

Improving the efficiency of pasture breeding programs through the use of spatial analysis

B. D. Braysher¹, B. S. Dear¹, S. Boschma², B. R. Cullis¹ and G. A. Sandral¹

¹NSW Agriculture, Agricultural Institute, PMB, Pine Gully Rd, Wagga Wagga, NSW.

²NSW Agriculture, Tamworth Crop Improvement Centre, Tamworth, NSW.

Abstract

This paper demonstrates the improved ability to partition environment and treatment effects in cultivar selection programs using spatial analysis. Spatial analysis of a trial comparing 50 lines of *Medicago polymorpha* identified strong extraneous column effects which could not be adequately accounted for by analysis of variance (ANOVA). When compared to results from a traditional ANOVA, the spatial analysis, including adjustments for these effects, markedly changed the ranking of the lines, reduced the residual error variance and increased the ability to identify superior genotypes. The results demonstrate the potential of spatial analysis to increase the efficiency of pasture breeding and selection programs, particularly where there is significant extraneous variation masking treatment effects.

Key words

Spatial analysis, variogram, trend, plot correlation, autoregression.

INTRODUCTION

Spatial variation within a paddock or glasshouse can often be so large that it may mask or distort real differences between treatments. In the case of plant improvement programs it may reduce the ability to identify superior genotypes and release new cultivars, and increase the chance of misclassifying a superior genotype as inferior and vice versa. The use of spatial analysis (1) to account for trends in one or two directions is a valuable statistical technique to account for systematic variation in plant improvement trials. Although the technique has been widely adopted in crop breeding programs, its use in pasture breeding programs is still at an embryonic stage. This paper demonstrates the potential improvements of using spatial analysis. As an example an experiment selecting improved lines of *Medicago polymorpha* is used.

MATERIALS AND METHODS

Experiment 1

The trial was sown in 1999 at Coonamble, in northern New South Wales, and the treatments consisted of 50 *Medicago polymorpha* (burr medic) lines. Two additional occurrences of the control varieties (lines 49 and 50) were sown in each replicate, as these were commercial standards. The plots were arranged in a 27 row by 6 column layout. The plot size was 1m x 4.5m, with the 4.5m perpendicular to columns. The trial was designed using SPADES (2), with 3 replications, arranged so that a complete replicate occupied two adjacent columns.

The herbage yields for spring 1999 were independently assessed by two people scoring each plot using a score from 0-50. These were calibrated by selecting 20 plots ranging in yield, giving each a visual rating, cutting a 50cm x 50cm quadrat, drying then weighing the sample. The data analysed in this paper are the herbage yields obtained from this calibration converted into kg DM/ha. All models were fitted using the SAMM (3) module in S-Plus.

Analysis

Statistical models

The basis for the analysis of designed field experiments is the additive decomposition of the data (in this case, herbage yield) into separate components. This decomposition is usefully represented by the following symbolic model:

$$Y \sim \text{treatment} + \text{trend} + \text{error} \quad (\text{A})$$

There are two approaches to modelling these types of experiments:

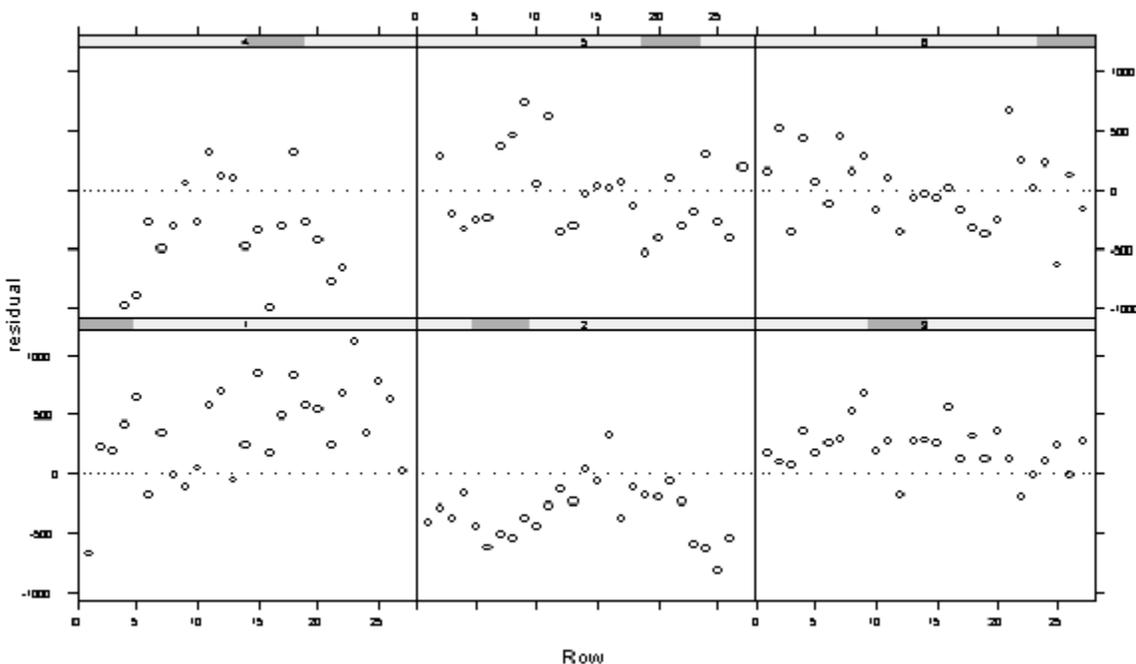
- The classical approach – ANOVA type analysis e.g. Randomised Complete Block (RCB)
- Newer approach – modelling the spatial variation as the trial requires

In the first we implicitly fit a model which contains a predetermined blocking effect, such as a replicate (rep) effect in the case of a RCB analysis. Figure 1 shows a plot of the residuals (data – treatment = trend + error) for these data. Recall that the trend model which matches the design assumes that the mean level of column 1 and column 2 is the same and constant along the column, the mean level of column 3 and column 4 is the same (but may be different to the mean level of column 1 and 2) etc.

This is clearly an erroneous assumption for these data. Thus the design-based analysis may be misleading if a variety ‘unluckily’ occurred in column 2 (rep 1) and column 4 (rep 2), say.

What spatial analysis (1) does, in contrast to the RCB analysis, is to use a modelling approach in which we attempt to identify, then model the extraneous variation or ‘trend’.

Figure 1. Columnwise plot of residuals from Randomised Complete Blocks (RCB) model for the Coonamble *M. polymorpha* trial.



Results

Three models were fitted to the data from the trial in our example. The first model was a RCB model fitting a random replicate effect as the ‘trend’ term in equation (A). The second model was the base spatial model, which consists of a first order autoregressive correlation term in both the row and column directions, to allow for correlation between adjacent plots. The plot of the residuals from the first model are presented in Figure 1 and the sample variogram (1) corresponding to the second model in Figure

2(a). These clearly indicate that a substantial amount of extraneous variation could be accounted for by including a column effect. Discussions with the researchers indicated that this was reasonable given previous paddock history. The base spatial model was thus updated to include a column effect. Any remaining variation should reflect local, or small scale variation which can be adequately modelled using the first order autoregressive correlation (1) terms that were contained in the base spatial model. The variogram in Figure 2(b), which corresponds to the final model does not show any major extraneous or global trends.

In Table 1, there is a comparison of the log-likelihoods (1) for each of the three models. This is a means of detecting whether the addition of random terms to a nested model have had a significant effect. The figure is related to the residual sum of squares. Model 1 (the RCB model) is not comparable to model 2 or 3 because the models are not nested i.e. models 2 and 3 do not contain a replicate effect. It shows that the addition of the column effect increases the likelihood by almost 5, far greater than the critical value for 1 degree of freedom, meaning that it has made a significant improvement.

Figure 2. Variograms showing residuals for (a) the base spatial model and (b) the final spatial model, with column effects added.

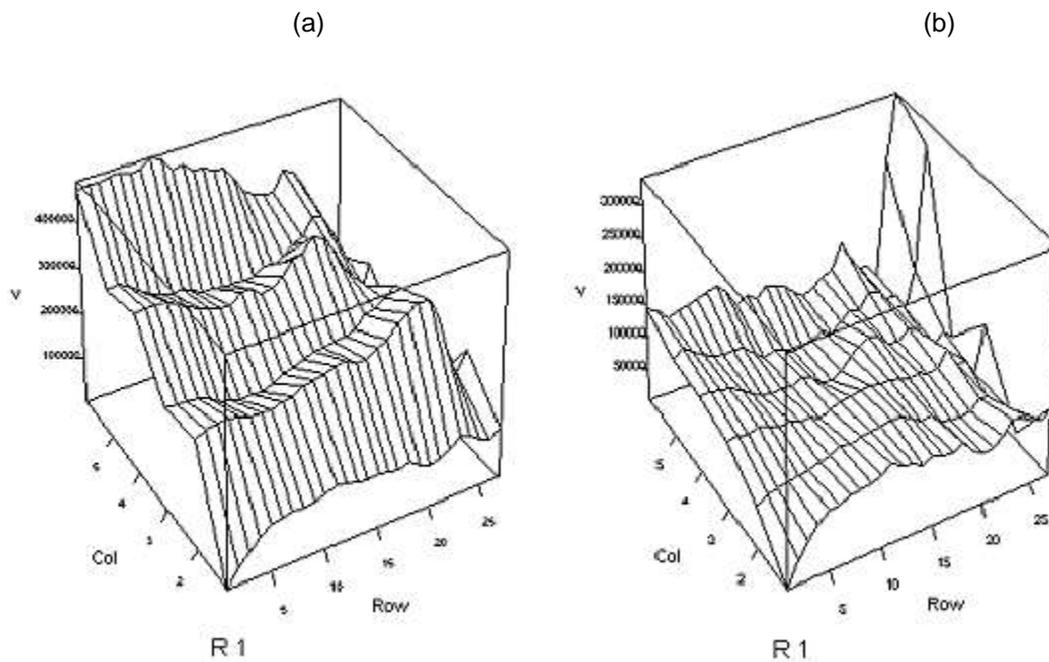


Table 1. Table of log-likelihoods for each model fitted. Models separated by a double line are not comparable.

Model no.	Model	No. of variance parameters	Log-likelihood	Change
1	RCB (treatment + rep)	2	-712.43	
2	Base spatial model: row and column correlation terms fitted	3	-675.57	

DISCUSSION

Analysis of the data using RCB analysis failed to adequately account for the extraneous effects within the paddock. The use of spatial analysis was however able to adjust for these trends and remove a greater proportion of the variation due to non-treatment effects.

The 16 top ranked lines from the RCB analysis and spatial analysis are presented in table 2. Accounting for spatial variation substantially changed the ranking of the treatments, for example moving line 27 from 4th to 1st position and line 45 from 13th to 3rd. Line 26 dropped from 2nd to 10th and line 9 dropped from 5th to 16th. The average standard error of differences was almost halved, from 417.5 to 234.4.

In our example it is possible that lines 26 and 9 may have been falsely promoted and line 45 may have been wrongly ignored. In general there were marked differences between the ranking of the RCB treatment means and the Spatial Analysis treatment means. It transpired that line 26 occurred in columns 1, 3 and 6 and line 45 occurred in columns 2,4 and 6. Recall from Figure 1 that columns 1 and 3 had high residuals and columns 2 and 4 had low residuals.

Spatial analysis has for some time now been used in crop breeding for variety selection, but has not been widely adopted in pasture breeding or agronomy. The benefits of spatial analysis will be greatest where paddock trial sites are subject to variation in weed densities, stubble residues or spray drift, or where experimental methodology such as harvesting direction or changes in sampling technique have unintentionally induced trends within a trial. However, it is still preferable if these types of problems causing extraneous trends are avoided.

Table 2. The mean herbage yields (kg DM/ha) of the top 16 *M. polymorpha* lines following Randomised Complete Block analysis versus Spatial Analysis at the Coonamble trial.

<i>Medicago</i> line	RCB Mean	<i>Medicago</i> line	Spatially Adjusted Mean
11	2635	27	2382
26	2545	11	2359
38	2409	45	2253
27	2317	38	2238
9	2219	50	2199
33	2204	33	2091
39	2179	34	2044
50	2151	2	2028

3	2129	16	2004
32	2055	26	2004
1	2044	39	1964
43	2036	4	1909
45	2011	37	1905
37	1908	32	1903
42	1867	42	1887
46	1863	9	1860
Av SED	417.5	Av SED	234.4

Large adjustments in the ranking of lines cause a loss of belief in the data. To avoid larger adjustments it is therefore preferable to use a design that is more robust to the types of variation often present in field trials. For example, a row-column design with neighbour balance would perhaps have been more suitable. Current research is underway to incorporate this facility into SPADES (2). An improved design which matches the trend more closely would leave the ranking of the treatment means unchanged as well as improving the precision of the analysis i.e. decreasing the average standard error of differences and the overall residual error variance.

References

1. Gilmour, A.R, Cullis, B.R., Verbyla, A.P. (1997). Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments, *Journal of Agricultural, Biological and Environmental Statistics*, **2**, 269-293
2. Coombes, N.E., Gilmour, A.R. (1999). SPADES reference manual, NSW Agriculture Biometrics Department.
3. Butler, D.G., Gilmour, A.R. (1999). SAMM reference manual, Queensland Department of Primary Industries.