# Sugar Research & Development Corporation
# Final Report

| | |
|---|---|
| **Project No** | UQ023 |
| **Title** | Optimisation of experimental design and analysis for variety trials to maximise genetic gain |
| **Restriction** | Normal/Unrestricted |
| **Project Start** | 1 July 1996 |
| **Project Completion** | 30 March 2001 |

## Administration


| | | | |
|---|---|---|---|
| **Name** | Mr Kerry Johnston | Phone | 3365-7493 |
| **Position** | Administrative Officer | Fax | 3365-4455 |
| **Organisation** | The University of Queensland | | |
| **Address** | Research and Postgraduate Studies<br>The University of Queensland | | |
| **Town/Suburb** | Brisbane | | |
| **State** | QLD 4072 | | |


## Chief Investigator(s)


| | | | |
|---|---|---|---|
| **Name** | Professor Kaye Basford | Phone | 3365-2810 |
| **Position** | Head, School of Land and Food Sciences | Fax | 3365-1177 |
| **Organisation** | The University of Queensland | | |
| **Address** | The School of Land and Food Sciences<br>The University of Queensland | | |
| **Town/Suburb** | Brisbane | | |
| **State** | QLD 4072 | | |

| | | | |
|---|---|---|---|
| **Name** | Dr Mac Hogarth | Phone | 3331-3316 |
| **Position** | Manager, Special Projects | Fax | 3871-0383 |
| **Organisation** | Bureau of Sugar Experiment Stations | | |
| **Address** | PO Box 86<br>Indooroopilly | | |
| **Town/Suburb** | Brisbane | | |
| **State** | QLD 4068 | | |

## Funding

| Milestone | Date | SRDC Funding | Task Description |
|---|---|---|---|
| 1 | 1/7/96 | 25,000 | Sign agreement; appoint PhD student |
| 2 | 1/12/96 | 14,955 | Understand current methods |
| 3 | 1/5/97 | 15,000 | Undertake preliminary analysis |
| 4 | 1/12/97 | 15,955 | Model inter-plot competition* |
| 5 | 1/5/98 | 14,955 | Include spatial variation# |
| 6 | 1/12/98 | 14,000 | Investigate early generation trials |
| 7 | 1/5/99 | 11,500 | Hypothesis testing and model selection |
| 8 | 30/3/00 | 1,000 | Completion of project and report |
| | | **$112,365** | |

*****The later milestones were revised.
**#**The original project was extended for six months with an additional $12,500 paid in the 1999/2000 financial year. The date for the completion of the project was extended to 30/3/01 (without funding) in the expectation that Vince Matassa's PhD thesis would be submitted by that time. Although most chapters are in draft form, it is yet to be finalised.

## Confidentiality

While this report is available for the purposes of academic study, it contains material that remains the intellectual property of Vince Matassa as it is part of his PhD thesis to be submitted to the University of Queensland (UQ). The Bureau of Sugar Experiment Stations (BSES) and UQ will be joint owners of the remaining intellectual property arising from this project.

## Disclaimer

The Research Organisation is not a partner, joint venturer, employee or agent of SRDC and has no authority to legally bind SRDC, in any publication of substantive details or results of this Project.

## Abstract

The analysis of field experimental data usually assumes independent observations. This method is often inappropriate if the data contain strong spatial trends. Methods of analysis (commonly used in cereal trials) which account for the spatial variability existing within a field were shown to improve the precision of the estimates of clonal effects in sugarcane variety trials. As inter-plot competition was a major contributing factor biasing clonal estimates, a model which accounted for the competitive interaction between neighbouring clones was developed. Both factors could be incorporated into a more general model that jointly assessed inter-plot interference and fertility gradients. Two different estimation procedures (marginal likelihood and profile likelihood) for the parameters in the underlying models were investigated in a simulation study. Based on the criteria of average bias and mean-squared error, there was a slight preference for the marginal likelihood estimators. An empirical investigation of the relative efficiency of the augmented randomised complete block design versus the spatial unreplicated design in early generation sugarcane breeding programs was also undertaken. This indicated that the spatial design was to be preferred with respect to the bias in the parameter estimates, but using other criteria for design evaluation there were very few advantages in choosing the more complicated spatial design.

## Non-technical Summary

The analysis of field experimental data usually assumes independent observations. This method of analysis is often inappropriate if the data contain strong, unaccounted for spatial trends. More recently, methods of analysis have been proposed that account for the spatial variability existing within a field experiment. They have been commonly utilised in cereals but rarely applied to perennial crops such as sugarcane. This method of analysis was demonstrated on a replicated sugarcane trial conducted within the Bureau of Sugar Experiment Stations (BSES) breeding program. The improvement in precision of the estimates of clonal effects relative to those for the classical analysis of variance (currently adopted by BSES) was illustrated, as were simplistic diagnostic tools to identify a suitable model for the environmental gradient within such a trial.

It was realised, by the spatial analysis of existing early generation BSES sugarcane breeding trial data, that inter-plot competition is a major contributing factor biasing clonal estimates. Thus in order to maximise genetic advance, an analysis of these trials taking account of the competitive effects between the neighbouring clones was required. A model accounting for the competitive interaction between neighbouring clones was developed. The application and robustness of this model to existing data was investigated and the benefits of this proposed model over existing methods of analysis were assessed.

Classes of models of specific relevance for the analysis of sugarcane field experiments were then outlined and derived from the specification of a more general model. Three categories of models were specified, namely models that assess inter-plot interference or soil fertility gradients only and models that jointly assess the presence of both inter-plot interference and fertility gradients. The marginal and profile likelihood estimation procedures, for the parameters of interest for the general model specification, were defined, along with the associated scores, and information matrix. Techniques of estimation for the parameters of these trend, competition and competition plus trend models were evaluated in a series of computing simulation experiments. Based on the criteria of average bias and mean-squared error of the parameter estimates, there was a slight preference for the marginal likelihood procedure.

Meanwhile, an empirical investigation of the relative efficiency of the augmented randomised complete block design versus the spatial unreplicated design in early generation sugarcane breeding programs was investigated by a series of simulation experiments. The study compared, by various criteria, the consequences of adopting either the augmented or the spatial design on the outcomes of selection. The validity and efficiency of the clonal estimates, as determined by the method of spatial analysis on simulated data, was also addressed for the both the designs. This indicated that the spatial design was to be preferred with respect to the bias in the parameter estimates, but using other criteria for design evaluation there were very few advantages in choosing the more complicated spatial design.

## Background of the research project

While BSES spends considerable resources annually in assessing the performance of clones in unreplicated early-stage trials, it was unclear if the best possible statistical designs were being used and the most appropriate methods of analysis were being applied to the resulting data. The success of a breeding program depends heavily on early-stage trials as superior clones must be identified for inclusion in later-stage selection trials. On the other hand, later-stage trials are characterised by a large number of clones in each replicate and it is difficult to maintain uniformity among experimental units because of soil variability. This heterogeneity is a major factor affecting the efficiency of selection and it is important to consider designs to take this into account. It is important at every stage of selection that promising clones are not discarded from the breeding program.

A primary concern of sugarcane field experimenters is to obtain accurate estimates of clonal means and clonal differences. Local variations in fertility gradients and the presence of inter-plot interference in sugarcane experiments are major contributing factors that can bias variety differences and inflate the estimate of the residual variation. More importantly, both these factors cause associations (or dependencies) between neighbouring plots, for which a classical method of analysis is inappropriate. Therefore, in order to maintain genetic progress, it becomes important to eliminate the bias of the clonal estimates caused by these effects, by perhaps imposing designs of increasing complexity, such as spatial or interference designs, or by the use of statistical modelling tools.

Methods of analysis that model local variations in plot values due to either fertility gradients (nearest neighbour or spatial methods) or direct interference between plots are appearing in the statistical literature. Although these methods can provide important gains in precision for variety contrasts over classical methods, most account for, or make adjustments to, variety contrasts for either fertility gradients or interference between neighbouring plots but not both. Little research effort has focused on formulating statistical models that jointly account for spatial correlation amongst neighbouring plots due to fertility gradients and inter-plot interference. They need to be investigated.

A commonly adopted design for early generation sugarcane field experiments by sugarcane breeders at the Bureau of Sugar Experimental Stations (BSES) in Australia is the augmented design. For such a design approximately 10% of the total number of plots are allocated to check clones which are randomly arranged in a standard complete or incomplete block design, augmented with a number of unreplicated experimental clones. Recently, the use of spatial unreplicated designs for the evaluation of unreplicated and repeated check varieties in plant breeding programs has been proposed in the literature. These designs have been shown to have high efficiency when using the minimum average variance of the contrasts between the unreplicated and the repeated check varieties as an optimality criterion for design selection. An investigation into the relative efficiency of the conventional, namely the augmented randomised complete block design, against the spatial unreplicated design in the BSES early generation sugarcane breeding program is of importance to the sugarcane industry.

## Original objectives

1. To obtain more accurate estimates of clonal performance in unreplicated early-stage trials by evaluating a range of statistical techniques,
2. To determine whether to use replication routinely in early-stage trials to maximise genetic gain, and
3. To investigate alternative incomplete block designs to alpha and lattice designs to provide an improved method for adjusting for soil variability in later stage trials.

As with all research, the results of the initial investigations affected the subsequent direction of the project. Thus the second and third objectives were modified after the first year (and successful completion of the first three milestones), when it became apparent that more effort was needed to determine appropriate designs for and analysis of unreplicated early-stage sugarcane trials. Hence it seems best to list the refined objectives, the methodology for their investigation, and the subsequent results.

## Refined objectives

1. To obtain more accurate estimates of clonal performance in unreplicated early-stage trials by evaluating a range of statistical techniques,
2. To assess whether the uses of single-row plots in sugarcane variety experiments are affected by inter-plot competition using statistical modelling techniques,
3. To investigate the relative efficiency of a conventional design, namely the augmented randomised complete block design, against the spatial unreplicated design in the BSES early generation sugarcane breeding program,
4. To define and develop models that fall into three categories, namely models that assess inter-plot interference or soil fertility gradients only and models that jointly assess the presence of both inter-plot interference and fertility gradients in sugarcane variety trials, and
5. To evaluate both a conventional and a new likelihood-based estimation procedure, for classes of models of specific interest to field experimenters in the sugarcane industry.

## Methodology (objective 1)

The application of spatial analysis techniques to field data has made a major contribution to the elimination and detection of unforeseen trends. This method of analysis is commonly utilised in cereals but rarely applied to perennial crops such as sugarcane. The method of spatial analysis was investigated for sugarcane and demonstrated on data from a replicated field experiment.

## Results and discussion (objective 1)

Spatial analysis of the chosen example resulted in identifying extraneous, global and natural trend that otherwise would have been unnoticed if the more classical method of analysis which is commonly adopted by BSES was performed (Matassa *et al*., 1998). By applying this newer procedure, there was a considerable decrease in the magnitude of the error variance relative to the classical analysis. Consequently, the clonal effects were estimated with greater accuracy and precision.

The spatial modelling approach adopted used the variogram and plots of the trend either in the row or column direction. We recommend that these graphical diagnostic tools be used to assist in the selection of an appropriate model for the description of trend. The strategic framework of spatial analysis detailed in the literature and utilised in this work is recommended for the spatial analysis of sugarcane experiments.

This part of the research demonstrated spatial methods for the analysis of sugarcane trials. Although the potential benefits and advantages of utilising this procedure are well documented in other crops, its use in sugarcane had been limited. This seemed to be partly due to the complexity of the methodology and the previous unavailability of user-friendly software. These problems have been overcome and the spatial analysis of field experiments can now be performed by the use of the program ASREML (Gilmour *et al*., 1996). The simplicity in fitting different spatial models and the display of various graphical diagnostics enables the routine analysis of sugarcane experiments with this methodology.

## Methodology (objective 2)

An experiment was conducted to assess whether single-row plots in sugarcane varietal experiments were affected by inter-plot competition. The competition model using neighbouring plot yields as a measure of competitive ability was fitted to both the entire experiment and sub-sections of it to provide evidence as to whether the observed plot yield in single-row plot data was influenced by its surrounding neighbours. For a particular chosen clone, estimates of clonal performance derived from this model were evaluated and compared to the mean value of the central two rows of six-row plots. Precision in detecting real clonal differences was investigated.

## Results and discussion (objective 2)

Results showed that the bias caused by inter-plot competition in single-row plot trials can be easily quantified by a statistical modelling approach (Matassa *et al*., 1999).

Furthermore, results suggested that adopting experimental designs that yield better precision amongst clonal comparisons, such as spatial designs, will result in a more accurate estimate of the competition coefficient and predicted pure stand yields.

It was shown that competitive effects amongst clones are present for clones evaluated in single-row plots. While plot yield is a useful measure of competitive ability in sugarcane experimental data, the use of other trait information, such as height of the cane, should not be ignored.

This part of the project evaluated the efficiency of a competition model on sugarcane single row plot data and indicated a moderate improvement in the predicted pure stand estimates of the clonal yields over the unadjusted clonal mean yields. However, it appeared that greater precision in the estimates of clonal effects would be obtained if the competition model was also able to take account of environmental trend gradients that spatial methods were able to detect. This indicated that further research was required.

Nevertheless, the statistical analysis of single-row sugarcane data should be examined for the presence of inter-plot competition by the modelling approach detailed in Matassa *et al*. (1999). When evidence of inter-plot competition is found in such trials, adjustments should be made to the clonal effects. Selection should then be carried out on these adjusted values.

## Methodology (objective 3)

Six different experiments were considered in a simulation study to evaluate the robustness of the augmented against a spatial design under various data generating schemes. In this study, the chosen designs were ones that researchers would find sensible given the constraints of the experimental field conditions and the total number of replicated and unreplicated clones available for evaluation. Consequently, the replicated (check) clones were placed in the simulated field layout in such a way that they, firstly, satisfied the theoretical optimality criterion given in the literature for the spatial design, and secondly, were arranged in a reasonable manner given the blocking structure of the augmented design. The study compared the consequences of adopting either design on the outcomes of selection. The precision and accuracy of the clonal estimates, as determined by spatial analysis of the simulated data, were also addressed for both designs.

## Results & Discussion (objective 3)

Based on various criteria for design comparison, the simulation results showed that there was little difference between the spatial designs and the augmented design for the experimental situation considered (Matassa and Basford, 2000, Matassa *et al*., 2001a). The exception was for average efficiency where there was a gain in efficiency when using spatial designs for estimating comparisons among clonal effects. However, it should be noted that the augmented design did possess good neighbour properties and was near optimal. Further, given the number of control treatments (replicated treatments) and the size of most unreplicated experiments, it follows that the ‘amount’ of possible neighbour structure is quite limited.

Nevertheless, the spatial designs offer theoretical advantages over the augmented design as their properties involve the neighbour balance structure as well as the placement of the repeated check varieties in the field. Uncontrolled fertility gradients in early generation sugarcane trials are common and hence the sensible arrangement of the check clones in order to estimate the error variance with minimal bias is important. Spatial designs specifically arrange check varieties based on the assumed underlying correlation structure.

Further investigation into the efficiency of the augmented against the spatial design is required before a final recommendation can be made for the routine use of the spatial designs for the evaluation of a large number of unreplicated experimental clones for sugarcane breeding programs. More importantly, the performance of other augmented design plans against the spatial design needs investigation.

## Methodology (objective 4)

The method of spatial analysis, generally associated with geostatistics and random field theory, has primarily been focused on modeling local trend gradients within a field experiment by considering (covariance) models for the plot error components as a realization of a continuous spatial random function. Estimation procedures based on variants of the maximum likelihood procedure (e.g. the marginal likelihood and profile likelihood functions), weighted and generalized least squares have been suggested for estimating the plot error covariance parameters when the observed plot responses are expected to behave as a continuous spatial process. However, as demonstrated by various simulation studies, bias can be considerably reduced if the estimation is based on the marginal likelihood, rather than these alternative procedures of estimation.

Nearest neighbour methods, an extension of (discrete) time-series theory in a spatial context, account for the correlation in the plot error components indirectly, either by differencing or by using the plot error components from neighbouring plots as a concomitant variable. The full maximum likelihood and the profile (adjusted) likelihood functions have been suggested for estimating the parameters for these models, but have received relatively little attention.

Several different models have been suggested in the literature to assess the presence of inter-plot interference in field experiments. Models that contain a spatially lagged dependent variable, i.e. models that take into account overlap (or treatment drift) effects from adjacent plots, have been considered. Although the profile likelihood function is usually employed for the estimation of parameters in these (non-linear) models that contain the same parameter in the mean and the covariance structure, recent investigations have shown that inefficient and inconsistent parameter estimates can be obtained. These models are structurally equivalent to the class of inter-plot interference models recommended by Besag and Kempton (1986) for the assessment of inter-plot competition in field experiments. Alternative estimation procedures, other than the profile likelihood function, have been recommended for estimating the parameters of such models. These estimation procedures include the marginal likelihood function and methods that make adjustments to either the profile likelihood or profile score function.

Models containing a spatially lagged dependent variable and a spatially autocorrelated error term are well developed and practiced in the fields of geography, regional science and econometrics. Estimation procedures associated with these models, based on the profile likelihood function, marginal likelihood function, instrumental variables and a generalised spatial two-stage least squares procedure have been effectively applied. However, at present, knowledge about these estimation procedures for such models is incomplete and diffuse.

## Results & Discussion (objective 4)

A general parametric model that forms a framework for the modeling of various forms of dependencies arising in agricultural field experiments, namely inter-plot interference and environmental trend gradients, was constructed. By constraining various parameters, three particular sub-models were derived:

1. Models that assess inter-plot interference (competition),
2. Models that assess soil fertility gradients (trend), and
3. Models that jointly assess inter-plot interference and fertility gradients (competition plus trend).

Methods of estimation required to yield asymptotically unbiased and consistent estimates of the parameters of interest (with associated score matrix and information matrix) for these various models were constructed (Matassa *et al*., 2001b).

## Methodology (objective 5)

Classes of models of specific relevance for the analysis of field experiments were outlined and derived from the specification of a more general model. Three main experiments, using either the marginal or profile likelihood functions, for the competition, trend, and competition plus trend models were performed. For each, Monte Carlo simulations were undertaken allowing for different design matrices, i.e. treatment effects only, and block plus treatment effects.

## Results & Discussion (objective 5)

For the competition model, the average biases for the two methods of estimation (marginal likelihood and profile likelihood) were roughly equivalent. The best method varied considerably with the choice of design matrix (block plus treatment or treatment only) and the combination of parameter values. For the model incorporating trend only, the marginal likelihood method appeared to perform better than the profile likelihood when the design matrix included block plus treatment effects. However no best method could be distinguished when the design matrix incorporated treatment effects only.
For the competition plus trend model, there were some conflicting results with perhaps a preference for the marginal likelihood estimation procedure.

Overall, based on the criteria of average bias and mean-squared error of the parameter estimates, there was a slight preference for the marginal likelihood procedure. See Matassa *et al*. (2001b) for full details and discussion.

## Assessment and Recommendations

The recommendations to the sugar industry are as follows:

1. Undertake spatial analysis of data from early generation field trials (with graphical diagnostic tools to assist in the selection of an appropriate trend model),
2. Incorporate inter-plot competition into these models,
3. Use marginal likelihood estimation procedures, and
4. Consider the adoption of spatial designs even though well-constructed augmented designs will be reasonably satisfactory.

The potential benefits to the industry are:

1. More accurate estimates of clonal performance in early-stage trials so that plant breeders are less likely to discard promising clones, and
2. More precise estimates of traits of interest in later stage trials to improve selection efficiency.

## List of publications

Besag, J. and Kempton, R.A. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics* **42**, 231-251.

Gilmour, A., Thompson, R., Cullis, B.R. and Welham, S. (1996). ASREML. Technical Report, NSW Agriculture, Wagga Wagga, NSW, Australia.

*#Matassa, V.J., Basford, K.E., Stringer, J.K. and Hogarth, D.M. (1998). The application of spatial analysis to sugarcane variety trials. *Proceedings of the Australian Society of Sugarcane Technologists* **20**: 162-168.

*Matassa, V.J., Basford, K.E. and Jackson, P. (1999). Intergenotypic competition in single-row plots of sugarcane variety trials. *Proceedings of the Australian Soceity of Sugarcane Technologists* **21**: 234-240.

*Matassa, V.J. and Basford, K.E. (2000). Simulated evaluation of an augmented against a spatial design for early stage selection in sugarcane breeding. *Proceedings of the Australian Society of Sugarcane Technologists*, **22** (poster).

*Matassa, V.J., Basford, K.E., Chan, B. and Eccleston, J. (2001a). Comparison of the augmented and spatial unreplicated experimental designs in sugarcane breeding under a spatial model – a simulation study. Submitted to *Australian and New Zealand Journal of Statistics*.

*Matassa, V.J., King, M. and Basford, K.E. (2001b). Estimation methods of non-linear spatial models for the analysis of agricultural field experiments. (In preparation)

* Material arising from this project.
\# Vince Matassa won the William Kerr Bursary Award for the best student paper at the conference for his presentation of this material.

When the submitted manuscript and the one in preparation are accepted for publication, copies will be forwarded to SRDC. Similarly, a copy of Vince's PhD thesis will be forwarded to SRDC in due course.