; l(row), fixed linear row effect; AR1, one-dimensional auto-regressive process; AR1 × AR1, two-dimensional auto-regressive process; s(row), a random curvature component for rows, spline row effect; s(col), a random curvature component for columns, spline column effect. *Note* that a spline effect is fitted in two components – a fixed linear trend and a random curvature component

- 5) standard spatial (SS, model with only AR1×AR1 or AR1 and genotype effect terms);
- 6) standard spatial fitted with measurement error (SSM);
- 7) preferred (PF).

AR1 × AR1 is the separable first-order autoregressive process along both the rows and columns and is the key term in the two-dimensional spatial model; AR1 stands for the separable first-order autoregressive process along only the rows or columns (only along the rows in this study) and is the key term in the one-dimensional spatial model. In the above models, line effects were treated as fixed, and all the other terms were treated as random, unless otherwise stated.

RCB and IB models were fitted as described by Cochran and Cox (1957), while RC and LRC models followed those of Williams and Matheson (1994). All 24 row-column trials were analysed as both RC and LRC models, although only 6 of them were designed as LRC trials. This practice was based on the technique of postblocking described by Ainsley *et al.* (1985) in which artificial blocks are imposed on the plots and the trial analysed as if these were the true blocks. The one- and two-dimensional spatial analyses of Gleeson and Cullis (1987) and of Cullis and Gleeson (1991) were used for the long-face and row-column trials, respectively, to produce the SS and SSM models called spatial models (Grondona *et al.* 1996; Gilmour *et al.* 1997). The approach of Gilmour *et al.* (1997) was used for the PF search, in which the SS model was first fitted as an initial model. Then the variograms (a graphical display of the semi-variance of the difference between plots h rows or columns apart) and the related diagnostics such as log-likelihood ratio (LOG), AV, EEMS, SED, EMS, auto-correlation between plots within a row, auto-correlation between plots within a column and residual plots were used to facilitate the search for the PF. The PF has a target of a flat variogram, largest LOG, lowest SED and lowest EMS.

Assessment of the methodologies

The impact of these experimental designs and analytical methods was assessed by relative efficiency in terms of the size of the error – the improvement in precision or efficiency over the conventional RCB model. The practical importance of these alternative models is evaluated by their impact on the selection of the top lines based on grain yield.

Relative efficiency

Relative efficiency (RE) is used to assess the improvement in precision of the alternative models over the RCB model. It is measured in terms of AV, EEMS or SED as follows:

$$RE_{AV} = 100 \times (AV_{RCB}/AV_{AT})$$

$$RE_{EEMS} = 100 \times (EEMS_{RCB}/EEMS_{AT})$$

$$RE_{SED} = 100 \times (SED_{RCB}/SED_{AT})$$

where the subscript AT denotes any of the alternative models to the RCB model. These quantities were calculated because they are widely used for varietal comparisons and/or for measuring the precision or efficiency of experiments (Patterson and Hunter 1983; Cullis and Gleeson 1991; Kempton *et al.* 1994; Zimmerman and Harville 1991). However, the results of RE expressed in EEMS are not reported here because $RE_{AV} = RE_{EEMS}$, which can be shown by the linear relationship between the two: AV is twice the EEMS divided by the number of replicates (Cochran and Cox 1957).

Selection of superior lines

Three popular measures of similarity (Snijders *et al.* 1990; Everitt 1993), namely Simple Matching coefficient (M), Czekanowski coefficient (D) and Jaccard coefficient (J), were considered for comparing each of the alternative models with the RCB model in se-

lection of the top lines. When selection intensity is fixed at a particular level, corresponding to the top 0.05, 0.10, 0.15, 0.20 or 0.40 proportion of the lines, these coefficients are defined as:

$$M = \frac{a+d}{a+b+c+d} \quad D = \frac{2a}{2a+b+c} \quad J = \frac{a}{a+b+c}$$

where a is the number of lines selected by both the alternative and the RCB model, b is the number of other lines selected by the alternative model, c is the number of other lines selected by the RCB model and d is the number of lines rejected by both models. Thus, the total number of lines grown in the trial is $a + b + c + d$ and, in this case, $b = c$ for a defined selection proportion.

The definition of M gives more emphasis on ‘double negatives’ (d), the number of lines rejected by both models, while the definitions of D and J emphasise the ‘double positives’ (a), the number of lines selected by both models (Everitt 1993). In a molecular genetics study, Nei and Li (1979) measured the similarity of two genetic populations as the ratio of the number of nucleotide fragments shared by both populations to the average number of fragments in these two populations, which is the Czekanowski similarity coefficient in a different form. For our study of percentage agreement in selection of the top lines, D and J are more appropriate than M , with D being somewhat conservative due to its less emphasis on the number of lines selected by either model but not both. From the plant breeding point of view, it would be more appropriate to measure the extent to which the selections from the two models agree when the top lines are selected separately for each of the models. It can be shown that the Czekanowski coefficient meets this purpose after a simple transformation:

$$D = \frac{2a}{2a+b+c} = \frac{a}{\frac{(a+b)+(a+c)}{2}} = \frac{a}{a+b} = \frac{a}{a+c},$$

which is the percentage of the number of lines selected by both models over the number of lines selected by either the alternative model or the RCB model (as $b = c$ for a given selection intensity). Hence, the Czekanowski coefficient is the appropriate similarity measure for the objectives of this study. Following these considerations, the Czekanowski coefficient, averaged over all 33 trials, was used to present the main findings from this study.

Results and discussion

Preferred model compositions

The preferred models (PFs) fitted for all 33 trials are presented in Table 1. With few exceptions the PF contained the spatial terms (AR1 × AR1 or AR1), suggesting that spatial adjustments are important in modelling our trial data. This agreed with the results of Cullis and Gleeson (1991), Grondona *et al.* (1996) and Gilmour *et al.* (1997), who found that the best model for each data set was always one with spatial terms (AR1×AR1 or AR1). Design information such as replicate, row and/or column effects in various forms was also used in almost all preferred models. This means that in searching for the PF, the design information and spatial adjustment of the trends both need to be considered.

Basford *et al.* (1996) concluded that the best model is that which combines design information and smooths the spatial trends. Federer (1998) also emphasised the importance of combining the spatial analysis and blocking analysis of experimental results in the search for the ‘best’ model. Spline row and/or column terms were included in the PF for almost half of the trials, which

Table 2 A comparison of two measures of relative efficiency for different models in Trial 11

Measure ^a	RCB ^b	IB	RC	LRC	SS	SSM	PF
AV	0.137	0.121	0.074	0.072	0.081	0.081	0.061
SED	0.369	0.348	0.271	0.268	0.284	0.284	0.247
RE _{AV}	100.0	112.9	185.2	190.0	168.7	169.0	223.3
RE _{SED}	100.0	106.3	136.1	137.8	129.9	130.0	149.0

^a AV, The average variance of difference between line means; SED, The average standard error of difference between line means; RE_{AV}, relative efficiency measured in AV; RE_{SED}, relative efficiency measured in SED

^b RCB, randomised complete block model; IB, randomised incomplete block model; RC, row-column model; LRC, latinised row-column model; SS, standard spatial model with only AR1 × AR1 (or AR1) + genotype effect terms; SSM, SS model fitted with measurement error term; PF, preferred model

demonstrates the presence of global trends in these trials (Gilmour *et al.* 1997). Many of these global trends could not have been adequately modelled by the traditional blocking structures in the wheat breeding trials.

Relative efficiency

Measures of relative efficiency

For each of the models, the AV, RE_{AV}, SED and RE_{SED} were estimated to provide a measure of precision and relative efficiency between different models of the same trial. As an example, in Trial 11 (Gatton 1994) an increase in the precision of the PF model over the RCB model was demonstrated by a value of 0.076 (0.137–0.061) when measured in AV, but this increase was 0.122 (0.369–0.247) when measured in SED (Table 2). Similarly for the same trial, the improvement in efficiency of the PF model relative to the RCB model was 49% if estimated by SED, but it was 123% if estimated by AV, which is approximately 2.5 times the first quantity (Table 2). The comparisons of RE for other trials or models with these two quantities show similar trends. This is a typical reflection of the different ways of expressing RE in the literature where both versions are commonly quoted. The use of different bases for computing relative efficiency can make the comparison of experimental results in terms of relative efficiencies difficult. Hence, it is essential that an appropriate measure of RE is identified and used in the evaluation of different models in such experiments.

Yates (1939) and Cullis and Gleeson (1989) described the efficiency of a particular analysis as the AV from the RCB analysis divided by the AV from that analysis. Patterson and Hunter (1983) used the same measure of efficiency. In a consideration of the relative merits of two designs of the same experiment, Cochran and Cox (1957) used the average experiment error over all treatment comparisons to indicate the net gain in precision. Magnussen (1990) evaluated the merits of RCB and various spatial models in terms of SED and blocking efficiency (BE). Cullis and Gleeson (1991) used both SED and EEMS for the comparison of models. Gleeson (1997) suggested the use of the average SED as a summary statistic for model comparisons. In comparing dif-

ferent modelling processes, Lill *et al.* (1988) presented evidence to suggest that the average standard error of a difference between variety means (SED) for the spatial model is a reasonably unbiased estimate of the true SED. Binns (1987) argued that when comparing two methods of analysis it is not sufficient merely to compare the error mean squares because the precision of treatment comparisons measured by SED is the key quantity.

From a practical point of view, RE measured in terms of the SED (RE_{SED}) seems more relevant than measuring it in terms of AV (RE_{AV}) because the SED is directly used for comparisons between line means, has the same scale as the original attributes and includes the efficiency factor. Furthermore, RE_{SED} seems more appropriate as a percentage as it is based on a quantity whose scale is the same as that of the original data. On the other hand, the AV, which is the square of SED, is indirectly used by plant breeders for varietal comparison and its meaning is not so obvious. Yau (1997), in adopting SED, argued that this statistic is more readily understandable than the AV. If AV were used in the expression of RE, it would exaggerate the actual size of the efficiency. Hence, we recommend the use of SED and RE_{SED} in such studies.

Designs with different blocking structures

The average RE_{SED} for the IB model was higher for the 9 long-face trials (139%) than for the row-column trials (108%) with an overall mean of 116% (Table 3). This implies that the row-column trials are relatively more homogeneous in their experimental units within the same replicates than the long-face trials. The smaller mean RE for the row-column trials is comparable to the efficiency of 108% in the wheat study of Grondona *et al.* (1996). This suggests that for the row-column trials, blocking from one dimension might not be sufficient to bring about an improvement in efficiency. The IB model did not exhibit an improvement in efficiency over the RCB model for many of the row-column trials.

Both RC and LRC models produced an average increase in efficiency of more than 10% for the row-column trials, which is an improvement over the IB model in the same trials. The RC and LRC models did not show too much difference for the row-column trials as a whole. However, a large difference was observed in

Table 3 Relative efficiency in percentage in terms of the average standard error of difference between line means (RE_{SED}) for different models fitted in each of the 33 trials

Trial	IB ^a	RC	LRC	SS	SSM	PF
1	147.9	—	—	178.8	179.2	181.0
2	101.3	—	—	99.7	100.8	101.3
3	278.2	—	—	356.6	358.2	369.8
4	102.0	—	—	49.1	102.8	103.1
5	140.6	—	—	174.8	176.2	176.2
6	112.6	—	—	109.4	113.7	120.1
7	110.3	—	—	56.9	111.2	112.0
8	124.6	—	—	123.9	137.8	138.6
9	135.3	—	—	170.4	170.9	173.7
10	109.7	110.4	110.8	126.7	126.7	138.4
11	106.3	136.1	137.8	129.9	130.0	149.0
12	116.5	121.1	126.0	112.0	112.0	131.2
13	113.0	113.0	121.7	135.1	135.1	141.8
14	109.1	113.3	120.1	129.3	129.3	156.4
15	121.8	125.1	127.0	144.9	144.9	157.7
16	146.7	150.3	174.9	189.3	189.3	188.7
17	100.9	101.9	102.6	96.9	98.2	113.9
18	101.9	101.9	105.6	110.8	110.8	120.0
19	100.0	100.0	102.3	105.9	105.9	119.1
20	102.1	103.0	103.0	100.4	100.4	113.3
21	100.8	106.4	106.4	96.6	97.8	125.7
22	100.0	101.6	101.4	100.2	100.2	108.8
23	100.0	101.7	101.8	102.9	102.9	108.8
24	101.2	107.2	107.8	99.1	99.1	140.8
25	133.9	135.2	140.7	144.6	144.8	140.7
26	114.5	127.7	135.2	109.5	124.0	152.8
27	100.4	100.4	100.4	104.6	103.3	122.3
28	100.0	102.8	102.8	97.5	99.0	110.3
29	101.5	102.7	102.8	97.9	100.4	107.1
30	101.9	101.9	101.9	104.4	105.0	113.5
31	100.0	100.5	100.5	86.9	86.9	103.5
32	100.0	101.2	101.2	100.0	100.0	107.7
33	100.0	100.1	100.1	93.7	93.7	106.3
Mean for the long-face trials	139.2	—	—	146.6	161.2	164.0
SE for the long-face trials	18.3	—	—	30.8	26.8	27.8
Mean for the row-column trials	107.6	111.1	113.9	113.3	114.2	128.3
SE for the row-column trials	2.5	2.9	3.8	4.7	4.7	4.4
Mean for all the trials	116.2	—	—	123.2	127.0	138.0
SE for all the trials	5.7	—	—	9.3	9.1	8.4

^a For abbreviations see Table 2

most of the QLD row-column trials. The majority of the improvement in efficiency of the LRC model over the RC model comes from the QLD row-column trials where 6 out of 9 trials were latinised row-column designs. Only 2 NSW trials (Trial 25 and 26) indicated a superiority of the LRC model over the RC model, and the difference (5–7%) is relatively smaller than that for the QLD trials.

These results show that good experimental designs were necessary and that gain in efficiency can be realised from the recovery of row and column information only if the trials have been designed with such a structure. The difference between the RC and LRC models is that the latter has an extra long column or long row term. This may imply that one could hardly expect a benefit from analysing the variation in a particular blocking structure if the trials have not been designed as such. This result disagrees with those of several researchers who favoured the technique of postblocking. Patterson and Hunter (1983) analysed data from trials designed as RCB as if they had been drawn up as row-column designs and found an improvement in efficiency. In trials

with an array of two dimensions there are many possible blocking arrangements. Robinson *et al.* (1988) investigated the effects of making adjustment for rows, columns, parts of rows, double rows and various combinations for row plus column effects. They commented that the technique is useful for investigations such as theirs, but it should not, in general, be used for the routine analysis of trials. Ainsley *et al.* (1985) showed that post-blocking a RCB design as an IB design is weakly valid under randomization and gained 16% increase in efficiency. The effect of a possible bias in the analysis of any single trial will be substantially reduced when results are averaged over a number of trials. Federer (1998) argued that at the stage of data analysis, any information should be taken into account if it helps to smooth the trend. In the present study, a significant improvement has been found for analysing the trials as a LRC for 4 out of 18 row-column trials which were not designed as a LRC. Therefore, the validity and application of the postblocking technique for wheat trials in the Northern Grains Region of Australia requires further investigation.

Spatial analysis

There was little difference between the two spatial models SS and SSM in efficiency, except for the long-face trials where the mean RE for SSM was about 10% higher than that of SS model (Table 3). Out of the 24 row-column trials, only 1 (Trial 26) showed an improved efficiency of the SSM over the SS model. This indicates that fitting a measurement error term to the spatial model does result in a gain in the trial efficiency in some trials. The spatial models SS and SSM in general produced a much higher efficiency than the IB model for the long-face trials, while the improvement in efficiency was greatly reduced for the row-column trials. The extreme cases were Trials 17, 21, 28, 31 and 33 where the SS and SSM models were slightly less efficient than the RCB model, obviously due to the lack of blocking or design information in the models. This defect was overcome in Trials 4, 7 and 29 where the switch from SS to SSM model compensated for the design information lacking in the SS model. Hence, both good experimental designs and appropriate analytical procedures are important to obtain a high efficiency for these wheat breeding trials. The average REs of the spatial models are similar to those of the RC and LRC models when calculated for all the row-column trials. This demonstrates the advantages of both more effective experimental designs and advanced analytical methods over the more conventional RCB model. The combination of well-designed experiments and spatial analysis would yield an improvement in mean RE compared to using one approach by itself.

The mean RE of all 33 trials amounted to 123% and 127% for SS and SSM models, respectively, compared to the average RE of 130% for the spatial models in the Grondona *et al.* (1996) study. Magnussen (1990), Kempton *et al.* (1994) and Basford *et al.* (1996) all found a clear advantage in the spatial models over the RCB model in terms of efficiency measured in SED.

The preferred model

The preferred model (PF) gave a mean RE of almost 164% over the RCB model for the long-face trials, which is similar to the average RE (157%) of the best spatial model in the study of Grondona *et al.* (1996), while for the row-column trials this figure was much lower (128%). There are two cases in the long-face trials where the PF does not seem to show any advantage over the RCB model. The mean RE of the PFs is always higher than any of the alternative models from the more effective designs (IB, RC and LRC) and the spatial models (SS and SSM) in each type of trial. Gilmour *et al.* (1997) showed that their PFs gained a mean improvement in RE_{SED} of 7% over the lattice design models in a range of South Australian (SA) wheat variety trials. The mean RE for the PFs over all 33 trials in our study is 138%, which is comparable to the RE_{SED} of 143% in a similar study of wheat variety trials (Binns 1987) but smaller than the av-

erage RE_{SED} of 157% achieved by Grondona *et al.* (1996) in wheat variety trials. Cullis and Gleeson (1991) reduced SED from 86.9 for the RCB model to 43.2 for the best spatial model, a RE of more than 200% in the analysis of one dataset. Gilmour *et al.* (1997) achieved a RE_{SED} as high as 249% in their demonstration of the spatial models with SA wheat data.

In our study, 11 of 33 trials had a RE higher than 140%, the highest being 370% for Trial 3. Therefore, application of these analytical methods is important in achieving high precision or efficiency in these wheat variety trials and should be used whenever possible in the Northern Grains Region. Basford *et al.* (1996) suggested a unified model for the analysis of variety trials which includes all of the one- and two-dimensional spatial models as well as the random block and row plus column models. The analysis of such designed trials can be conducted as an error variance modelling exercise in which the most appropriate variance model is fitted (be it spatial, block or a mixture of spatial and block). However, time availability is a practical constraint in applying these methodologies in plant breeding. Experience in agriculture seems to indicate that a more orderly model choice is possible and that models of a certain type tend to yield comparable estimates of spatial terms (Kempton 1984). Grondona *et al.* (1996) found a practical procedure in which a fixed number of simple models including spatial models are searched for the preferred model.

The difference between long-face trials and row-column trials in RE for alternative models may reflect the effects of the shape of the field layout of the trial plots. The row-column trials extend a shorter distance spatially than the long-face trials and hence may be less likely to be affected by field trends of various forms. Kempton *et al.* (1994) also found that the improvement in efficiency of these analyses is higher when plots form a single linear array than when they form a row-column array.

Selection of superior lines: effect of different selection intensities

Each of the alternative models was compared with the RCB model by using the average across all trials of the Czekanowski coefficients which measure the average percentage agreement in selection of the top lines (Table 4). None of the models showed full agreement with the RCB at any selection intensity when only row-column trials or all 33 trials were analysed. Among the alternative models, the PF had the lowest agreement with the RCB model, followed by the spatial models SS and SSM. The models based on different designs (IB, RC and LRC) which used only the design information showed more agreement with the RCB. The average percentage agreement remained fairly constant when the selection proportion decreased from 0.40 to 0.15. It dropped sharply when the top 10% lines were selected for all models and returned to a higher figure when a 5% selection intensity was imposed. The

Table 4 Average percentage agreement (%) in selection of the top lines based on the Czekanowski coefficient for the comparison of the models with the RCB model for five selection proportions (P)

P	IB ^a	RC	LRC	SS	SSM	PF
0.40	94.1 (91.3)	92.0	91.6	90.3 (88.3)	89.9 (88.3)	87.8 (86.0)
0.20	97.2 (93.5)	95.8	96.6	89.1 (85.0)	88.2 (86.0)	86.2 (84.0)
0.15	95.8 (88.5)	95.7	95.7	92.6 (84.7)	91.5 (84.7)	86.2 (80.9)
0.10	90.0 (82.9)	85.0	86.7	85.0 (77.1)	81.7 (79.1)	76.7 (73.3)
0.05	94.1 (88.5)	91.2	91.2	88.2 (78.9)	88.2 (82.7)	79.4 (76.9)

^a For abbreviations, see Table 2

ranking of percentage agreement for each model remained the same across all selection intensities.

These results indicate that spatial analysis is the key factor that contributes to the difference in the lines selected between the alternative models and the RCB model. For the IB, RC and LRC models, where there is no $AR1 \times AR1$ or $AR1$ term, the selections did not differ as much as for the spatial models which did have this term. This indicates that the application of these methodologies would have an impact on the selection of superior lines in these trials. The impact depends on the selection proportions or intensities. The interrelationships between selection intensities and the alternative models require further investigation.

Conclusions

Both effective experimental designs and spatial analyses can have an important role in improving the efficiency or precision of the experiment. For the row-column trials, analytical models from different designs (incomplete block, row-column and latinised row-column) can improve efficiency by an average of 8%, 11% and 14%, respectively. Spatial models can improve efficiency by an average of 13% and 14% for a standard spatial model and a standard spatial model fitted with a measurement error, respectively. The postblocking technique appears to have limited application in our trials. The preferred model, which is usually the one that combines the design information and spatial adjustments, can produce an average improvement in efficiency of 28%. Averaged over all trials, the same model improved efficiency by 38%.

None of the alternative models showed full agreement with the randomised complete block model in terms of the selection of the top lines. The preferred model showed the poorest agreement with the randomised complete block model among the alternative models. The agreement was poorest at high selection intensities for all models, with the preferred model holding the poorest agreement at all selection proportions. Thus, these methodologies would have an impact on selection decisions in plant breeding. The option for plant breeders to apply the advanced designs and analytical methods will depend on their modelling proficiency, time availability, importance of the selections, precision required and breeding stage. Further work is underway to evaluate the impact

of these alternative selection decisions in an independent series of trials.

Acknowledgments The contributions of the field trial staff at the Leslie Research Centre, Toowoomba, and the Plant Breeding Institute, Narrabri, are gratefully acknowledged. The senior author is greatly indebted to his colleague A/Prof. Shu Fukai for valuable comments on the manuscript. This research was funded by the Grain Research Development Corporation.

References

- Ainsley AE, Paterson LJ, Patterson HD (1985) A method for predicting the efficiency of incomplete block trials. *Biometrics* 43:55–59
- Basford KE, Williams ER, Cullis BR, Gilmour A (1996) Experimental design and analysis for variety trials. In: Cooper M, Hammer GL (eds) *Plant adaptation and crop improvement*. CAB Int, Wallingford, UK, pp 125–138
- Binns MR (1987) Practical use of neighbour methods and alternatives in the analysis of field trials. *Can J Plant Sci* 67:477–489
- Brownie C, Gumpertz ML (1997) Validity of spatial analyses for large field trials. *J Agric Biol Environ Stat* 122:1–23
- Cochran WG, Cox GM (1957) *Experimental designs*, 2nd edn. John Wiley & Sons, New York
- Cooper M, Fox PN (1996) Environmental characterisation based on probe and reference genotypes. In: Cooper M, Hammer GL (eds) *Plant adaptation and crop improvement*. CAB Int, Wallingford, UK, pp 529–547
- Cullis BR, Gleeson AC (1989) Efficiency of neighbour analysis for replicated variety trials in Australia. *J Agric Sci* 113: 233–239
- Cullis BR, Gleeson AC (1991) Spatial analysis of field experiments—an extension to two dimensions. *Biometrics* 47:1449–1460
- Everitt BS (1993) *Cluster analysis*, 3rd edn. Halsted Press, New York
- Federer WT (1998) Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle design experiments. *Biometrics* 54:471–481
- Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for natural and extraneous variation in the analysis of field experiments. *J Agric Biol Environ Stat* 2:269–293
- Gilmour AR, Cullis BR, Welham SJ, Thompson R (1998) *AS-REML program user manual*, NSW Agriculture, Orange Agricultural Institute, Forest Road, Orange, NSW, Australia
- Gleeson AC (1997) Spatial analysis. In: Kempton RA, Fox PN (eds) *Statistical methods for plant variety evaluation*. Chapman and Hall, London
- Gleeson AC, Cullis BR (1987) Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics* 43:277–288
- Grondona MR, Crossa J, Fox PN, Pfeiffer WH (1996) Analysis of variety yield trials using two-dimensional separable ARIMA processes. *Biometrics* 52:763–770

- Kempton RA, Seraphin JC, Sward AM (1994) Statistical analysis of two-dimensional variation in variety yield trials. *J Agric Sci* 122:335–342
- Lill WJ, Gleeson AC, Cullis BR (1988) Relative accuracy of a neighbour method for field experiments. *J Agric Sci* 111:339–346
- Magnussen S (1990) Application and comparison of spatial models in analysing tree genetics field trials. *Can J Forest Res* 20:536–546
- Nei M, Li WH (1979) Mathematical model for structuring genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76:5269–5273
- Nguyen NK, Williams ER (1993) An algorithm for constructing resolvable row-column designs with $v \leq 100$. *Aust J Stat* 35:363–370
- Patterson HD, Hunter EA (1983) The efficiency of incomplete block designs in National List and Recommended List cereal variety trials. *J Agric Sci* 101:427–433
- Robinson DL, Kershaw CD, Ellis RP (1988) An investigation of two-dimensional yield variability in breeders' small plot barley trials. *J Agric Sci* 111:419–426
- Snijders TAB, Dormaar M, van Schuur WH, Dijkman-Caes C, Driesen G (1990) Distribution of some similarity coefficients for dyadic binary data in the case of associated attributes. *J Class* 7:5–31
- Williams ER, John JA (1989) Construction of row and column designs with contiguous replicates. *Appl Stat* 38:149–154
- Williams ER, Matheson AC (1994) Experimental design and analysis for use in tree improvement. CSIRO Information Services, East Melbourne, Australia
- Yates F (1939) The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Ann Eugen* 9:135–156
- Yau SK (1997) Efficiency of alpha-lattice designs in international variety yield trials of barley and wheat. *J Agric Sci* 128:5–9
- Zimmerman DL, Harville DA (1991) A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics* 47:223–239