**BSES**

*Advancing Sugar*

FINAL REPORT

SRDC PROJECT BS75S

EVALUATION OF METHODS OF ESTIMATING

BREEDING VALUE OF SUGARCANE

PARENTAL CLONES

by

J K STRINGER

SD96006

Principal Investigator:     J K Stringer
Research Officer
BSES
PO Box 86
INDOOROOPILLY Q 4068
Ph : (07) - 3377 3037

**BSES Publication**
**SRDC Final Report SD96006**                              **August 1996**

# GLOSSARY

| | |
|---|---|
| BLUP | Best Linear Unbiased Prediction |
| BLP | Best Linear Prediction |
| OLS | Ordinary Least Squares |
| LSM | Least Squares Adjusted Means |
| GLS | Generalised Least Squares |
| BLUE | Best Linear Unbiased Estimates |
| MM | Marginal Means |
| MINQUE | Minimum Norm Quadratic Unbiased Estimators |
| ML | Maximum Likelihood |
| REML | Restricted Maximum Likelihood |
| NRM | Numerator Relationship Matrix |
| MME | Mixed Model Equations |
| GCA | General Combining Ability |
| SCA | Specific Combining Ability |
| GLMM | Generalised Linear Mixed Model program |
| $H_N$ | Narrow Sense Heritability on a Family Mean Basis |
| GAREML | BLUP statistical program which incorporates Giesbrecht's algorithm to estimate REML variance components |

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# 1    SUMMARY

BSES uses an empirically based index to assess the breeding potential of sugarcane (Saccharum spp.) parents and this may take many years to provide reliable estimates. A rapid and efficient method of assessing the breeding value of parents in early stage families is needed to increase the rate of population improvement.

Early stage family selection trials are typically highly unbalanced and so data analysis cannot be undertaken by ordinary least squares approaches. Statistical techniques such as Best Linear Unbiased Prediction (BLUP) were specifically developed to allow prediction of breeding values from messy dairy cattle data sets. Although the theory could be adapted to other breeding programs, there have been few applications in plants. Hence, it was necessary to review the theory of BLUP to determine its suitability for application to sugarcane data. A discussion on BLUP necessitates a discussion on the related technique of Best Linear Prediction (BLP). These techniques are compared to several fixed effects approaches and the advantages of BLUP and BLP are discussed.

Once the suitability of BLUP for analysing sugarcane data was established, it was necessary to determine if the predicted breeding values based on BLUP differed from current BSES methods. The current empirical method used by BSES is outlined and reasons for investigating alternative methods are given. Family selection data from the southern Queensland breeding program were then used to study three univariate methods of predicting breeding value: the family mean, the BSES empirical method and the Best Linear Unbiased Prediction (BLUP).

Initially data from the 1989 southern Queensland series were analysed assuming a North Carolina design I. BLUP estimates were correlated with the arithmetic mean and all correlations were high regardless of the narrow sense heritabilities.

In subsequent analyses data were analysed assuming an incomplete diallel mating design. Using data from the 1990-1991 series, BLUP estimates of breeding value for 234 parents were calculated. For cane yield, CCS and net merit grade (selection index) the BLUP estimates for each parent were used to predict the performance of 79 families in the 1992 series. Predicted performance from the BLUP estimates was correlated with performance in the 1992 series. The correlations were strong, particularly for relative CCS (0.64).

For net merit grade only, breeding values based on the empirical formula were correlated with the BLUP estimates for the 79 families in the 1992 season. The correlation was moderate (0.48).

The BLUP analyses provided BSES plant breeders with a simple and rapid means of combining data from a wide range of sources to identify superior parents. BLUP appears to be as effective as the BSES empirical method or family mean for identifying superior clones. Given that the empirical method incorporates information from up to ten years whereas the BLUP results were based on few years data, it is probable that BLUPs will increase the rate of population improvement, which would have major benefits for the plant breeding program and the sugar industry.

As BLUP can accommodate complicating factors such as inbreeding, it was important to quantify the level present in BSES family selection data. Inbreeding coefficients for all families except three were less than 0.3. The level of inbreeding was correlated with agronomic performance. Correlations were negligible although future studies on inbreeding are suggested.

## 2    BACKGROUND

The current method of assessing breeding value is empirical and combined several years of selection data, agronomic performance data, and disease ratings into an index. Although most of the data used are objective, some selection is based on subjective estimates of yield.

Each year approximately 300 new clones enter the parental collection. An objective method can be applied to stage 1 families is needed so that inferior parents are identified and removed from the collection. This will help to increase the rate of population improvement.

## 3    OBJECTIVES

* To develop a routine methdology for analysing unbalanced family performance records using BLP or BLUP

* Obtain more precise estimates of breeding value through the application of BLP or BLUP

## 4    INTRODUCTION

Sugarcane is a highly heterozygous clonally propagated crop. Modern commercial sugarcane varieties are complex interspecific hybrids of the genus *Saccharum*. Varieties are polyploid and highly aneuploid, with the number of chromosomes ranging from 100 to 130 (Sreenivasan *et al.* 1987). No simply inherited traits have been demonstrated, and sugarcane improvement by breeding has been based on quantitative genetic theory. Selection is also complicated by the long crop cycle, and the time required before assessment of traits such as ratooning ability.

Identification of superior genotypes to be used as parents is a key component of many plant breeding programs. Sugarcane breeders from BSES use an empirical formula for assessing the breeding value of a parent. This method combines several years of agronomic performance data, breeding information and disease ratings into an index. Although this method is comprehensive, it takes many years to reliably estimate the breeding potential of a parent.

Each year approximately 300 new clones enter the parental collection. A method is needed which can be applied to early stage family selection trials, so that inferior parents can be rapidly identified and removed from the collection. This would result in a more efficient breeding program and thus increase the rate of population improvement.

Data from early stage family selection trials are typically highly unbalanced and so classical statistical approaches such as ordinary least squares (OLS) are not appropriate. Best Linear Unbiased Prediction (BLUP) is a proven technique in animal breeding research for obtaining

precise estimates of breeding value from messy data sets. BLUP allows data from a diverse range of mating designs, relatives, traits and precisions to be combined into a single breeding value for each trait and genotype (White and Hodge 1989).

BLUP has been used extensively in animal breeding research but there are few applications in plants. It was necessary to review the underlying theory of BLUP (section 9 ) to assess its suitability for analysis of sugarcane data. To understand the theory of BLUP, an investigation into Best Linear Prediction (BLP) is essential (section 7). The advantages of these approaches over fixed effects models are discussed in sections 5-7.

Once the suitability of BLUP for analysing sugarcane data was established, it was necessary to determine if the prediction of breeding values from BLUP differed from current BSES methods. The BSES empirical method of estimating breeding value is outlined and reasons for investigating other methods are given (section 12). BLUP is then applied to several sugarcane data sets and problems encountered in its application are outlined (sections 12-14). BLUP estimates of breeding value are compared to current BSES methods (section 14) and recommendations for future research are given (section 17).

## 5    FIXED, RANDOM AND MIXED MODELS

The most frequently used statistical models in the analysis of agricultural data (regression, analysis of variance and analysis of covariance) are characterised by an emphasis on estimation and hypothesis testing and there is strict classification of model parameters into fixed and random effects (Stroup 1989). Fixed effects are those factor levels which are deliberately chosen to be included in an experiment and inferences apply only to these treatments. Each treatment of interest is actually included in the experiment and any repetitions of the experiment must include the same treatments or treatment combinations (Hays 1981). Fixed effects are estimated by taking the differences between the treatment means and grand mean. Random effects are those factor levels which represent a random sample of all possible levels. Different treatment levels may be observed over repetitions of the same experiment. Inferences apply to an entire set of distinct treatments or factor levels, including some not actually observed. Random effects are measured largely through variance component estimation.

Conventional linear model theory has treated the distinction between fixed and random effects as being quite sharp. Subsequently, procedures appropriate for the analysis of a completely fixed or random model are well developed and are extensively documented in standard statistical texts. Most agricultural experiments have both fixed and random effects and hence have a mixed model aspect. However, in the literature they are treated as 'something of occasional importance which create difficulties in data analysis' (Stroup 1989).

Statistics has tended to concentrate heavily on fixed effects models and the subsequent data analysis by OLS. According to Stroup (1989), the most compelling advantage of the mixed model is generalisation of linear model theory (section 9.2). Even in cases where OLS assumptions hold, it is still possible to use the general linear model since, in mixed model theory, fixed effects models are simply special cases and their analysis reduces to usual OLS methodology (Stroup 1989).

In the following section, methods of combining data from multiple progeny tests in order to obtain estimates of breeding value of parents are discussed. Most treat the problem of ranking parents as one of estimating fixed family effects (White and Hodge 1989). Marginal means and weighted least squares are two examples of analytical techniques that treat breeding values as fixed effects to be estimated from the data. An approach widely adopted in animal breeding is to treat the breeding values as random effects to be predicted rather than fixed effects to be estimated. BLP and BLUP are two analytical techniques which treat breeding values as random effects. In addition to requiring predictions of the random breeding values, fixed effects (such as herd, site) need to be estimated. Hence, a mixed model approach is necessary. The most successful application of mixed model methodology in agriculture is Henderson's work on BLP and BLUP in dairy cattle breeding (Stroup 1989).

# 6    METHODS OF COMBINING DATA FROM MULTIPLE PROGENY TESTS

## 6.1    Fixed effects models

One of the simplest approaches to combining data from multiple progeny tests is to calculate a mean across all individuals within a family. This is considered as an estimate of general combining ability (GCA) or half the breeding value of the parent involved (White and Hodge 1989). Family means may be expressed as deviations from the site mean or a genetic standard. This can adjust for differences among sites but breeding values will be biased if families are not represented by the same number of observations in all blocks. To overcome this problem, family means are often expressed as the *means of the plot means*. This involves finding the average for a family in a given block and then averaging across blocks. When families are equally represented in all blocks the mean of the plot means is equal to the least squares adjusted mean (LSM) for that family (Freund and Littell 1981; Milliken and Johnson 1984).

When families do not appear in all blocks, LSMs can be used to calculate unbiased estimates of family effects. LSMs are the family marginal means that would be expected had the design been balanced. Least squares estimates assume that there is no family by environment interaction and that each progeny test has the same precision. If the number of replications varies from one site to another, then the method of least squares will produce biased estimates as each progeny test will not have the same precision. If families are ranked on the basis of their mean values then those with the highest and lowest score will generally be represented in the fewest number of tests (White and Hodge 1989).

In the method of *weighted least squares* or *generalised least squares* (GLS), data are weighted in inverse proportion to the variance for the site. As a result, family means from precise progeny tests have a greater influence on the estimate of family effects or GCA than those from imprecise tests (White and Hodge 1989). Weighted least squares analysis accounts for tests with varying precision and replication and differences in site means.

GLS estimates of fixed effects are more efficient than corresponding OLS estimators (Harville 1984). Hence, this allows for more precise information from 'conventionally-sized' experiments or an equivalent amount of information from experiments with fewer observations (Stroup 1989).

Another method which takes into account different test precisions is *shrunken least squares*. Estimates of family effects are `shrunk' or regressed towards the grand mean by an amount which depends on the replication of each family and the overall variance (Cotterill *et al.* 1983). Family effects for poorly replicated families are shrunken closer to zero than those from well replicated families.

Differences in environmental conditions among sites may result in the site mean being correlated with the variance. This relationship is undesirable as this would lead to family, site and block effects being multiplicative rather than additive (Cotterill *et al.* 1983). A log transformation may remove the relationship between mean and variance and achieve additivity. However, a problem with any type of transformation is how to interpret the results after back transformation.

To alleviate the problem of unequal variances across sites, the family means can be converted to standard normal deviates. These standardised scores are then averaged across sites. Since, in the standardisation process, different weights are applied, this may prevent a higher percentage of families being selected from sites with larger variances (Cotterill *et al.* 1983).

## 6.2    Random effects models

The techniques described above treat breeding values as fixed effects to be estimated from the data. The inferences from these techniques only apply to the set of parents present in the study and not to the broader population (White and Hodge 1989).

An approach used extensively in animal breeding is to consider the breeding values as random effects to be predicted rather than fixed effects to be estimated (White and Hodge 1988). Two analytical techniques which treat breeding values as random effects are BLP and BLUP. In BLP, fixed effects can be accurately estimated by simple approaches such as arithmetic averages. However, when the fixed effects need to be estimated simultaneously with the random effects, then the technique of BLUP will be more appropriate.

## 6.3    Fixed vs random effects models

When data from multiple progeny tests are balanced and equally precise for all sites and all parents, the fixed and random analytical procedures will result in the same genotypic rankings (White and Hodge 1989). BLP and BLUP can use information from related experiments to obtain improved estimates of variances. This has important efficiency implications for research conducted over multiple locations on a long-term basis (Stroup 1989). BLP and BLUP also allow greater flexibility in analysis. Data from a diverse range of mating designs, relatives, traits and precisions can be combined into a single breeding value for each trait and genotype (White and Hodge 1989). It is in these 'messy' data situations that BLP and BLUP have many desirable properties.

### 6.4   Scope of theory

The theory of BLP and BLUP can be quite complex but to simplify the mathematics, only additive genetic models will be considered in subsequent sections. It is also assumed that the parents are unrelated, non-inbred and from unselected populations. In section 10, methods of accommodating these complications in the BLUP model are discussed.

## 7   BEST LINEAR PREDICTION

### 7.1   Advantages of BLP

The theory of BLP was developed by Charles Henderson in the 1960s. It was used extensively within the dairy industry until the steady increase in computing power meant that the more computationally difficult technique of BLUP could be considered. It has now been superseded by BLUP within the dairy industry. Very few applications of BLP were found outside of the animal breeding literature although it has gained some acceptance in the forestry area.

BLP is essentially a selection index procedure which was first developed by Hazel (1943) although there are important differences between the two methods of estimating breeding values. In the selection index approach it is assumed that for each parent there is equal quantity and quality of data and so a single vector of coefficients is applied to all parents (White and Hodge 1989). In direct contrast to this is the BLP approach, in which a different set of weights is developed for each parent (section 7.2) which takes into account the replication and reliability of the data for each genotype (White and Hodge 1989). Hence, in BLP, the error variance of the predicted breeding values is minimised.

BLP also exhibits superiority over OLS procedures and this can be seen by examining how it handles various sources of complications in progeny data which include variability in precision among tests, tests of varying ages due to different planting dates and unequal replication of families across sites (White and Hodge 1988). In BLP, the variance among predicted breeding values increases as parents are in more precise and larger number of tests. As the quality and quantity of information increases, the error variance of prediction decreases and BLP predictions of breeding value tend to cover a wide range. They are not shrunk or regressed towards the mean as the heritabilities for each family will be high. Hence in BLP, a large proportion of the higher ranking parents will be based on higher quality and quantity of data (White et al. 1986). This could be a problem when new parents are introduced in a testing program. Unless they are superior then they will have difficulties competing with a parent that has been in many tests.

The opposite occurs with least squares estimates such as marginal means. When parents have been in very few progeny tests, the marginal means estimate of breeding value will have a large variance. Thus, if marginal means were used to rank parents there would be a tendency for those identified as superior to have been represented in the fewest number of progeny tests (Quaas et al. 1984).

The advantage of BLP over marginal means can easily be seen by examining Figures 1 and 2. Figures 1 and 2 (from White and Hodge 1989) are based on hypothetical data from balanced open pollinated loblolly pine progeny tests.
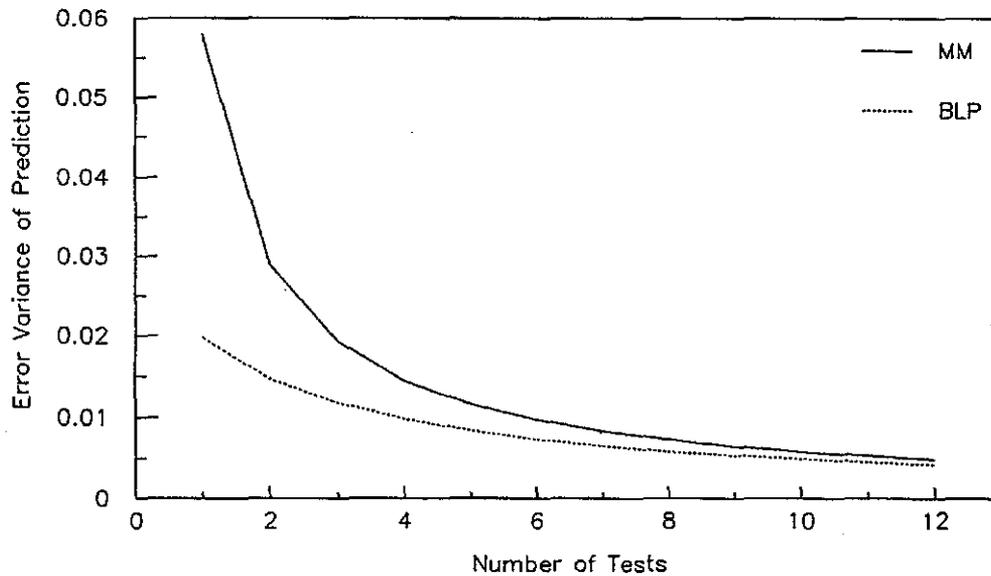


**Fig. 1.** Error variance among breeding values predicted from Best Linear Prediction (BLP) and Marginal Means (MM)

BLP predictions always have lower error variance than those based on marginal means (Fig. 1). Also the marginal mean estimates of breeding value have least precision (Fig. 1) when they have the largest spread (Fig. 2). BLP estimates of breeding value are in direct contrast to this. With BLP, as the number of tests increases the predictions have a smaller error variance (Fig. 1) and there is a larger spread among predictions (Fig. 2). For the data in Figures 1 and 2, it can be seen that when all parents have been in at least 4 or 5 tests, both the error variance and variance among predictions becomes relatively constant and the two methods result in similar parental rankings with comparable precision (White and Hodge 1989).

## 7.2    Theory of BLP

The objective of BLP is to use **y**, an n x 1 vector of observations, to predict **g**, a q x 1 random vector of breeding values for each of the s parents and q traits (White and Hodge 1988). It is assumed that the vector **y** is distributed jointly with **g**. As the exact form of the joint distribution is never known, all that is assumed in the derivation of BLP is that the following fixed effects and variance-covariance matrices are known (White and Hodge 1989).

E (**g**)          =          $\tau$ :  a q x 1 vector of the expected values of the breeding values being predicted. This is often assumed to be equal to zero.
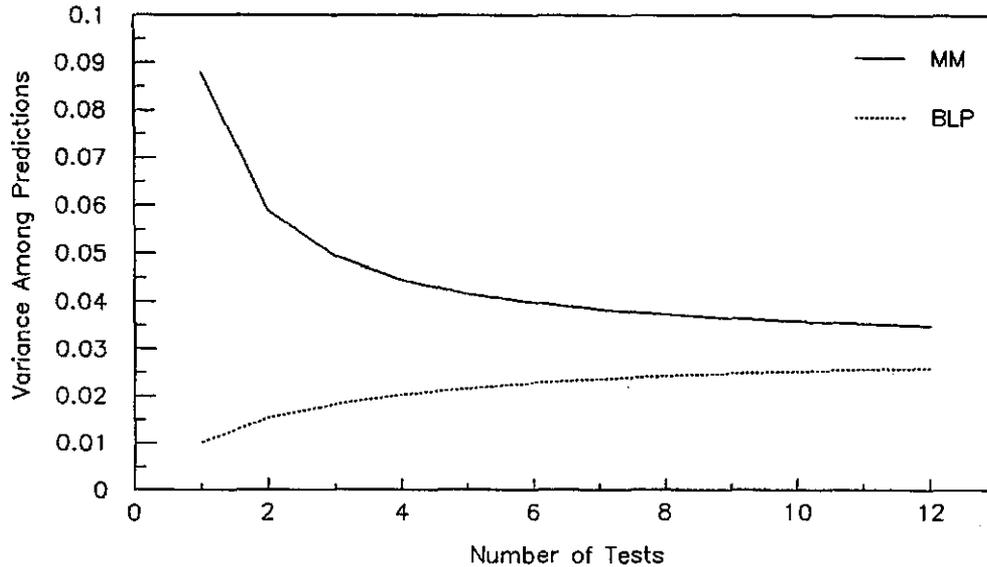
**Fig. 2.** Variance among breeding values predicted from Best Linear Prediction (BLP) and Marginal Means (MM)

| | | |
|---|---|---|
| E (**y**) | = | α : the expected values of the n x 1 vector of data values (the first moments). The elements of α represent fixed effects such as site, blocks and treatment effects. |
| Var (**y**) | = | **V** : an n x n variance-covariance matrix among the data values (the second moments). |
| Cov (**y**, **g**') | = | **C** : a n x q matrix of covariances between the data values and the breeding values being predicted. |

(White and Hodge 1989)

In BLP, the prediction of **g** is a linear function of the data which minimises the error variance of predictions, $E(\hat{g} - g_i)^2$ from all possible solutions:

$$\hat{g} = \tau + C'V^{-1} (y - \alpha).$$

(Quaas *et al.* 1984)

This equation is analogous to linear regression in two ways. Both linear regression and BLP only consider linear functions of the data. Additionally, $C'V^{-1}$ can be considered as Cov (**g**,**y**) 'divided by' var (**y**) which is similar to the regression of **g** on **y** (Thompson 1979). In BLP a different set of coefficients is developed for each trait and parent, which depends on the quality and quantity of the data. For families with high phenotypic variability in relation to additive genetic variance as found in imprecise tests, the predicted breeding values will be regressed towards zero (closer to mean) as the heritability will be low (White and Hodge 1989).

### 7.3    First and second moments

In the theory of BLP it is assumed that the first moments are known. As this does not occur in practice, they are estimated by OLS procedures (section 6). This may involve a) calculating the family mean for each block, site, year etc or b) expressing the family means as deviations from the site mean or a genetic standard. This adjustment attempts to correct for differences among sites and/or years. Once precise and unbiased estimates of the fixed effects are obtained, the genetic effects can be predicted.

The derivation of BLP also assumes that the variances and covariances are known. As this is never the situation in practice, they have to be estimated by direct or indirect methods. If the data set represents a large random sample of the population, then the variance of the family means can be calculated directly from the given data. It is also possible to calculate the variances of the family means from the variance components. The variance components can be obtained from the expectations of the mean squares (Method of Moments) or via a method such as Restricted Maximum Likelihood if the data are unbalanced (section 8.3). If reliable estimates of the second moments cannot be obtained from the given data because it is highly unbalanced or is insufficient, then data from related experiments may be used. These experiments may be repeated over several locations and/or years and hence may provide precise estimates of the variances of the family means.

## 8    METHODS OF ESTIMATING VARIANCE COMPONENTS

There are several methods for estimating variance components. All procedures yield the same estimators when the data are balanced but give different results when the design is unbalanced. Methods discussed here include:

- Method of moments
- Minimum norm quadratic unbiased estimators (MINQUE)
- Maximum likelihood (ML)
- Restricted maximum likelihood (REML)

### 8.1    Method of moments

The method of moments (or ANOVA based estimators) is the simplest method of calculating variance components and is recommended for designs that are balanced. Four steps should be followed:

1)    Calculate the mean squares for each random effect in the model
2)    Determine the expected value of each mean square
3)    Equate the observed value of each mean square to its expectations
4)    Solve the set of equations to obtain estimates of the variance components

The method of moments procedure can result in negative variance components. For most balanced designs the method of moments estimates have a minimum variance property and they are unbiased (Milliken and Johnson 1984).

## 8.2    MINQUE

MINQUE is similar to the method of moments in that functions of the sums of squares are equated to their expectations. However, in MINQUE, quadratic forms are considered. According to Milliken and Johnson (1984), MINQUE estimators are difficult to compute but efficient algorithms for the computations are now available. MINQUE estimators are unbiased and have a minimum variance property.

## 8.3    ML/REML

The first method of estimating variance components for unbalanced data was ML estimation (Henderson 1984). This was developed by Hartley and Rao in 1967 (cited in Henderson 1988). The ML technique uses the assumed distribution of the observations to construct a likelihood function which depends on the model parameters (Milliken and Johnson 1984). ML estimators are those values of the parameters that maximise the $\log_e$ of the likelihood function.

A modified ML method is called REML. REML estimators are obtained by iterating on MINQUE (Henderson 1984). REML is the preferred method of variance component estimation in animal breeding (Meyer 1991) and many other areas of research. In REML, ML estimators of the variance components are obtained for the part of the model which is free of fixed effects (Patterson and Thompson 1971). Elimination of the fixed effects removes the bias in ML estimates. Both ML and REML are iterative processes and may take many iterations to converge especially if the data are highly unbalanced.

# 9    BEST LINEAR UNBIASED PREDICTION

## 9.1    BLP versus BLUP

There are situations where unbiased and precise estimates of fixed effects cannot be obtained by OLS procedures. This may occur if the data are highly unbalanced or genetic standards have not been planted to allow an adjustment for variation among sites. In these instances BLUP is used to simultaneously estimate the fixed effects and predict the breeding values.

BLUP was developed by Henderson (1963, 1973 and 1977a,b) to allow analysis of dairy cattle data sets which typically are highly unbalanced. When evaluating sires based on the performance of their offspring, unbiased and precise fixed effects are difficult to estimate for the following reasons:

• The animals are so highly selected and related that these effects must be accounted for;

• Culling results in more data being available on higher performance animals;

• Large amounts of data are often missing.

(Henderson 1973,1974)

These complications coupled with the fact that dairy cattle data sets are typically large, led to the development of BLUP which involves best linear unbiased estimates (BLUE) of the fixed effects through GLS and BLUP of the random effects.

## 9.2    Linear model

When analysing data from an experiment, a linear model is frequently assumed to be a suitable approximation. It is also assumed that the observations (y) are a random sample from a population which has a multivariate normal distribution. Analyses based on this distribution have desirable statistical properties (Sorensen and Kennedy 1984). One of these features is that the multivariate normal distribution is defined completely by its mean and second moments (Henderson 1984).

A typical linear model is as follows:

$$y = X\beta + e \tag{1}$$

where **y** is a vector of observations; ß is a vector of fixed effects; **X** is the associated design/regression constants and **e** is a vector of errors.

Many experimenters have supplemented equation (1) with additional terms and assumptions. The following model is called a mixed model:

$$y = X\beta + Zu + e \tag{2}$$

where

y       is an n x 1 vector of observations. Var(y) = **V** which is an n x n matrix of the variances and covariances among the observations;

β       is a t x 1 vector of unknown fixed effects due to blocks, sites or any other 'nuisance' effect;

X       is an n x t design/regression matrix that relates the fixed effects in β to the elements in **y**. **X** contains values of 0 and 1;

u       is an s x 1 vector of random effects. In the case of BLUP, **u** is a vector of additive genetic effects or parental breeding values. E(**u**) = 0 while the Var(**u**) = **G**. **G** is an s x s non-singular matrix containing the variances and covariances among the random effects.

Z       is an n x s design matrix containing values of 0 and 1 that relate the random genetic effects in **u** to the elements in **y**; and

e       is an n x 1 vector of random effects associated with components of experimental error. E(**e**) = 0 and the Var(**e**) = **R**. **R** is an n x n non-singular matrix containing variances among the experimental error components.

(White and Hodge 1989)

## 9.3 Mixed model theory

Under a random mating system and assuming that the genetic and error components of equation (2) are uncorrelated ie $E(\mathbf{u},\mathbf{e}) = 0$ then:

$$\text{Var}(\mathbf{y}) = \mathbf{V}$$

$$= \text{Var}(\mathbf{X}\beta + \mathbf{Zu} + \mathbf{e})$$

$$= \text{Var}(\mathbf{Zu}) + \text{Var}(\mathbf{e})$$

$$= \mathbf{Z}\,\text{Var}(\mathbf{u})\mathbf{Z}' + \text{Var}(\mathbf{e})$$

$$= \mathbf{ZGZ}' + \mathbf{R}$$

and

$$E(\mathbf{y}) = \text{mean of } \mathbf{y} = \mathbf{X}\beta.$$

(White and Hodge 1989)

Usually we do not know the values of $\mathbf{G}$ and $\mathbf{R}$ but make certain assumptions about the form of these matrices. Frequently $\mathbf{R}$ is assumed to be the identity matrix ($\mathbf{I}$) or a matrix containing diagonal elements that are equal and off diagonal elements that are zero. That is, the elements of $\mathbf{e}$ are assumed to have equal variances and be uncorrelated (Henderson 1984). On many occasions $\mathbf{G}$ is also a diagonal matrix. Despite these simplifications there is a problem with the mixed model. $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$ is often large and non-diagonal. Consequently, $\mathbf{V}^{-1}$ is difficult or impossible to compute. However, it was shown by Henderson (1973) that

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}. \qquad (3)$$

While this may seem a complicated method of calculating $\mathbf{V}^{-1}$, it should be noted that $\mathbf{R}^{-1}$ and $\mathbf{G}^{-1}$ are often easy to compute. Hence, this method of computing $\mathbf{V}^{-1}$ may have important advantages. $\mathbf{V}^{-1}$ can also be obtained by writing equations, known as Henderson's Mixed Model Equations (MME). For a single trait and a single measurement being made on an individual the MME is:

$$\begin{vmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{vmatrix} \begin{vmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{vmatrix} = \begin{vmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{vmatrix}$$

(4)

(Henderson 1950 cited in Henderson 1984)

### 9.3.1 Generalised inverses

In BLUP, one of the objectives is to determine a linear unbiased estimator of ß or linear functions of ß say k'ß. There may be many different estimates of $\hat{\text{ß}}$ and the one chosen is the GLS estimate of k'ß denoted by k'$\hat{\text{ß}}$ where $\hat{\text{ß}}$ is a solution to:

$$\hat{\text{ß}} = (\mathbf{X'V^{-1}X})^- (\mathbf{X'V^{-1}y}).\tag{5}$$

(Henderson 1984)

Such an estimator is called the BLUE of ß since it results in the minimum sampling variance.

$(\mathbf{X'V^{-1}X})^-$ is the generalised inverse of $(\mathbf{X'V^{-1}X})$. If a matrix (or matrix product in this case) is singular then true inverses where

$$\mathbf{MM^{-1} = M^{-1}M = I}$$

do not exist. In such a matrix, the rows (or columns) are linearly dependent which leads to a determinant of zero and hence no inverse exists. However, singular matrices do have generalised inverses (or g inverses for short) and for a matrix, S, this is designated as S⁻. An example from Quaas *et al.* 1984 will clarify this.

Suppose we have a matrix, S:

$$\begin{vmatrix} 5 & 2 & 3 \\ 2 & 2 & 0 \\ 3 & 0 & 3 \end{vmatrix}.$$

Since the second and third columns add up to the first there is linear dependency among the columns. The S matrix has a rank of 2 instead of 3 and thus no inverse exists. It does however, have a g inverse, which is designated as S⁻. The easiest g inverse to compute is obtained by deleting rows and columns to get a full rank subset which is inverted and then filled in with zeros where the elements were deleted.

If we consider the subset matrix of S which has the third row and column deleted:

$$\begin{vmatrix} 5 & 2 \\ 2 & 2 \end{vmatrix}.$$

The inverse of this matrix is:

$$\begin{vmatrix} 2/6 & -2/6 \\ -2/6 & 5/6 \end{vmatrix}.$$

Thus, a g inverse of S is:

$$
\begin{vmatrix}
2/6 & -2/6 & 0 \\
-2/6 & 5/6 & 0 \\
0 & 0 & 0
\end{vmatrix} .
$$

For any singular matrix **S**, Penrose (1955) showed there is a unique matrix **K** which satisfies the following conditions:

1) SKS = S
2) KSK = K
3) (KS)' = KS
4) (SK)' = SK

In the current application a matrix will be termed a g inverse if **SKS = S**. The g inverses do not have the most desirable property of true inverses ie SK ≠ I. It can be shown that in the example by Quaas *et al.* 1984, multiplication of the matrices:

$$
\begin{vmatrix}
5 & 2 & 3 \\
2 & 2 & 0 \\
3 & 0 & 3
\end{vmatrix}
\begin{vmatrix}
2/6 & -2/6 & 0 \\
-2/6 & 5/6 & 0 \\
0 & 0 & 0
\end{vmatrix}
\begin{vmatrix}
5 & 2 & 3 \\
2 & 2 & 0 \\
3 & 0 & 3
\end{vmatrix} .
$$

gives

$$
\begin{vmatrix}
5 & 2 & 3 \\
2 & 2 & 0 \\
3 & 0 & 3
\end{vmatrix} .
$$

If different rows and columns of **S** were deleted then a different g inverse would be obtained. Unlike non-singular matrices which have unique inverses, singular matrices have a family of g inverses.

### 9.3.2 Properties of BLUP

As defined by Henderson (1974) the BLUP of **u**, denoted by û, is:

$$
\hat{u} = GZ'(y - X\hat{\beta}) \tag{6}
$$

and $\hat{\beta}$ is a solution to

$$\hat{\beta} = (X'V^{-1}X)^{-} (X'V^{-1}y) \ .$$

Equation (6) has two powerful properties:

1)      The error variance of predictions, $E(u_i - \hat{u}_i)^2$ is minimised.
2)      The genetic evaluations are unbiased. Hence, $E(\hat{u}) = 0$.

## 10    EXTENSION OF BLUP

In much of the theory in the proceeding sections, it was assumed that the individuals were unrelated, non-inbred and from unselected populations. It was also assumed that an additive genetic model was appropriate. If any of these assumptions are violated then this complicates the application of BLUP. In animal breeding where much of the theory was developed, the application of BLUP is complicated by several factors:

- Data sets are typically large and unbalanced. This creates difficulties for inversion of matrices.

- The animals are so interconnected that all records back to the base population should be considered;

- The animals are culled and this results in more data being available on higher performance animals.

(Henderson 1973,1974)

Much of the animal breeding literature discusses methods for handling these complications and methods of achieving computational efficiency. Some of the difficulties in applying BLUP are discussed in the subsequent sections.

### 10.1   Estimating second moments

In the theory of BLUP it is possible to completely specify the individual components of the of C and V matrices. When the data are from various locations or the individuals are planted at different times, BLUP can correctly weight and scale the data for the different error and/or genetic variances. However, when heterogeneous second moments are coupled with large data sets, computations become difficult. Thus, to allow calculations to be more feasible, a model commonly used in the dairy cattle industry is the Reduced Animal Model which is a modification of equation (4) with $G = A\sigma^2_a$ and $R = I\sigma^2_e$ (Quaas and Pollak 1980). A is a symmetric matrix of size s x s and is known as the numerator relationship matrix (NRM) which specifies the additive genetic relationships among the parents. The diagonal elements are equal to $1 + F_i$ where $F_i$ is the inbreeding coefficient of the $i$th parent and the off-diagonal elements are equal to the numerator of Wright's (1922) coefficient of relationship between the $i$th and $j$th parents (Kennedy and Sorensen 1988). I is the identity matrix while $\sigma^2_a$ and $\sigma^2_e$ are scalars representing the additive genetic and error variances and are assumed to apply to all observations in y. Even if A is large, its inverse can be written directly without ever

constructing **A** (Henderson 1976). It only requires a list of individuals and their pedigree information. According to Everett and Jones (1985) an $A^{-1}$ with an order of 10000 can be calculated in a matter of seconds. Henderson discovered this in the mid 70s and this subsequently lead to the development of BLUP for multi-trait evaluation (section 10.4).

Henderson's MME for the Reduced Animal Model is:

$$\begin{vmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z + A^{-1}\lambda} \end{vmatrix} \begin{vmatrix} \hat{\beta} \\ \hat{u} \end{vmatrix} = \begin{vmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{vmatrix}.$$

(7)
(Henderson 1963)

where

$$\lambda = \sigma^2_e / \sigma^2_a.$$

When the variances are assumed to be equal but in fact differ, then prediction errors will result (White and Hodge 1989). If the data are not scaled to take into account that the variability is due to experimental error and not genetic differences, then too many selections from sites (or herds) with higher error variances will result (White and Hodge 1989). In dairy cattle breeding too many selections frequently result from herds with higher error variances.

## 10.2 Effects of inbreeding and coancestry

BLUP can use genetic information from relatives to reduce the error variance of predicted breeding values. In an additive genetic model, the effects of inbreeding are taken into account in **V** (equation (3)). However, in many applications of BLUP the additive genetic relationships are specified in the NRM which can be readily inverted for large data sets. In this matrix, all observations are assumed to have the same genetic and error variances. When homogeneous error and genetic variances cannot be assumed, the usefulness of the NRM is lessened and in these cases the elements of **V** should be directly specified (White and Hodge 1989). However, the inverse of **V** may be difficult to derive for large data sets.

## 10.3 Non-additive model

In the previous sections an additive genetic model for BLUP evaluation was assumed. However, Henderson (1977b) also developed a method for BLUP prediction using a non-additive model which took into account the effects of dominance. Computationally, all that is required is expansion of equation (2) to accommodate the additional random factor. The mixed model incorporating dominance is:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Zd} + .... + \mathbf{e} \qquad \text{Equation (8)}$$

where **d** is a vector of dominance genetic effects and all other terms are as previously defined (Henderson 1988).

Recall that Var(**u**) = **G** which is frequently assumed to be equal to $A\sigma^2_a$ (section 10.1). Now var(**d**) = $D\sigma^2_d$ and **D** is referred to as the dominance relationship matrix while $\sigma^2_d$ is the dominance genetic variation. **D** can be calculated from the elements of **A**.

Henderson's MME for the genetic model including dominance is as follows:

$$
\begin{vmatrix}
X'R^{-1}X & X'R^{-1}Z & X'R^{-1}Z \\
Z'R^{-1}X & Z'R^{-1}Z + A^{-1}/\sigma^2_a & Z'R^{-1}Z \\
Z'R^{-1}X & Z'R^{-1}Z & Z'R^{-1}Z + D^{-1}/\sigma^2_d
\end{vmatrix}
\begin{vmatrix}
\hat{\beta} \\
\hat{u} \\
\hat{d}
\end{vmatrix}
=
\begin{vmatrix}
X'R^{-1}y \\
Z'R^{-1}y \\
Z'R^{-1}y
\end{vmatrix}.
$$

(9)

(Henderson 1988)

Although there is an easy method for inverting **A**, no easy solution for inverting **D** is available especially if the data set is large. Henderson (1985) showed that a smaller set of equations suffices to compute BLUP.

When inbreeding depression exists its effects are not accounted for by **D** but rather through incorporating the inbreeding coefficient as a covariate in the model (Kennedy and Sorensen 1988).

## 10.4   Multiple-trait BLUP

Depending on the size of the data set, computing skills and facilities available, it is possible to use BLUP on:

1)      Individual trait data to predict on an individual trait basis

2)      Individual trait data to predict on an aggregate genotype basis through the application of economic weights

3)      All data to predict on an individual trait basis

4)      All data to predict on an aggregate genotype basis

The different uses of BLUP are referred to as methods.

Method 1 is the simplest method and may be undertaken on simple or compound variables. In method 2, BLUP estimates for each parent are obtained on an individual trait basis. Depending on the economic importance of each trait, the BLUPs are weighted by coefficients to obtain a single BLUP for each parent. It is important to note that when economic weights are applied after predictions to obtain an aggregate genetic value for each parent, the predictions are still BLUE.

With methods 1 and 2, correlations among traits are not considered. In methods 3 and 4, consideration of the genetic and environmental covariances among the traits may reduce the error variance of the predicted breeding values.

The simultaneous estimation of breeding values for several traits is termed *multi-trait BLUP* and is outlined in Henderson and Quaas (1976). According to Arnason (1986) and Pollak (1988), multi-trait BLUP has several advantages and these include:

- If the traits involved are correlated, correctly specifying the genetic and environmental covariance structure among the traits will decrease the error variance of the predicted breeding values. The greatest increase in accuracy of selection is for those individuals with fewest records.

- It can evaluate traits on which the measurements may not have yet been recorded. This is possible through the correlated performance of the individuals's progeny.

- It can use all observations regardless if some data are missing by using correlations among traits.

- It can eliminate selection bias. If a single trait analysis is undertaken and this is not the trait on which selection was practised, then the estimates of breeding value will be biased. In a multi-trait approach inclusion of both traits will eliminate much of the selection bias (Quaas *et al.* 1984).

In multi-trait BLUP, many equations have to be solved. This results in substantial increase in computing resources in comparison to single trait procedures. The value of the gain in accuracy must therefore be weighed against the cost of increased computational effort (Arnason 1986). Computing expenses are likely to decrease as computers become more powerful and solve even larger sets of equations at lower costs. If computing resources or suitable programs are not available to undertake multi-trait BLUP, method 2 provides an appropriate alternative. Method 1 using calculated variables which consider two or more other variables is another option. Obviously, these latter two methods do not consider the correlation among traits.

## 10.5 Effects of selection

BLUP can account for previous cycles of selection providing the relationships between all parents and their relatives on which selection was based are included in the NRM (Pollak and Quaas 1981; Eriksson *et al.* 1981). If all relationships are not included or incorrectly specified, prediction errors will result. Another approach to removing selection bias is to use genetic groupings which represent the different sub-populations over time (Taylor 1981 cited in Parnell 1984). These groups would be treated as fixed effects in the BLUP model.

# 11 APPLICATIONS OF BLUP

## 11.1 Range of applications

BLUP was developed by Henderson to allow prediction of breeding values for dairy cows from large data sets which are usually particularly messy. Hence, it has been used extensively within the dairy industry and many of these applications were discussed in previous sections. BLUP has also been applied to other areas of animal breeding research, particularly beef cattle (Quaas and Pollak 1980; Graser and Hammond 1985), pigs (Belonsky and Kennedy 1988) and horses (Tavernier 1988).

Although the theory could easily be adapted to other breeding programs, there have been few applications of BLUP to plants. One of the most easily understood discussions on the theory of BLUP and potential applications in forestry breeding was by White and Hodge (1989). Although they discuss the theory and problems encountered in the application of BLUP, they have applied BLP to much of their data as they were able to obtain precise estimates of fixed effects.

In a study by Hill and Rosenberger (1985), several different methods of combining data from a series of alfalfa trials at a single location were evaluated. Methods of assessment were divided into three groups:

1)    *Percent of checks* - yields for each cultivar in each year were assessed relative to the genetic checks in a trial;

2)    *Summation of differences* - the yield for each cultivar was expressed as a deviation from the trial mean and check mean based on all trials. These deviations were summed over all trials for each cultivar;

3)    *BLUP* - several different forms of BLUP were tested and these varied in complexity depending on assumptions about the form of $R^{-1}$ and $G^{-1}$; for example the BLUP models ranged from $R^{-1}$ equal to the identity matrix and $G^{-1}$ not being part of equation 4 to a model where the components of $R^{-1}$ and $G^{-1}$ were estimated from a two-way (cultivars and trials) analysis of variance.

Of the 14 trials evaluated, 11 had yield data for each of three years. The various methods of assessment were conducted by omitting, one at a time, each of the trials and using the estimates to predict the data for the trial that was excluded (Hill and Rosenberger 1985). The optimal method was the one with the smallest prediction error.

Most of the BLUP methods were highly superior to other methods. The BLUP version in which the environmental and genetic variances were estimated for each trial was only slightly better than the summation of differences method. The method where data were assessed relative to genetic checks had the largest prediction errors.

## 11.2 Applications within the sugar industry

Within the sugar industry, the only known studies applying BLUP was by Chang and Milligan (1992a,b). They compared the effectiveness of BLUP to three other techniques as methods of cross appraisal. Chang and Milligan (1992a,b) did not attempt to estimate breeding value of parents, so their objective was different.

## 12 BSES METHOD OF ASSESSING BREEDING VALUE

### 12.1 Background to Australian sugar industry

Within Australia, approximately 39 million tonnes of sugarcane are produced annually with a value of $1.5 billion to the Australian economy. Approximately 95% of the production is in Queensland while the remainder is located in northern New South Wales.

The BSES is the principal research and extension organisation serving both the agricultural and milling sections of the Queensland Sugar Industry. One of the main objectives of BSES is to improve the productivity and profitability of the Australian Sugar Industry, and breeding is recognized as one of the most important mechanisms for achieving this. The importance of plant breeding to the Queensland sugar industry was assessed by Hogarth (1976). He estimated that from 1948 to 1975 plant breeding contributed about 50% to increasing yields by 1.9% per annum.

### 12.2 Factors which affect breeding value

The success of a breeding program relies heavily on effectively identifying superior parents to be used in the crossing system. The current BSES method of estimating breeding value is empirical and the following information is used when assessing the breeding potential of parents:

- Agronomic performance,
- Previous breeding performance, and
- Disease resistance status.

#### 12.2.1    Agronomic performance

Clones are initially planted in the parental collection if they exhibit superior phenotypic performance in yield trials which are conducted over a number of years and locations. Clones are assessed relative to commercial standards and the results are tabulated in a selection index called *net merit grade*. This is based on sugar yield, CCS, fibre quality and visual grade. After a number of years in the crossing program, the breeding potential of the clones can be evaluated and their continued usage in the parental collection will depend on that performance (Hogarth and Skinner 1987).

About 100 foreign commercial clones are imported annually and after two years in quarantine, they are evaluated in the parental collection. Despite their phenotypic value and lack of adaptation to Australian conditions, these foreign clones are assessed in experimental crosses as they may have valuable genes (Hogarth and Mullins 1989).

### 12.2.2 Breeding performance

Many experimental crosses are evaluated each year at five major Sugar Experiment Stations (Meringa, Herbert, Burdekin, Central and Bundaberg) and the progeny are evaluated. The aim is to identify those crosses with specific combining ability (SCA) coupled with high GCA. These crosses are called *proven* which means that more seedlings from that cross should be planted.

The system for identifying proven parents is based on either:

1)  *Selection rate* - the percentage of original seedlings selected and replanted in more advanced stages of testing (Hogarth and Skinner 1986). Selection in original seedlings is based on a sucrose measurement and visual assessment of yield. According to Skinner *et al.* (1977) selection rate combines for each family all the factors - mean, variance and type of distribution that are likely to be important in the identification of superior clones. However, it was discovered in the early 1970s, that greater progress could be made by family selection rather than selection of individuals (Hogarth 1971). At the time this could not be done due to inadequate weighing facilities.

2)  *Family selection* - whole families are rejected or selected based on mean performance (Falconer 1960). Family selection is particularly effective for traits with low heritability such as cane yield (Cox and Hogarth 1993). Until the development of mobile weighing equipment in the late 1980s, a visual assessment of yield was used to determine elite clones or families. However, it is now possible to obtain an efficient and objective evaluation of families (Bull *et al.* 1992).

### 12.2.2.1 Selection rate

For a single planting of a cross, the ratio of the selection rate of the cross relative to the selection rate for the whole population is termed the *cross ratio* (Hogarth and Skinner 1986). Up to ten years of selection data may be used when assessing whether a cross is proven or not. If a family is planted many times in the ten year period, a weighted cross ratio is calculated. Cross ratios are weighted according to the number of seedlings planted in the original population and more emphasis is given to the later stages of selection.

If the cross ratio is 0.9 or higher the cross is called *proven*. Proven cross status is allocated according to the actual cross ratio as follows:

## Table 1    Assessment of proven cross status

| Cross ratio | Proven cross status |
|---|---|
| < 0.9 | Failed |
| 0.9 - 1.5 | 1 |
| 1.6 - 2.5 | 2 |
| 2.6 - 3.5 | 4 |
| > 3.5 | 6 |

Thus, to determine the breeding potential of a parental clone, two factors are considered:

- The number of proven crosses made with that parent, and
- The number of non proven (failed) crosses with that parent.

### 12.2.3    Disease resistance ratings

Information on the reaction of parental clones to various diseases is also taken into account when assessing breeding value. Disease ratings are assigned to a clone on a 1-9 scale where 1 is resistant and 9 is susceptible.

In the empirical formula, disease ratings are in the form of an adjustment factor which is designed to equal zero if a clone has a rating of five for all diseases (Hogarth and Skinner 1987). The form of the adjustment depends on the region from which the parental clone comes and thus more important diseases in a particular area are given greater weighting.

The formula used to calculate a breeding value estimate of a clone on a particular station is:

$$Y = \text{Net merit grade} + \text{disease adjustment} + 1.5 \sqrt{X_1} + 0.3 \sqrt{X_2} - 0.3 \sqrt{X_3} \tag{10}$$

where

| | | |
|---|---|---|
| $Y$ | = | breeding value estimate |
| $X_1$ | = | sum of the proven cross status for the clone for that station |
| $X_2$ | = | sum of the proven cross status for the clone for all other stations |
| $X_3$ | = | number of failed (non proven) crosses that the clone has been tried in for that station. |

(Hogarth and Skinner 1986)

The coefficients in equation (10) were derived from a series of multiple regression analyses. After many regression analyses, the coefficients were optimised.

In recent years, experiments were undertaken to assess the importance of each component in the breeding estimate formula (BSES unpublished data). The results of these experiments

indicated that, when clones have been used in a large number of crosses, greater emphasis should be given to breeding performance rather than net merit grade. For new clones there is no information on breeding potential and the best available information is net merit grade.

## 12.3 Improvements to breeding program

It is recognized that more precise estimates of breeding value will result in a more efficient breeding program and thus increase the rate of population improvement. BSES plant breeders have not been satisfied with the method of calculating breeding value for some time but have found it difficult to make improvements. Directly related to the problem of estimating breeding value is the design of the breeding system. The current system has a very unbalanced mating design and as such is not very amenable to conventional statistical analyses. This unbalanced design is partly caused by unreliable flowering and poor pollen fertility at the main breeding station of Meringa in North Queensland (Berding and Skinner 1987).

With the aim of improving the breeding population through more accurate and precise estimates of breeding value, statistical techniques such as BLUP have been investigated.

## 13    ANALYSES ASSUMING NORTH CAROLINA DESIGN I

### 13.1    Quantitative genetic assumptions

The interpretation of genetic parameters in subsequent sections depends on the validity of several assumptions. These are:

1)      Diploid inheritance,
2)      Absence of linkage,
3)      Absence of epistasis,
4)      Random choice of parents from a random mating population,
5)      No maternal effects.

In sugarcane, all assumptions except the absence of maternal effects may be violated (Brown *et al.* 1968 and Hogarth 1968, 1971). Although the genetic assumptions may be invalid for sugarcane, it is more important to know if this results in large errors in the estimation of variance components. This was the subject of a study by Hogarth (1977). In particular, the study concentrated on the assumption of no epistasis. Failure to meet the assumptions was assessed by comparing independent estimates of genetic variance components based on different genetic assumptions. The study was designed to test only large violations as the variance components had high standard errors.

Hogarth (1977) collected data on brix, tonnes cane per hectare, weight per stalk and number of stalks. For all traits except tonnes cane per hectare, additive genetic variance was more important than dominance genetic variance. For cane yield the two forms of genetic variance were approximately equal.

Hogarth (1977) found that brix was characterised by a low coefficient of variation and so was the most appropriate trait on which to base conclusions. Variance components based on the

different assumptions gave similar results for brix which suggested that violation of assumptions had little effect on estimation (Hogarth 1977). However, for the other traits, the different methods of calculating variance components gave varying results which suggested that some assumptions were violated. Epistatic variance was important for weight per stalk but not for the other traits. For tonnes cane per hectare it was difficult to determine which assumptions were invalid. It was proposed by Hogarth (1977) that failure to meet the assumption of regular diploid inheritance may have affected estimation.

It is assumed in quantitative genetic experiments that parents are randomly chosen from a random mating population. This is unlikely to be fulfilled in sugarcane experiments as varieties are included in the parental collection on the basis of superior agronomic or breeding performance. The effect of invalidating the assumption will be to increase or decrease the genetic variance, so that a bias in the estimate of genetic parameters is inevitable. In the pursuing analyses many families are included and these can be considered representative of the breeding population. As inferences apply not only to the parents in the study but to the broader breeding population, the breeding values are considered random effects and hence techniques such as BLUP are appropriate.

## 13.2 Further assumptions

None of complications, such as the effects of selection and inbreeding which make BLUP more computationally difficult, are considered in the analyses detailed in subsequent sections. It is assumed that the parents involved are unrelated and hence the level of inbreeding equals zero. This may not be valid but as this is the first application of BLUP to sugarcane data sets in Australia, it is appropriate to commence at a preliminary level. The level of inbreeding in the BSES parental collection is quantified in section 15.

## 13.3 Trial details

As the weighing of families is a recent initiative, objective family selection data sets are somewhat limited. However, several analyses have been undertaken.

Family selection data from the 1989 southern Queensland series were available for this study. Two trials conducted on blocks K4E and K11 at the Bundaberg Sugar Experiment Station consisted of 50 and 81 families, respectively. Both trials had varying replication which ranged from one to several replicates of each family. Family plots contained 10 clones each in a single row, 2 metres in length.

Data measured at harvest of the plant crop included cane yield (t/ha) of each family plot, average CCS of the 10 clones in a family plot and tonnes sugar per hectare. Tonnes sugar per hectare is a derived variable and is calculated as follows:

(Cane yield * CCS) /100

## 13.4 The model

Data from K4E and K11 were analysed assuming a nested or North Carolina Design I (Comstock and Robinson 1952). The linear model is:

$$Y_{ijk} = \mu + m_j + (f/m)_{jk} + e_{ijk}$$

where

$Y_{ijk}$ is the phenotypic observation on the $i$th replicate from the $j$th male and $k$th female
$\mu$ is the population mean
$m_j$ is a random variable associated with the $j$th male $\sim$ NID $(0,\sigma^2_m)$
$(f/m)_{jk}$ is a random variable associated with the $k$th female within the $j$th male $\sim$ NID $(0,\sigma^2_{f/m})$
$e_{ijk}$ is the random plot error associated with the $i$th replicate from the $j$th male and $k$th female
$\quad \sim$ NID $(0,\sigma^2_e)$

$i = 1,...,r; j = 1,...,m$ and $k = 1,...,f$, where $r$ refers to the number of replicates, $m$ refers to the number of males and $f$ refers to the number of females.

For sugarcane the reverse model is also appropriate:

$$Y_{ijk} = \mu + f_j + (m/f)_{jk} + e_{ijk}$$

where

$Y_{ijk}$ is the phenotypic observation on the $i$th replicate from the $j$th male and $k$th female
$\mu$ is the population mean
$f_j$ is a random variable associated with the $j$th female $\sim$ NID $(0,\sigma^2_f)$
$(m/f)_{jk}$ is a random variable associated with the $k$th male within the $j$th female $\sim$ NID $(0,\sigma^2_{m/f})$
$e_{ijk}$ is the random plot error associated with the $i$th replicate from the $j$th female and $k$th male
$\quad \sim$ NID $(0,\sigma^2_e)$
$i = 1,...,r; j = 1,...,m$ and $k = 1,...,f$, where $r$ refers to the number of replicates, $m$ refers to the number of males and $f$ refers to the number of females.

## 13.5 Features of GLMM

Using a generalised linear mixed model program called *GLMM* (Blouin and Saxton 1990) BLUE of the fixed effects and BLUP of the random effects can be obtained. Using GLMM only one dependent variable can be specified. Hence, multi-trait BLUP cannot be considered. GLMM allows for a range of models with main, cross, nested, cross nested and nested crossed effects.

## 13.6 Results

GLMM offers the user a range of methods for the calculation of variance components. As REML is the preferred method in many areas of research, it was used in all calculations.

For both trials, narrow sense heritabilities ($H_n$) on a family mean basis were calculated from the REML variance components. As the families within each trial had unequal number of replicates, the error variance had to be divided by an approximate value to express the heritabilities on a family mean basis (Steel and Torrie 1980; page 151).

For each trial, the BLUP estimates of each male (or female in the second model) were correlated with the arithmetic means. Results of the analyses undertaken on the data from blocks K4E and K11 are given in tables 2 and 3, respectively.

**Table 2**    **Narrow sense heritabilities on a family mean basis and correlations between BLUP estimates and means for data from block K4E**

*(a) Model:    Males and Females/males*

| Variable | $H_n$ | Correlation coefficient |
|---|---|---|
| Cane yield | 0.90 | 0.99 |
| CCS | 0.91 | 0.99 |
| Tonnes sugar per hectare | 0.72 | 0.99 |
| Net merit grade | 0.39 | 0.96 |

*(b) Model:    Females and Males/females*

| Variable | $H_n$ | Correlation coefficient |
|---|---|---|
| Cane yield | 0.19 | 0.97 |
| CCS | 0.10 | 0.95 |
| Tonnes sugar per hectare | 0.28 | 0.97 |
| Net merit grade | 0.34 | 0.96 |

**Table 3**    **Narrow sense heritabilities on family mean basis and correlations between BLUP estimates and means for data from block K11**

*(a) Model:    Males and Females/males*

| Variable | $H_n$ | Correlation coefficient |
|---|---|---|
| Cane yield | 0.05 | 0.92 |
| CCS | 0.90 | 0.98 |
| Tonnes sugar per hectare | 0.43 | 0.97 |
| Net merit grade | 0.67 | 0.96 |

*(b) Model:    Females and Males/females*

| Variable | $H_n$ | Correlation coefficient |
|---|---|---|
| Cane yield | 0.07 | 0.97 |
| CCS | 0.23 | 0.94 |
| Tonnes sugar per hectare | 0.10 | 0.97 |
| Net merit grade | 0.09 | 0.95 |

## 13.7  Discussion

In tables 2 and 3 the narrow sense heritabilities on a family mean basis for cane yield were typically low except for those in table 2a. This suggests that non-additive genetic variance is important for this character. This agrees with the findings of Hogarth (1977, 1987). For CCS there were conflicting results for narrow sense heritabilities in the two tables. For this variable additive genetic variation is usually more important than non-additive variation (Hogarth 1977, 1987).

It is interesting to note the differences in narrow sense heritabilities for cane yield from table 2a and 2b. In table 2a where the narrow sense heritability was high there were some very high and very low breeding values with the range being -23 to 23 tonnes per hectare. In table 2b cane yield had a low narrow sense heritability and the BLUP estimates of breeding values were regressed closer to the mean. The range in this case was -4 to 10 tonnes per hectare.

Despite the range in narrow sense heritabilities, the correlations between the BLUP estimates and means were always high. This is not surprising as the families were not replicated over sites. It is when data are from multiple sources which are particularly messy, that BLUP would be expected to exhibit superiority.

### 13.8  Problems with North Carolina Design I analyses

As given by Kearsey (1965) there are two assumptions which must be met prior to using a North Carolina Design I:

1)      Each male should be mated to an equal number of females (vice versa in the reverse model).

2)      A different groups of females should be used with each male.

In the current data sets there were unequal numbers of females/male. Steel and Torrie (1980, page 151) describe a method for handling the analysis of such data. Additionally, a female was often used more than once with a male which means that this no longer constitutes a random sample. This will introduce a bias into the variance components.

Although the assumptions for the North Carolina I were violated to some degree, the analyses provided useful preliminary results. At the time, a lack of suitable software for the analysis of BLUP meant that it was not possible to undertake the desired analyses.

## 14      ANALYSES ASSUMING INCOMPLETE DIALLEL DESIGN

### 14.1  The model

A more appropriate mating design for the family selection data is an incomplete diallel (Griffing 1956) without reciprocals where:

$$Y_{ijk} = \mu + g_j + g_k + s_{jk} + e_{ijk}$$

where

$Y_{ijk}$ is the phenotypic observation for progeny in the $i$th replicate from the $jk$th cross
$\mu$ is the population mean
$g_j$ is the random variable associated with the GCA of the $j$th female ~ NID $(0,\sigma^2_{gca})$
$g_k$ is the random variable associated with the GCA of the $k$th male ~ NID $(0,\sigma^2_{gca})$
$s_{jk}$ is the random variable associated with the SCA of parents $j$ and $k$ ~ NID $(0,\sigma^2_{sca})$
$e_{ijk}$ is the random plot error associated with the $i$th replicate from the $jk$th cross, ~ NID $(0,\sigma^2_{e})$

$i = 1,..,r;\ j = 1,..,f$ and $k = 1,..,m$, where $r$ refers to the number of replicates, $f$ refers to the number of females and $m$ refers to the number of males.

### 14.2  Features of GAREML

An extensive investigation of available software identified packages suitable for the analysis of data with an incomplete diallel design. Many of the programs were specifically designed for animal breeding and hence were not really appropriate.

Of the programs examined, *GAREML*, written by Dr Dudley Huber formerly of the Forestry Department from University of Florida and now at Weyerhaueser in Arkansas, was the most applicable for sugarcane data. GAREML is useful for estimating variance components and predicting breeding values. It applies the method of Giesbrecht's algorithm to estimate REML variance components and the theory developed by Henderson to obtain BLUP estimates of the random effects (Huber 1993).

GAREML can be applied to half-sibs, full-sibs and disconnected sets of genetic entries. These mating designs may be planted in single or multiple locations with one or more entries per plot. BLUP estimates of the parental GCA and SCA are obtained in addition to the GLS solutions of fixed effects. Multi-trait analyses cannot be undertaken at present but should be available in the near future (Huber pers comm).

GAREML is computationally efficient and can use hard disk space as virtual memory. This latter feature is particularly important with large data sets as inversion of matrices in the BLUP calculations is memory intensive.

## 14.3  Preliminary analysis of family selection data

The acquisition of GAREML provided BSES plant breeders with one of the first opportunities to estimate breeding values by statistically analysing family selection data. A highly unbalanced mating design coupled with subjective data had precluded the use of any conventional statistical techniques. Hence, it was difficult to compare the effectiveness of BLUP to other techniques that are used by BSES. The current method of assessing family selection data is based largely on net merit grade (section 12.2.1). Families with a mean net merit grade above a certain value are advanced to the next stage of selection. Since 1988, clones from families that have a high sugar content in early selection trials are also used as parents (Cox and Hogarth 1993).

Several data sets have been analysed but few have contained families replicated over years and/or locations. Data analysed have been characterised by the unequal replication of families within and across sites. The analyses have provided plant breeders with a simple and rapid means of combining data from a wide range of sources to identify superior parents. One of the BSES breeders used BLUP estimates to supplement existing information to select families for the central Queensland breeding program (McRae pers comm).

## 14.4  Comparison of predictions using BLUP and current methods

### 14.4.1  BLUP predictions vs family mean predictions

Theoretically, BLUP is expected to result in more accurate and precise estimates of breeding value than the current BSES methods. However, it is important to compare cross predictions based on the two methods before routine application of BLUP. Core family selection data for the 1988-1992 series seedlings from the southern Queensland breeding program based at Bundaberg were available for this study. Each family plot consisted of 10 clones each in a single, two metre row.

Data measured at plant crop harvest included:

- Cane yield of the whole family plot
- Average CCS of the 10 clones in a family plot
- Net merit grade.

Families were planted in several blocks with varying replication on the Bundaberg Sugar Experiment Station. To adjust for block differences, cane yield and average CCS of the family plots were assessed relative to genetic standards.

It was planned to use all available family selection data for this study. However, the dimensions of GAREML were constrained to 400 parents, so the study was restricted to the 1990-1991 series data. This data set contained 234 parents which were represented in 345 crosses. Many parents were used repeatedly as a male or female parent so estimates of GCA should be reliable.

GAREML was used to calculate BLUP estimates of breeding value for each of the 234 parents. For each variable, the BLUP estimates of each parent were used to predict the performance of 79 families present in the 1992 series. Many of these families were unique to the 1992 series; only some were present in the 1990-1991 series data. Hence, it was not possible to compare actual performance in 1990-1991 to that in 1992.

Predicted performance from the BLUP estimates was correlated with relative cane yield, relative CCS and net merit grade in the 1992 series. The correlations were found to be:

0.41 for relative cane yield ($p < 0.01$)
0.64 for relative CCS ($p < 0.01$) and
0.28 for net merit grade ($p < 0.05$) .

The correlation coefficient was highest for relative CCS. For this variable, additive genetic variation is usually more important than non-additive variation (Hogarth 1987) and hence performance is more predictable. The results are encouraging especially since family performance was based on data from one site and year. The analyses should be repeated using more family information from a second site. This will be harvested later in the year. Depending on the magnitude of family x location interaction, this may result in more reliable estimates of family performance and hence the correlations may improve.

Data from a BSES research project in southern Queensland, denoted as BS45S, was used to compare the predictions based on family mean performance and BLUP estimates from one stage of selection to actual family mean performance in the next stage.

In BS45S, 36 stage 2 (second stage of selection) families were planted on several blocks with varying replication on the Bundaberg Sugar Experiment Station. Each family plot consisted of 12 clones each planted in a single, two metre row. The same 36 families were planted in stage 3 (third stage of selection) on one block on the Bundaberg Station. This trial was a randomised complete block design with three replicates. Family means were based on 4 clones each planted in a single, ten metre row.

The stage 2 families were assessed in plant crop only. As the families were planted on several blocks it was necessary to assess performance relative to genetic standards. Plant crop data available included:

- Relative cane yield of the whole family plot
- Average relative CCS of the 12 clones per family plot
- Net merit grade.

The stage 3 families were harvested in plant and first ratoon crops and data collected included:

- Cane yield of the whole family plot
- Average CCS of the 4 clones per family plot
- Net merit grade.

Family mean performances in stage 2 were correlated with family means in stage 3.

Using the stage 2 data, GAREML was used to calculate BLUP estimates of breeding value for each parent. For each variable, the BLUP estimates for each parent were used to predict the performance of the 36 families involved in the study. Predicted performance from the BLUP estimates of the stage 2 were correlated with actual performance in stage 3.

Using the stage 3 data, GAREML was used to calculate BLUP estimates of breeding value for each parent. For each variable, the BLUP estimates for each parent were used to predict the performance of the 36 families. Predicted performances from the BLUP estimates based on data from stages 2 and 3 were correlated.

The results for plant, first ratoon and (plant + first ratoon)/2 are given in table 4.

**Table 4**      **Correlations between family performance in stages 2 and 3 for BS45S based on family means and BLUP predictions**

*(a) Stage 2 plant crop vs Stage 3 plant crop*

| Variable | Correlations between family mean performances in stages 2 and 3. | Correlations between predicted family performance in stage 2 based on BLUPs and family means from stage 3. | Correlations between predicted family performance based on BLUPs from stages 2 and 3. |
|---|---|---|---|
| Cane yield | 0.69** | 0.73** | 0.66** |
| CCS | 0.62** | 0.50** | 0.39** |
| Tonnes sugar per hectare | 0.54** | 0.60** | 0.62** |
| Net merit grade | 0.57** | 0.52** | 0.53** |

### (b) Stage 2 plant crop vs Stage 3 first ratoon crop

| Variable | Correlations between family mean performances in stages 2 and 3. | Correlations between predicted family performance in stage 2 based on BLUPs and family means from stage 3. | Correlations between predicted family performance based on BLUPs from stages 2 and 3. |
|---|---|---|---|
| Cane yield | 0.43** | 0.43** | 0.17 |
| CCS | 0.54** | 0.44** | 0.50** |
| Tonnes sugar per hectare | 0.28 | 0.24 | 0.19 |
| Net merit grade | 0.31 | 0.25 | 0.24 |

### (c) Stage 2 plant crop vs Stage 3 (plant + first ratoon)/2

| Variable | Correlations between family mean performances in stages 2 and 3. | Correlations between predicted family performance in stage 2 based on BLUPs and family means from stage 3. | Correlations between predicted family performance based on BLUPs from stages 2 and 3. |
|---|---|---|---|
| Cane yield | 0.62** | 0.63** | 0.64** |
| CCS | 0.64** | 0.52** | 0.61** |
| Tonnes sugar per hectare | 0.45** | 0.46** | 0.45** |
| Net merit grade | 0.52** | 0.45** | 0.51** |

** - $p < 0.01$

The results in table 4 were not surprising given that the same families were assessed in stages 2 and 3. Regardless of the prediction method used, the correlations for each trait were similar. The only exceptions to this were CCS in table 4(a) and cane yield in table 4(b). Correlations in table 4(b) were lower than in table 4(a). This was as anticipated as the predictions in stage 2 were based on plant crop data and in stage 3 on first ratoon information.

The results in table 4(b) and (c) give an estimate of the repeatability of the methods to predict family performance. A prediction method with higher correlations between successive harvests of a crop should indicate the future performance of the family more reliably (Chang and Milligan 1992b). Although the correlations are confounded by the different stages of selection, they were similar irrespective of the prediction method used (except for cane yield in table 4b). This suggests that neither method is superior.

It was believed that a technique such as BLUP that considers the family variance in addition to the mean would be superior to the other techniques. However, most correlations based on BLUPs in stages 2 and 3 were very similar to the correlations for the other techniques. Hogarth (1971) found that the mean was of more relevance in determining the importance of a cross than within family variance.

### 14.4.2 Comparison of breeding values based on BSES empirical formula and BLUP

It is important to compare the breeding value estimates based on the BSES empirical formula (equation (10)) to the BLUP method. Family selection data from the southern Queensland breeding program based at Bundaberg were available for this study.

Using the BSES empirical formula for estimating breeding value and the most recent information on clonal breeding potential, net merit grade and disease status, breeding values for the 79 families described above were calculated. As in the previous study on southern Queensland family selection information, BLUP estimates based on all data from the 1990-1991 series were used to predict performance of the 79 families in the 1992 series.

As the empirical formula has net merit grade as the only agronomic performance variable contributing to the estimation of breeding value, it was appropriate to centre the study on this character.

The correlation between breeding values based on the empirical formula and BLUP estimates was 0.48 ($p < 0.01$). This is quite high when considering that the correlations between actual net merit grade and the two methods of estimating breeding value were:

  0.28 for the BLUP method ($p < 0.05$), and
  0.24 for the empirical formula ($p < 0.05$).

The low correlation between actual net merit grade and breeding value based on the empirical method was not surprising. When parental clones have been used in few experimental crosses, the empirical formula places little emphasis on breeding performance and relies heavily on than net merit grade. In these instances, a low correlation between net merit grade and breeding value based on the empirical method would be expected. It is when there is extensive clonal breeding information that the breeding values from the empirical formula would correlate highly with net merit grade for families.

The low correlation between actual net merit grade and predicted performance from the BLUP estimates may be explained by the fact that net merit grade is a calculated variable. It is based largely on cane yield and CCS. Although genetic variation for CCS is largely additive, for cane yield non-additive genetic variation is more important (Hogarth 1987). Hence, net merit grade performance is difficult to predict.

## 15    ESTIMATION OF INBREEDING COEFFICIENTS

In the calculation of BLUP estimates, genetic information from related crosses can be used to reduce the error variance of predicted breeding values. The relationships can be specified in the NRM. Although with GAREML it is not possible to specify the NRM, an investigation of inbreeding was considered worthwhile as an insufficient amount of work has been done in sugarcane to quantify its effect on family performance. BSES plant breeders assume that it is detrimental to cross performance and so it is generally avoided.

Core family selection data from the southern Queensland breeding program were used in this study. Of the 684 families from the 1988-1992 series, 368 had BSES varieties as both parents. The study was restricted to these as BSES has more extensive pedigree information on their varieties than those from CSR or overseas countries. For many clones the pedigree could be traced to include varieties used in the initial interspecific hybridisation studies undertaken by Dutch breeders in Java.

Inbreeding coefficients were calculated using Proc Inbreed within SAS (SAS Institute Inc 1992). The frequency distribution of the inbreeding coefficients is given in Figure 3.

As the families in this study were planted on several blocks on the Bundaberg Station, it was necessary to assess performance relative to genetic standards. Data available included:

- Relative cane yield
- Relative CCS
- Net merit grade.
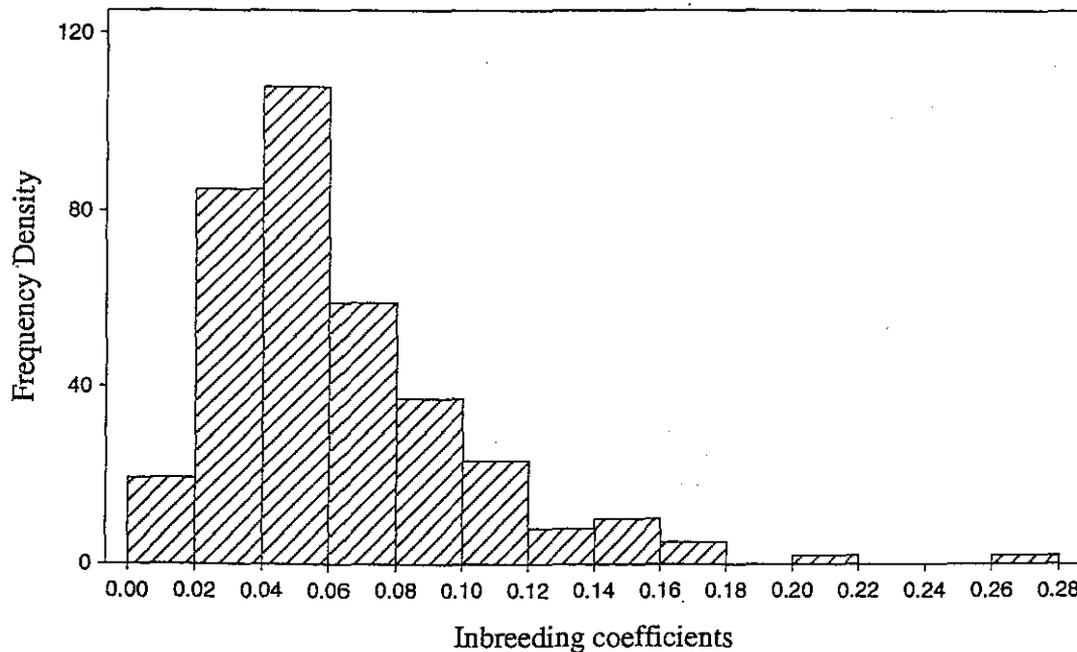


**Fig. 3. Frequency distribution of inbreeding coefficients for BSES families from the 1988-1992 southern Queensland series**

To determine the effect of inbreeding on performance, the inbreeding coefficients were correlated with these variables. The following results were obtained:

-0.02 for relative cane yield
-0.03 for relative CCS, and
-0.03 for net merit grade.

All inbreeding coefficients were low which is not surprising since BSES avoids inbreeding. Only three families had inbreeding coefficients higher than 0.2. Of these, two had above average performance for relative cane yield, relative CCS and net merit grade. The remaining family was above average for relative CCS but below average for the remaining variables.

There was no significant relationship between the inbreeding coefficients for the families and any of the variables analysed (p > 0.05). This does not agree with findings of Hogarth (1980). He found that cane yield exhibited substantial inbreeding depression. In the study by Hogarth (1980) there were controlled levels of inbreeding which were obtained by careful selection of clones for the 5 x 5 diallel cross. Inbreeding depression was considered important if the selfs had much lower means than the crosses. For cane yield there was severe inbreeding depression but CCS showed no such depression.

## 16    DIFFICULTIES

One of the main difficulties in this project was finding a suitable package for the BLUP analyses. The majority of programs available are specifically designed for animal breeding and are not really appropriate. GAREML was identified as the most suitable for analysis of sugarcane data, but its development is still in the early stages. GAREML cannot incorporate information on relatives.

## 17    CONCLUSIONS/FUTURE RESEARCH

Based on the limited amount of family selection data analysed in this project, BLUP appears to be as effective as the current BSES empirical formula for identifying superior clones to be used in the parental collection. Given that the empirical method incorporates information from up to ten years whereas the BLUP results were based on few years data, it is probable that BLUPs will increase the rate of population improvement which would have major benefits for the plant breeding program and the sugar industry.

In the current project it was difficult to compare breeding value estimates based on the BSES empirical formula and the BLUP method. An effective way of comparing BLUP to the empirical method, is to contrast the performance of crosses selected from parents based on the two techniques. SRDC funded project BS119S, 'Best linear unbiased prediction as a method for predicting cross potential' aims to research this. In the first phase of this project, it is planned to use the results from the BLUP analyses undertaken on all available family selection data from the Burdekin, Bundaberg and Mackay areas in conjunction with the current empirical method to nominate 100 crosses to be made in the 1994 and 1995 crossing seasons. The breakdown will be as follows:

40 crosses based on BLUP results
40 crosses based on the empirical method
20 random crosses

The crosses will be planted at Ayr, Mackay and Bundaberg in 1996 and assessed in plant crop in 1997. Yield and CCS will be measured at harvest. Performance of the BLUP crosses will be compared to those based on the empirical method. Both will be compared to the performance of the random crosses.

In this project none of the complications, such as multiple traits, effects of inbreeding and non-additive genetic variance, which make BLUP more computationally difficult were considered. Attempts were made to quantify the effect of inbreeding on cross performance and breeding value. However, low levels of inbreeding were found in most of the families examined. Within BSES, inbreeding is assumed to be detrimental to cross performance and so is generally avoided. The only effective way to study the effect of inbreeding, is to conduct an experiment where the levels of inbreeding are controlled. Such a study is difficult because of male sterility and incompatibility of some clones when crossed (Hogarth 1980). However, an experiment investigating inbreeding is planned as part of project BS119S.

## 18    INTELLECTUAL PROPERTY ARISING FROM THE RESEARCH

There are no matters of any description arising from the research in project BS75S, that pertain to intellectual property rights.

## 19    PUBLICATIONS

Stringer, J. K. (1993). Best Linear Unbiased Prediction as a Method of Estimating Breeding Value of Sugarcane Parental Clones. MAgST Project Report. The University of Queensland, Brisbane, Australia.

Stringer, J. K., McRae, T. A. and Cox, M., C. (1996). Best Linear Unbiased Prediction as a Method of Estimating Breeding Value in Sugarcane. In 'Sugarcane: Research Towards Efficient and Sustainable Production' (Eds J. R. Wilson, D. M. Hogarth, J. A. Campbell and A. L. Garside. (CSIRO Division of Tropical Crops and Pastures: Brisbane, Australia). In Press.

## 20    REFERENCES

Arnason, T. (1986). Estimation of breeding values for multi-trait selection. *World Review of Animal Production* **22**, 75-84.

Belonsky, G. M., and Kennedy, B. W. (1988). Selection on individual phenotype and best linear unbiased predictor of breeding value in a closed swine herd. *Journal of Animal Science* **66**, 1124-31.

Berding, N., and Skinner, J. C. (1987). Traditional breeding methods. In 'Copersucar International Sugarcane Breeding Workshop'. pp. 269-320. (Copersucar: Piracicaba, Brazil.)

Blouin, D. C., and Saxton, A. M. (1990). 'General Linear Mixed Models - GLMM Version 1.0.' (Department of Experimental Statistics, Louisiana State University: Baton Rouge, Louisiana.)

Brown, A. H., Daniels, J., and Latter, B. D. H. (1968). Quantitative genetics of sugarcane I - Analysis of variation in a commercial hybrid sugarcane population. *Theoretical and Applied Genetics* **38**, 361-9.

Bull, J. K., Hogarth, D. M., and Basford, K. E. (1992). Impact of genotype x environment interaction on response to selection in sugarcane. *Australian Journal of Experimental Agriculture* **32**, 731-7.

Chang, Y. S., and Milligan, S. B. (1992a). Estimating the potential of sugarcane families to produce elite genotypes using univariate cross prediction methods. *Theoretical and Applied Genetics* **84**, 662-71.

Chang, Y. S., and Milligan, S. B. (1992b). Estimating the potential of sugarcane families to produce elite genotypes using bivariate prediction methods. *Theoretical and Applied Genetics* **84**, 633-9.

Cockerham, C. C. (1963). Estimation of genetic variances in statistical genetics and plant breeding. In 'Statistical Genetics and Plant Breeding'. (Eds W. D. Hanson and H. F. Robinson.) pp. 53-93. (NAS-NRC Publication No. 982: Washington, DC.)

Comstock, R. E., and Robinson, H. F. (1952). Estimation of average dominance of genes. In 'Heterosis'. (Ed J. W. Gowen.) pp. 494-516. (Iowa State University Press: Ames, Iowa.)

Cotterill, P. P., Correll, R. L., and Boardman, R. (1983). Methods of estimating the average performance of families across incomplete open-pollinated progeny tests. *Silvae Genetica* **32**, 28-32.

Cox, M. C., and Hogarth, D. M. (1993). Progress and changes in the south Queensland variety selection program. *Proceedings of the Australian Society of Sugar Cane Technologists* **15**, 251-5.

Eriksson, J. A., Schaeffer, L. R., and Wilton, J. W. (1981). 'Estimating Sire's Breeding Value for Weaning Weight and Gain On Test By Different BLUP Procedures, Using Selected Records - A Simulation Study, Rep No 5.' (Department of Animal Breeding and Genetics: Uppsala, Sweden.)

Everett, R. W., and Jones, L. P. (1985). Breeding values - their calculation and potential use. In 'The Challenge: Efficient Dairy Production, Proceedings of the Conference Organised by the Australian and New Zealand Societies of Animal Production, March 25-28 1985, Albury-Wodonga, Australia'. pp. 406-19. (Australian Society of Animal Production: Warragul, Victoria.)

Falconer, D. S. (1960). 'Introduction to Quantitative Genetics.' (Oliver and Boyd: Edinburgh, London.)

Freund, R. J., and Littell, R. C. (1981). 'SAS for Linear Models.' (SAS Institute: Cary, North Carolina.)

Graser, H. U., and Hammond, K. (1985). Mixed model procedures for the Australian beef industry I. Multiple-trait model for estimation of breeding values for 200-day and final weights of cattle. *Australian Journal of Agricultural Research* **36**, 527-35.

Griffing, B. (1956). Concept of general and specific combining ability in relation to diallel crossing systems. *Australian Journal of Biological Sciences* **9**, 463-93.

Harville, D. A. (1984). Discussion of papers by Wahaba, Journel and Henderson. In 'Statistics An Appraisal'. (Eds H. A. David and H. T. David.) . (Iowa State University Press: Ames, Iowa.)

Hays, W. L. (1981). 'Statistics for the Social Sciences.' 3rd Edn. (Holt, Rinehart and Winston Inc: New York.)

Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics* **28**, 476-90.

Henderson, C. R. (1963). Selection index and expected genetic advance. In 'Statistical Genetics and Plant Breeding'. (Eds W. D. Hanson and H. F. Robinson.) pp. 141-63. (NAS-NRC Publication No. 982: Washington, DC.)

Henderson, C. R. (1973). Sire evaluation and genetic trends. In 'Animal Breeding and Genetics Symposium in Honour of J Lush'. pp. 10-41. (Animal Science Association of America: Champaign, Illinios.)

Henderson, C. R. (1974). General flexibility of linear model techniques for sire evaluation. *Journal of Dairy Science* **57**, 963-72.

Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**, 69-83.

Henderson, C. R. (1977a). Best least unbiased prediction of breeding values not in the model for records. *Journal of Dairy Science* **60**, 783-7.

Henderson, C. R. (1977b). Prediction of future records. In 'Proceedings of the International Conference On Quantitative Genetics'. (Eds E. Pollak, O. Kempthorne and T. B. Bailey.) pp. 615-38. (Iowa State University Press: Ames, Iowa.)

Henderson, C. R. (1984). 'Applications of Linear Models in Animal Breeding.' (University of Guelph: Ontario, Canada.)

Henderson, C. R. (1985). Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *Journal of Animal Science* **60**, 111-7.

Henderson, C. R. (1988). Progress in statistical methods applied to quantitative genetics since 1976. In 'Proceedings of the Second International Conference on Quantitative Genetics'. (Eds B. R. Weir, E. J. Eisen, M. M. Goodman and G. Namkoong.) pp. 85-90. (Sinauer Associates Inc: Sunderland, Massachusetts.)

Henderson, C. R., and Quaas, R. L. (1976). Multiple trait evaluation using relatives' records. *Journal of Animal Science* **43**, 1188-9.

Hill, R. R., and Rosenberger, J. L. (1985). Methods for combining data from germplasm evaluation trials. *Crop Science* **25**, 467-70.

Hogarth, D. M. (1968). A review of quantitative genetics in plant breeding with particular reference to sugarcane. *Australian Journal of Agricultural Research* **34**, 108-20.

Hogarth, D. M. (1971). Quantative inheritance studies in sugarcane II - Correlations and predicted responses to selection. *Australian Journal of Experimental Agriculture* **22**, 103-9.

Hogarth, D. M. (1976). New varieties lift sugar production. *Producers Review* **66**, 21-2.

Hogarth, D. M. (1977). Quantative inheritance studies in sugarcane III - The effect of competition and violation of genetic assumptions on estimation of genetic variance components. *Australian Journal of Agricultural Research* **28**, 257-68.

Hogarth, D. M. (1980). The effects of accidental selfing on the analysis of a diallel cross with sugarcane. *Euphytica* **29**, 737-46.

Hogarth, D. M. (1987). Genetics of sugarcane. In 'Sugarcane Improvement Through Breeding'. (Ed D. J. Heinz.) pp. 255-71. (Elsevier Science Publishers: Amsterdam.)

Hogarth, D. M., and Bull, J. K. (1990). The implications of genotype x environment interactions for evaluation of sugarcane families I - Effects on selection. In 'Genotype x Environment Interaction and Plant Breeding'. (Ed M. S. Kang.) pp. 335-44. (Louisiana State University: Baton Rouge, Louisiana.)

Hogarth, D. M., and Mullins, R. T. (1989). Changes in the BSES plant improvement program. *Proceedings of the International Society of Sugar Cane Technologists* **20**, 956-62.

Hogarth, D. M., and Skinner, J. C. (1986). Computerisation of cane breeding records. *Proceedings of the International Society of Sugar Cane Technologists* **19**, 478-91.

Hogarth, D. M., and Skinner, J. C. (1987). Computerisation of parental selection. In 'Copersucar International Sugarcane Breeding Workshop'. pp. 87-101. (Copersucar: Piracicaba, Brazil.)

Huber, D. (1993). Optimal Mating Designs and Optimal Techniques for Analysis of Quantative Traits in Forest Genetics. Ph.D. Dissertation, University of Florida, Gainesville, Florida.

Kearsey, M. J. (1965). Biometrical analysis of a random mating population: A comparison of five experimental designs. *Heredity* **20**, 205-35.

Kennedy, B. W., and Sorensen, D. A. (1988). Properties of mixed models methods for prediction of genetic merit. In 'Proceedings of the Second International Conference on Quantitative Genetics'. (Eds B. S. Weir, E. J. Eisen, M. M. Goodman and G. Namkoong.) pp. 91-103. (Sinauer Associates Inc: Sunderland, Massachusetts.)

Meyer, K. (1991). Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genetical Selection and Evolution* **23**, 67-83.

Milliken, G. A., and Johnson, D. E. (1984). 'Analysis of Messy Data Vol 1 Designed Experiments.' (Lifetime Learning Publications: London.)

Parnell, P. F. (1984). Best Linear Unbiased Prediction (BLUP) and the design of breeding programs. *Proceedings of the Australian Society of Animal Production* **15**, 513-6.

Patterson, H. D., and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545-54.

Penrose, R. A. (1955). A generalised inverse for matrices. *Proceedings of the Cambridge Philosophical Society* **51**, 406-13.

Pollak, E. J. (1988). New approaches to genetic evaluation of beef cattle. *Animal Breeding Opportunities, BSAP Occasional Report* **12**, 99-110.

Pollak, E. J., and Quaas, R. L. (1981). Monte Carlo study of genetic evaluations using sequentially selected records. *Journal of Animal Science* **52**, 257-64.

Quaas, R. L., and Pollak, E. J. (1980). Mixed model methodology for farm and rank beef cattle testing programs. *Journal of Animal Science* **51**, 1277-87.

Quaas, R. L., Anderson, R. D., and Gilmore, A. R. (1984). 'BLUP School Handbook.' (Animal Genetics and Breeding Unit, University of New England: Armidale, New South Wales, Australia.)

SAS Institute Inc (1992). 'SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements, Release 6.07.' (SAS Institute: Cary, North Carolina.)

Schultz, E. F. (1955). Rules of thumb for determining expectations of mean squares in analysis of variance. *Biometrics* **11**, 123-35.

Skinner, J. C., Hogarth, D. M., and Wu, K. K. (1977). Selection methods, criteria and indices. In 'Sugarcane Improvement Through Breeding'. (Ed D. J. Heinz.) pp. 409-53. (Elsevier Science Publishers: Amsterdam.)

Snedecor, G. W., and Cochran, W. G. (1967). 'Statistical Methods.' 6th Edn. (Iowa State University Press: Ames, Iowa.)

Sorensen, D. A., and Kennedy, B. W. (1984). Estimation of response to selection using least-squares and mixed model methodology. *Journal of Animal Science* **58**, 1097-106.

Sreenivasan, T. V., Ahloowalia, B. S., and Heinz, D. J. (1987). Cytogenetics. In 'Sugarcane Improvement Through Breeding'. (Ed D. J. Heinz.) pp. 211-53. (Elsevier Science Publishers: Amsterdam.)

Steel, R. G. D., and Torrie, J. H. (1980). 'Principles and Procedures of Statistics - A Biometrical Approach.' 2nd Edn. (McGraw-Hill Kogakusha Ltd: Tokyo.)

Stroup, W. W. (1989). Why Mixed Models ? In 'Applications of Mixed Models in Agriculture and Related Disciplines, Southern Co-operative Series Bulletin No 343'. pp. 1-8. (Louisiana Agricultural Experiment Station: Baton Rouge, Louisiana.)

Tavernier, A. (1988). Advantages of BLUP animal model for breeding value estimation in horses. *Livestock Production Science* **20**, 149-60.

Thompson, R. (1979). Sire evaluation. *Biometrics* **35**, 339-53.

White, T. L., and Hodge, G. R. (1988). Best linear prediction of breeding values in a forest tree improvement program. *Theoretical and Applied Genetics* **76**, 719-27.

White, T. L., and Hodge, G. R. (1989). 'Predicting Breeding Values with Applications in Forest Tree Improvement.' (Kluwer Academic Publishers: Dodrecht.)

White, T. L., Hodge, G. R., and Delorenzo, M. A. (1986). Best linear prediction of breeding values in forest tree improvement. In 'Proceeding of Statistical Considerations in Genetic Testing of Forest Trees, Southern Co-operative Series Bulletin 324'. pp. 99-122. (University of Florida: Gainesville, Florida.)

Wright, S. (1922). Coefficients of inbreeding and relationship. *American Nature* **56**, 330-8.

## 20 ACKNOWLEDGEMENTS