Theor Appl Genet (2007) 114:1319–1332
DOI 10.1007/s00122-007-0515-3

ORIGINAL PAPER

Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials

Helena Oakey · Arūnas P. Verbyla · Brian R. Cullis · Xianming Wei · Wayne S. Pitchford

Received: 7 September 2006 / Accepted: 25 January 2007 / Published online: 11 April 2007

Abstract A statistical approach for the analysis of multi-environment trials (METs) is presented, in which selection of best performing lines, best parents, and best combination of parents can be determined. The genetic effect of a line is partitioned into additive, dominance and residual non-additive effects. The dominance effects are estimated through the incorporation of the dominance relationship matrix, which is presented under varying levels of inbreeding. A computationally efficient way of fitting dominance effects is presented which partitions dominance effects into between family dominance and within family dominance line effects. The overall approach is applicable to inbred lines, hybrid lines and other general population structures where pedigree information is available.

Communicated by M. Cooper.

H. Oakey (✉) · A. P. Verbyla
School of Agriculture, Food and Wine, The University of Adelaide, PMB 1, Glen Osmond, SA 5064, Australia
e-mail: Helena.Oakey@adelaide.edu.au

A. P. Verbyla
e-mail: ari.verbyla@csiro.au

B. R. Cullis
Biometrics, NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, PMB, Wagga Wagga, NSW 2650, Australia
e-mail: brian.cullis@dpi.nsw.gov.au

X. Wei
BSES Ltd, Mackay Central Sugar Experiment Station, Peak Downs Highway, Te Kowai, PMB 57, Mackay Mail Centre, QLD 4741, Australia
e-mail: xwei@bses.org.au

W. S. Pitchford
School of Agriculture, Food and Wine, The University of Adelaide, Roseworthy, SA 5371, Australia
e-mail: wayne.pitchford@adelaide.edu.au

Introduction

Multi-environment trials (METs) are generally used at the later stages of crop breeding programs. They are important in assessing the suitability of lines in different environmental conditions, such as different locations, years or seasons for traits of interest. METs have two main aims, the selection of lines with superior performance and the selection of lines as potential parents for future crosses. Lines may be selected for specific environments or across a range of environments.

The selection of best performing lines is undertaken through well-designed breeding trials conducted across multiple environments and analyzed appropriately.

Mixed model approaches for analyzing METs are being used more widely (see Smith et al. 2005 for a recent review). Most of these approaches are based on classical quantitative genetics which partitions the phenotypic response into genetic line effects, environment effects, genetic line by environment interaction effects and within environment error effects. They differ in the treatment of the genetic line and environment effects as random or fixed and in the extent to which they define and explore the genetic line by environment interaction.

The suitability of lines as parents and the determination of preferable parental crosses has traditionally been carried out through specialized mating designs such as the diallel cross. These designs allow the partitioning of the genetic line effect into additive and non-additive line effects also

known as ''general combining ability'' and ''specific combining ability'', respectively (Griffing 1956). The additive effects or breeding values obtained for each line, measure the potential of a line as a parent (Falconer and Mackay 1996). The non-additive effects obtained for each line are associated with dominance and epistatic effects. Dominance genetic effects result from the interaction of alleles at a particular locus, whereas epistatic genetic effects result from the interactions between alleles at different loci. Specialized mating designs are however, often carried out in addition to MET analyses, and near or after the commercial release of lines therefore restricting their usefulness. These disadvantages often result in the suitability of lines as parents being assessed in the same way as the selection for best performing lines, that is, by examining their overall or total genetic effect.

The additive genetic effect is widely used in animal breeding programs to assess the potential of an animal as a parent, since it is not simple, nor practicable to replicate genotypes. The approach involves the incorporation of the pedigree information of animals into the analysis in the form of the additive relationship matrix $\boldsymbol{A}$ (Henderson 1976). When fitting non-additive effects in mixed linear models used to evaluate large pedigrees in animal breeding applications, a common simplifying is to ignore inbreeding and thus non-additive effects take the form of dominance and epistatic effects. Cockerham (1954) made theoretical developments for non-additive effects including dominance and epistatic effects under no-inbreeding. Henderson (1984, Chap. 29) shows how these results are applicable in practice by fitting a model which includes additive and non-additive effects.

In plant breeding trials, attempts at incorporating pedigree information into plant breeding trials have initially focused on special types of populations. Stuber and Cockerham (1966) give explicit theoretical results of genetic variances and covariances for hybrid relatives. Specifically, they consider the hybrid individuals produced from a cross between two separate parent populations. In Stuber and Cockerham (1966), the additive genetic effect of the hybrid individual is partitioned into two components, with each component relating to the additive genetic effect resulting from one of the parent populations. In addition, a dominance genetic effect of hybrid individuals is determined. Stuber and Cockerham (1966) however note that as a result of the partitioning of the additive genetic effect, more of the total genetic variance is assigned to the additive component and less to the dominance component. Bernardo (1994, 1996) applied these results to hybrid populations of maize. Lo et al. (1995) present theoretical developments for obtaining genetic means and covariances of a population composed of two pure breeds and their hybrid offspring, including dominance inheritance. Cockerham (1983) derives the covariance of relatives for individuals that are completely inbred, noting five relevant terms that make up the total genetic variances. These terms are additive variance, dominance variance, homozygous dominance variance, the covariance between additive and homozygous dominance effects and the inbreeding depression. Edwards and Lamkey (2002) apply this theoretical development to a maize population estimating all five terms.

Despite Cullis et al. (1989) acknowledging that pedigree information in the form of the additive relationship matrix can be incorporated into mixed model MET analysis readily, only recently have examples of this application to plant breeding programs surfaced. The use of the additive relationship matrix allows more general population structures to be considered. Panter and Allen (1995), Durel et al. (1998), Dutkowski et al. (2002), Davik and Honne (2005) and Crossa et al. (2006) all estimate additive effects using the additive relationship matrix. These papers however, fail to account for non-additive effects. Many authors (van der Werf and de Boer 1989; Hoeschele and VanRaden 1991; Lu et al. 1999) have indicated that accounting for non-additive effects in the genetic line effects might improve the estimation of additive effects resulting in less biased prediction. Costa e Silva et al. (2004) make some attempt at including dominance effects by including a between family effect (as would be applied in a diallel setting). The failure to account for non-additive effects in plant breeding trial settings appears to be mainly due to a lack of relevant theoretical developments for general population structures with varying levels of inbreeding. The theoretical developments that have been made are either for application in animal breeding programs, where when fitting non-additive effects in mixed linear models used to evaluate large pedigrees, a common simplifying is to ignore inbreeding, or for specialized populations, as discussed. Harris (1964) and later de Boer and Hoeschele (1993) do present the generalized genetic covariances between individuals allowing for varying levels of inbreeding. They give results for the variance of individuals explicitly in terms of the coefficients of parentage and inbreeding coefficients in these papers. However, until recently, theoretical developments of explicit results for the covariances between individuals under varying levels of inbreeding were lacking. Verbyla and Oakey (2006) have derived these explicit results under inbreeding. A copy of this research report is available from the corresponding author.

Recently, Oakey et al. (2006) presented a mixed model approach for single site analyses in which *both* the selection of potential parents for future breeding programs and promising commercial lines could be determined from the analysis of a standard crop breeding trial. The method involved partitioning the genetic effect of a line

into additive and non-additive effects. The additive line effects were so called breeding values and line performance was determined by combining both additive and non-additive effects in an overall or total genetic effect. This method was shown to be superior to the *Standard* analysis (which did not partition the genetic effect), with the total genetic effect having a lower prediction error variance on average than that obtained by the *Standard* analysis (Oakey et al. 2006).

In this paper, the partitioning of the genetic line effect used in Oakey et al. (2006) is extended to the analysis of METs, so that both the total genetic value of lines and the breeding values of potential parents can be determined. The approach is an extension to the MET model of Smith et al. (2001). With the view to analyzing non-inbred crops, non-additive effects are further partitioned into dominance and residual non-additive effects. Residual non-additive effects may include inbreeding depression effects, homozygous dominance effects, the covariance between additive and dominance effects and epistatic effects. No attempt to partition these latter terms is made here. In order to fit dominance effects, the dominance variance and covariance are derived under varying levels of inbreeding. This extension is crucial, for example, in the analysis of METs for partially-inbred hybrid crops such as sugarcane, sorghum and maize where the parents of lines are inbred and the commercial line themselves are F1 hybrids and for other populations with generalized structures. Since dominance genetic effects are determined by the parents of an individual line they can be used in determining whether a particular combination of parents is beneficial. The inclusion of non-additive residual effects that could account for enhanced or reduced performance of particular lines ensures that an overall or total genetic performance can be obtained from the same analysis.

The approach presented here is a mixed model form of a classical quantitative genetics model. It follows a long and ongoing tradition to attempt to model the gene to phenotype relationship (see Cooper and Hammer 2005 for a recent review).

Materials and methods

Motivating experiment

The data considered in this paper were taken from the joint sugar breeding program of BSES Ltd and the Commonwealth Scientific Industrial Research Organisation (CSIRO). A large number of clones were evaluated (and selected) in 2002 at two sites in South East Queensland in 'Stage 2' or clonal assessment trials (CATs). The CATs involved clones planted in a single 10 m plot, interspersed with multiple plots of (the same) four commercial lines in a grid-plot layout. Land availability at the sites governed the spatial layout and resulted in two contiguous row by column arrays of plots (subtrials) at the MQN site (site is synonymous with trial). A selected set of 80 clones from these CATs was then planted in four 'Stage 3' or final assessment trials (FATs) in 2003. Each FAT was designed as a latinized row-column design (John et al. 2002) with two replicates (and included additional plots of 25 commercial lines) using the software CycDesigN (Whitaker et al. 2006). Again land availability at each site necessitated two contiguous arrays or subtrials. Plots were four drill rows by 10 m with data recorded from the middle two drill rows to ameliorate the effect of inter-plot competition on the data. Hereafter clones are synonymous with lines. Table 1 presents a summary of information for each site including the design layout.

The pedigree of all of the lines in the clonal and final assessment trials and their parents was available resulting in pedigree information on 2,663 individuals, over several generations. The data considered here are plant cane measures of commercial cane sugar (CCS, %). CCS is an industry formula and estimates the percentage of recoverable sucrose in the cane on a fresh weight basis (BSES 1984).

Statistical model

Following Smith et al. (2001), an appropriate statistical model for $\boldsymbol{y}^{(n\times 1)}$ the full vector of data for individual plots is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{Z}_g\boldsymbol{g} + \boldsymbol{Z}_u\boldsymbol{u} + \boldsymbol{\eta} \qquad (1)$$

where $\boldsymbol{\tau}^{(b\times 1)}$ is the vector of fixed parameters and includes an overall mean performance for each site as well as subtrial specific modeling terms (see Gilmour et al. 1997) with associated design matrix $\boldsymbol{X}^{(n\times b)}$, $\boldsymbol{g}^{(mp\times 1)} = (\boldsymbol{g}_1^{\mathrm{T}},\ldots,\boldsymbol{g}_p^{\mathrm{T}})^{\mathrm{T}}$ is the vector of random genetic effects of the m lines in each of p sites, with associated design matrix $\boldsymbol{Z}_g^{(n\times mp)}$, $\boldsymbol{u}^{(c\times 1)}$ is the vector of random effects and includes extraneous environmental variation specific to each subtrial, and design or randomization based blocking factors (Cullis et al. 2006), $\boldsymbol{Z}_u^{(n\times c)}$ is its associated design matrix and $\boldsymbol{\eta}^{(n\times 1)} = (\boldsymbol{\eta}_1^{\mathrm{T}},\ldots,\boldsymbol{\eta}_t^{\mathrm{T}})^{\mathrm{T}}$ is the residual vector partitioned conformably within t subtrials. The residual vector $\boldsymbol{\eta}_s^{(n_s\times 1)}$ represents local stationary variation at the sth sub-trial ($s = 1,\ldots,t$). It is the sum of two independent vectors, $\boldsymbol{\varsigma}_s^{(n_s\times 1)}$ representing a spatially dependent mean zero random stationary process and $\boldsymbol{\zeta}_s^{(n_s\times 1)}$ a zero mean process representing measurement error. The measurement error term $\boldsymbol{\zeta}_s$ has variance $\sigma_s^2\boldsymbol{I}_{n_s}$ and the spatial dependent term $\boldsymbol{\varsigma}_s$ has variance $\sigma_{e_s}^2\boldsymbol{\Sigma}_s^{(n_s\times n_s)}$, where the matrix $\boldsymbol{\Sigma}_s = (\boldsymbol{\Sigma}_{c_s}\otimes\boldsymbol{\Sigma}_{r_s})$,

Table 1 Summary of the design layout and other details of the sugar example subtrials

Year	Site	Type	Clones[a]	Mean CCS (%)	Subtrial	Columns	Rows	Plots[b]
2002	BIN1	CAT	1236	11.37	1	30	46	1380
2002	MQN	CAT	1010	14.29	2	16	58	1144
					3	8	27	
2003	BIN2	FAT	105	13.52	4	14	8	224
					5	14	8	
2003	FMD	FAT	105	16.22	6	16	7	224
					7	16	7	
2003	ISS	FAT	105	13.98	8	14	8	224
					9	14	8	
2003	MYB	FAT	105	13.73	10	16	7	224
					11	16	7	

FAT final assessment trial, *CAT* clonal assessment trial, *CCS* commercial cane sugar

[a] Total number of clones planted for each site (across subtrials)

[b] Total number of plots for each site (across subtrials)

represents the kronecker product between auto-regressive processes of order one (AR1) in the column and row directions, respectively, for the sth sub-trial. Notice that $n = \sum_{r=1}^{p} n_r = \sum_{s=1}^{t} n_s$, where n_r is the number of observations in the rth site, and n_s is the number of observations in the sth sub-trial.

It is assumed that $\boldsymbol{g}$, $\boldsymbol{u}$ and $\boldsymbol{\eta}$ are pairwise independent with var $(\boldsymbol{u}) = \boldsymbol{G}_u$ and var $(\boldsymbol{\eta})$ is a block diagonal matrix with t blocks corresponding to subtrials, with block s of the form $\boldsymbol{R}_s = \sigma_s^2 \boldsymbol{I}_{n_s} + \sigma_{e_s}^2 \boldsymbol{\Sigma}_s$. The structure for genetic line effects is discussed below.

Thus, the genetic line effects $\boldsymbol{g}$ reflect the genetic variation, while the fixed $\boldsymbol{\tau}$, random $\boldsymbol{u}$ and residual $\boldsymbol{\eta}$ terms reflect the design and conduct of the trial, and as such, provide the underlying structure for non-genetic variation.

Modeling genetic effects

The vector of genetic line effects $\boldsymbol{g}$, termed total genetic value, is decomposed into three components: additive, dominance and residual non-additive genetic effects. Thus $\boldsymbol{g} = \boldsymbol{a} + \boldsymbol{d} + \boldsymbol{i}$, where the terms in the sum have obvious meaning and it is further assumed that these are mutually independent, zero mean Gaussian random vectors. It then follows that a sensible model for the variance of $\boldsymbol{g}$, referred to here as the *Extended* model for $\boldsymbol{g}$ is

$$\text{var}(\boldsymbol{g}) = \boldsymbol{G}_a \otimes \boldsymbol{A} + \boldsymbol{G}_d \otimes \boldsymbol{D} + \boldsymbol{G}_i \otimes \boldsymbol{I}_m \tag{2}$$

The additive, dominance and residual non-additive genetic variance matrices across sites are $\boldsymbol{G}_a$, $\boldsymbol{G}_d$ and $\boldsymbol{G}_i$, respectively. These matrices have diagonal elements that are the genetic variances for the individual sites and off-diagonal elements that are the genetic covariances between pairs of sites. The form of these matrices for the different genetic terms need not be the same. Variance models for $\boldsymbol{G}_\gamma$, $\gamma = a, d, i$, range from a compound symmetry structure where all sites have the same variance and all pairs of sites have the same covariance (Patterson et al. 1977); to a completely unstructured form for p sites of $p(p+1)/2$ parameters with different site variances and covariances between sites. Cullis et al. (1998) consider a model for $\boldsymbol{G}_\gamma$ that includes a separate variance for each site and the same covariance for pairs of sites. Smith et al. (2001), consider a factor analytic structure for $\boldsymbol{G}_\gamma$ with up to l factors ($l < p$), so that $\boldsymbol{G}_\gamma = \boldsymbol{\Lambda}_\gamma \boldsymbol{\Lambda}_\gamma^{\mathrm{T}} + \boldsymbol{\Psi}_\gamma$, with $\boldsymbol{\Lambda}_\gamma$ being a matrix of factor loadings at each of the p sites, and the matrix Ψ is a diagonal matrix with elements ψ_s the specific variance for site s.

Notice that the *Extended* model has the *Standard* model as a sub-model. In the *Standard* model $\boldsymbol{g}$ is not partitioned so that var $(\boldsymbol{g}) = \boldsymbol{G}_i \otimes \boldsymbol{I}_m$, where $\boldsymbol{I}_m$ is a $(m \times m)$ identity matrix. Thus, in the *Standard* model, an overall random genetic effect is fitted where lines are assumed independent.

The matrix $\boldsymbol{A}^{(m \times m)} = \{ A_{jk} \}$ is the known additive relationship matrix (Henderson 1976) between lines j and k, such that

$$A_{jk} = \begin{cases} 1 + F_j, & j = k \\ 2f_{jk}, & j \neq k \end{cases} \tag{3}$$

where $2f_{jk}$ is the numerator of the coefficient of relationship (Wright 1922), f_{jk} is the coefficient of parentage between lines j and k and F_j is the inbreeding coefficient of line j. The matrix $\boldsymbol{D}^{(m \times m)} = \{D_{jk}\}$ is the known dominance relationship matrix between line j who has parents Y and Z and line k, who has parents U and V, such that

$$D_{jk} = \begin{cases} 1 - F_j, & j = k \\ (f_{YU} f_{ZV} + f_{YV} f_{ZU})(1 - F_j)(1 - F_k), & j \neq k \end{cases} \tag{4}$$

where terms are defined as above. The case for $j = k$ was derived by Harris (1964) and for $j \neq k$ the result is derived in Verbyla and Oakey (2006), however a short summary of the derivation is provided in Appendix 2.

If the equalities $A_{YU} = 2f_{YU}$ and $F_j = 0.5A_{YZ}$ (Henderson 1976; Falconer and Mackay 1996) are noted, then an alternative formulation of Eq. 4 which allows us to determine $\boldsymbol{D}$ from the elements of $\boldsymbol{A}$ is

$$D_{jk} = \begin{cases} 1 - 0.5A_{YZ}, & j = k \\ 0.25(A_{YU}A_{ZV} + A_{YV}A_{ZU})(1 - 0.5A_{YZ})(1 - 0.5A_{UV}), & j \neq k \end{cases} \quad (5)$$

There are several special cases of interest. Firstly, consider completely inbred lines. Thus $A_{YZ} = A_{UV} = 2$ and Eq. 5 reduces to $D_{jj} = D_{jk} = 0$ so that there is no dominance. Cockerham (1954) considers lines that are not inbred. In this case $A_{YZ} = A_{UV} = 0$ and Eq. 5 reduces to:

$$D_{jk} = \begin{cases} 1, & j = k \\ 0.25(A_{YU}A_{ZV} + A_{YV}A_{ZU}) & j \neq k \end{cases} \quad (6)$$

For lines that are full sibs and have (some) inbreeding, Eq. 5 reduces to

$$D_{jk} = \begin{cases} 1 - 0.5A_{YZ}, & j = k \\ 0.25(A_{YY}A_{ZZ} + A_{YZ}^2)(1 - 0.5A_{YZ})^2 & j \neq k \end{cases} \quad (7)$$

while for full-sib lines which are not inbred, Eq. 7 reduces to

$$D_{jk} = \begin{cases} 1, & j = k \\ 0.25(A_{YY}A_{ZZ}) & j \neq k \end{cases} \quad (8)$$

Equation 8 was also given by Stuber and Cockerham (1966) and Cockerham and Weir (1984) but in terms of F_Y and F_Z.

For the experiment presented in this paper, the lines are hybrid crops so that either $A_{YZ} = 0$ or $0 \leq A_{YZ} < 1$. It is possible that $A_{YZ} > 0$ due to relationships between the parents of the lines, so that D_{jj} and D_{jk} are greater than zero.

Fitting the dominance genetic effect $\boldsymbol{d}$

The inverses of the additive and dominance relationship matrices are required for the mixed model equations (Henderson 1950) and therefore for the calculation of additive and dominance genetic effects. There are several algorithms (Henderson 1976; Quaas 1976; Meuwissen and Luo 1992) for the direct calculation of the inverse of the additive relationship matrix and therefore no obstacles to fitting this term. However, at present, there is no algorithm to calculate the inverse of the dominance relationship matrix directly. This is because the elements of the dominance relationship matrix are functions of the elements of the additive relationship matrix and it is difficult to see how the inverse of the dominance relationship matrix can be calculated directly without first calculating the additive relationship matrix. Therefore, for large data sets obtaining the inverse of the dominance relationship matrix (using conventional rules for inverting matrices) may be a limiting factor to the fitting of dominance genetic effects.

Hoeschele and VanRaden (1991) noted that the dominance relationship between two individuals is defined by the relationships between their parents. Individuals from the same family (i.e. same parents) therefore share the same dominance relationships. If a pedigree contains many individuals from the same family, the dominance relationship between these individuals can be summarized in a reduced form by considering two components; one relating to between family effects and the other relating to within family line effects (Hoeschele and VanRaden 1991). In the example presented here, the 2,267 lines are from 187 families. Thus, by partitioning $\boldsymbol{d}$ the potential information required to be input in the form of dominance relationships between lines can be reduced almost 130-fold from a potential maximum of 2,570,778 data points to a potential maximum of 19,845 data points. The inverse of the smaller between family dominance matrix thus can be obtained using conventional rules for inverting matrices with little difficultly.

Hoeschele and VanRaden (1991) suggested that the between family effects could be included in the model and the within family line effects be obtained by back-solving. Here we extend their approach by including both the between family effects and within family line effects in the model. This means that the total dominance effect is predictable. We also present the approach under inbreeding.

The vector of dominance effects $\boldsymbol{d}^{(mp \times 1)} = \{d_{jr}\}$, where d_{jr} is the dominance effect of the jth line ($j = 1,...,m$) in the rth site ($r = 1,...,p$), can thus be partitioned (without loss of information) into two mutually independent vectors: a vector of dominance effects relating to between family effects $\boldsymbol{d}_b^{(vp \times 1)} = \{d_{bqr}\}$, where d_{bqr} is the dominance between family effect for the qth family (with $q = 1,2,..., v$, $v < m$) in the rth site and a vector of dominance effects relating to within family line effects $\boldsymbol{d}_w^{(mp \times 1)} = \{d_{wjr}\}$, where d_{wjr} is the within family line effect for the jth line in the rth site.

A particular line j from family q and site r will have its dominance effect defined as

$$d_{jr} = d_{bjr} + d_{wjr} = d_{bqr} + d_{wjr}$$

where d_{bjr} is equivalent to the between family effect d_{bqr}, and d_{wjr} is as defined above. Thus $\boldsymbol{d}$ can be written as

$$\boldsymbol{d} = \boldsymbol{Z}_b\boldsymbol{d}_b + \boldsymbol{d}_w \tag{9}$$

where $\boldsymbol{Z}_b$ is a ($mp \times vp$) matrix relating lines to families within sites.

The between family dominance effect $\boldsymbol{d}_b$ has distribution $\boldsymbol{d}_b \sim N(\boldsymbol{0},\boldsymbol{G}_d \otimes \boldsymbol{D}_b)$, where $\boldsymbol{D}_b^{(v\times v)} = \{D_{bq_\alpha q_\beta}\}$ is the known between family dominance relationship matrix for families q_α and q_β with parents Y, Z and U, V, respectively. The dominance within family line effect $\boldsymbol{d}_w$ has distribution $\boldsymbol{d}_w \sim N(\boldsymbol{0},\boldsymbol{G}_d \otimes \boldsymbol{D}_w)$, where $\boldsymbol{D}_w^{(m\times m)} = \text{diag}\{D_{wj}\}$ is the known within family line dominance relationship matrix for individual j. The elements of $\boldsymbol{D}_b$ and $\boldsymbol{D}_w$ are now developed.

$\boldsymbol{D}_b$ is a symmetric covariance–variance matrix with diagonal terms which correspond to the between family variance and the off-diagonal terms which correspond to covariances between families. Hoeschele and VanRaden (1991) noted that if j and k are lines in the same family q with the same parents Y and Z (i.e. they are full sibs), that

$$\begin{aligned}\text{cov}(d_{bjr}, d_{bkr}) &= \text{cov}(d_{bqr}, d_{bqr})\\ \text{cov}(d_{bjr}, d_{bkr}) &= \text{var}(d_{bqr})\end{aligned} \tag{10}$$

(*Note*: here in addition to Hoeschele and VanRaden (1991), it is assumed that j and k are both from site r for completeness. However, because of the separable nature of var($\boldsymbol{d}$) the r could be omitted without loss of information.) Therefore, Eq. 10 indicates that the diagonal terms of $\boldsymbol{D}_b$ are defined by the covariances between full-sibs (Eq. 7). There are three scenarios depending on whether the parents Y and Z of family q_α are known. Thus the diagonal terms of $\boldsymbol{D}_b$ are:

$$D_{bq_\alpha q_\alpha} = \begin{cases} 0.25(A_{YY}A_{ZZ} + A_{YZ}^2)(1 - 0.5A_{YZ})^2, & \text{if both parents are known}\\ 0.25A_{YY}, & \text{if one parent, say, } Y, \text{ is known}\\ 0.25, & \text{if neither parent is known}\end{cases} \tag{11}$$

When one parent is known, it is assumed (as in the derivation of the $\boldsymbol{A}$ matrix by Henderson 1976), that $A_{ZZ} = 1$. When neither parent is known, in addition it is assumed that $A_{YY} = 1$ and $A_{YZ} = 0$.

The off-diagonal terms of $\boldsymbol{D}_b$ are non-zero only when both parents of both families q_α and q_β are known and are given by

$$\begin{aligned}D_{bq_\alpha q_\beta} = {} & 0.25(A_{YU}A_{ZV} + A_{YV}A_{ZU})(1 - 0.5A_{YZ})\\ & \times (1 - 0.5A_{UV})\end{aligned} \tag{12}$$

Hoeschele and VanRaden (1991) showed that the diagonal terms of $\boldsymbol{D}_w$ are defined as

$$\text{var}(d_{jr}) = \text{var}(d_{bjr}) + \text{var}(d_{wjr})$$

so that

$$\text{var}(d_{wjr}) = \text{var}(d_{jr}) - \text{var}(d_{bqr}) \tag{13}$$

(*Note*: again in addition to Hoeschele and VanRaden (1991), it is assumed that j and k are both from site r for completeness. However, because of the separable nature of var ($\boldsymbol{d}$) the r could be omitted without loss of information.) Using Eq. 13, the diagonal terms of $\boldsymbol{D}_w$, under three scenarios are

$$D_{wj} = \begin{cases} (1 - 0.5A_{YZ}) - 0.25(A_{ZZ}A_{YY} + A_{YZ}^2)(1 - 0.5A_{YZ})^2, & \text{if both parents are known}\\ 1 - 0.25A_{YY}, & \text{if one parent, say, } Y, \text{ is known}\\ 1 - 0.25 = 0.75, & \text{if neither parent is known}\end{cases} \tag{14}$$

When one parent is known, it is assumed (as in the derivation of the $\boldsymbol{A}$ matrix by Henderson 1976), that $A_{ZZ} = 1$. When neither parent is known, in addition it is assumed that $A_{YY} = 1$ and $A_{YZ} = 0$. Recall that $\boldsymbol{D}_w$ is a diagonal matrix, so the off-diagonal terms are zero.

The variance matrix of $\boldsymbol{d}$ can thus be written in terms of $\boldsymbol{D_b}$ and $\boldsymbol{D_w}$ is

$$\text{var}(\boldsymbol{d}) = \text{var}(\boldsymbol{Z}_b\boldsymbol{d}_b + \boldsymbol{d}_w) = \boldsymbol{Z}_b\boldsymbol{D}_b\boldsymbol{Z}_b^{\mathrm{T}} + \boldsymbol{D}_w = \boldsymbol{D} \tag{15}$$

Implementing this modeling strategy, $\boldsymbol{d}_b$ and $\boldsymbol{d}_w$ are fitted as separate random terms with $\boldsymbol{G}_d$ constrained to be equal for both terms. This implies for instance, in the case of a factor analytic structure for $\boldsymbol{G}_d$, that the factor loadings and the specific variances are constrained to be the same for both random terms. By partitioning the dominance effects $\boldsymbol{d}$ with symmetric dominance relationship matrix $\boldsymbol{D}$ of size ($m \times m$) the prediction of $\boldsymbol{d}$ becomes a reduced problem which will be more computationally feasible. This is because $\boldsymbol{D}_b$ is a symmetric matrix of size ($v \times v$), where v may be much smaller than m; and $\boldsymbol{D}_w$ is a diagonal matrix of size ($m \times m$).

The rules for obtaining the elements of $\boldsymbol{D}_b$ and $\boldsymbol{D}_w$ under the scenarios, both parents known, one parent known and no parents known presented above can be used as the basis of a computer program to form these matrices. The full or usual pedigree is needed to form $\boldsymbol{A}$ and a reduced pedigree

based on familial relationships needs to be used to form $\boldsymbol{D}_b$ and $\boldsymbol{D}_w$.

Estimation and fitting

When fitting the models described above, a hierarchical or incremental approach must be taken. In the first instance we fit the *Standard* model to determine the non-genetic or environmental parameters appropriate for each subtrial. Examination of diagnostics includes plotting a sample variogram for examining spatial covariance structure and plots of residuals against row(column) number for each column(row) (see Gilmour et al. 1997 for details) determines which (if any) spatial terms may be needed. Once an appropriate non-genetic model is determined, the genetic effects of the *Extended* model can be incorporated and fitted. There will be situations where one or more of the REML estimates of the additive, dominance and residual non-additive genetic variances are zero at a particular site; thus the particular component is not present. This also means that correlations between the sites with zero estimated variance and other sites cannot be estimated. To determine if genetic variance is present for each component at each site, a model which assumes zero correlations between sites is initially fitted for all three components. Variance models for $\boldsymbol{G}_a$, $\boldsymbol{G}_d$ and $\boldsymbol{G}_i$ can then be chosen which exclude sites with no estimable additive, dominance or residual non-additive variance, respectively.

For sites with positive variances we aim to fit a factor analytic structure as these have been shown to work well in practice (Smith et al. 2005). However, factor analytic structures can be difficult to fit. For a single factor model, simpler models should be used as a basis for initial parameter estimates. It is recommended that the model of Cullis et al. (1998) be fitted and initial values from this be used in the factor analytic structure with one factor. In particular, the environment variance estimates are used for initial estimates of the specific variances. If the number of sites is greater than 4, then a factor analytic structure with two factors can also be attempted with initial estimates based on the results of the one factor model.

For models which are not nested the goodness of fit of models is compared using the Akaike Information Criterion (AIC, Aiaike 1974). Models with smaller AIC values are superior in terms of fit and parsimony (number of variance parameters). The models discussed here are fitted using the software ASReml (Gilmour et al. 2006). Estimation of variance parameters is by residual maximum likelihood (REML, Patterson and Thompson 1971), using the average information REML algorithm (Gilmour et al. 2006). Given estimates of the variance components Empirical Best Linear Unbiased Estimates (E-BLUEs) are obtained for fixed effects and Empirical Best Linear Unbiased Predictors (E-BLUPs) for random effects. An example of the ASReml code to fit the final *Extended* model is included in Appendix 1. The relationship matrix or it's inverse is required. ASReml will calculate the inverse of the additive relationship matrix directly if supplied with the appropriate pedigree file. R code (R Development Core Team 2005) is available and can be obtained from the corresponding author to calculate the dominance relationship matrices. This will be incorporated into version 3 of ASReml (Gilmour et al. 2006).

Analysis

A summary of the models chosen to account for the non-genetic component of the data is presented in Table 2. The REML estimates of the spatial correlations (AR1 parameters) for columns and rows, respectively, are from Model 10 (see Table 3). All of the models had these same environmental or non-genetic terms fitted. Blocking terms fitted but not shown in this table included a subtrial effect and a site by replicate effect (see Appendix 1 for detail).

The genetic effects are summarized in Table 3, (many of the abbreviations for variance structure are consistent with ASReml syntax). Specifically, for each multi-environment analyses the structure of the site genetic variance matrix G_γ for each genetic component $\gamma = a, d, i$ is shown. Models 2 and 3 are equivalent to fitting a separate analysis at each site because they assume a separate variance for each site and no covariance between sites. Model 2 corresponds to the *Pedigree* model of Oakey et al. (2006) and partitions the genetic line effect into an additive and a general non-additive genetic effect. Model 3 further partitions the non-additive component into dominance and residual non-additive components. The *Extended* model for $\boldsymbol{g}$ is more appropriate here because the clones are F1 hybrids in contrast to the wheat example in Oakey et al. (2006), where lines were inbred and (assumed) homozygous. Thus the non-additive component can and should be partitioned. The remaining models fitted are MET analyses. Model 1 corresponds to a form of the *Standard* model such as that fitted by Smith et al. (2001) where $\boldsymbol{g}$ is not partitioned. The non-genetic terms fitted at each site (Table 2) are determined from this model and then used when fitting further models. Comparing the AIC of Models 1–3, both the single site analyses which partition the genetic line effect into components provide a better fit than the *Standard* MET analysis (Table 3).

Model 4 (Table 3) provides only an additive genetic component (Crossa et al. 2006) and is referred to here as the *Additive* model. Model 5 is the multi-environment extension of the *Pedigree* model of Oakey et al. (2006). Model 5 has a much lower AIC and is therefore a better fit than Model 4. Models 4 and 5 have been fitted for com-

Table 2 Non-genetic terms (excluding blocking terms, blocking terms fitted include subtrial and site by replicate) used in the MET analysis of the sugar example

Site	Subtrial	Random	Fixed	[a]Column AR1	[a]Row AR1
BIN1	1	Column, row		0.09	0.15
MQN	2			0.26	0.22
	3	Column		0.17	0.10
BIN2	4		lin(row), lin(column)	0.59	0.52
	5		lin(row)	0.06	0
FMD	6			0.36	0.21
	7			0.0	0.13
ISS	8			0	0.03
	9			0	0.13
MYB	10			0	0.33
	11	Row		0	0.02

[a] Column and row correlations presented were from the final model (Model 10, Table 3)

parison purposes only and are not recommended as the models of choice for F1-hybrid data. Model 5 is however appropriate if the data consist solely of fully inbred lines where the dominance component is assumed to be zero. Models 6–10 are all MET analyses which use the *Extended* model for the genetic line effect, but have different structures for the site genetic variance matrix $\boldsymbol{G}_a$, $\boldsymbol{G}_d$ and $\boldsymbol{G}_i$ for each of the genetic components $\boldsymbol{a}$, $\boldsymbol{d}$ and $\boldsymbol{i}$, respectively. Models 6 and 7 are the poorest performing *Extended* MET models. Model 6 is the *Extended* model of Patterson et al. (1977) and Model 7 is the *Extended* model of Cullis et al. (1998). All of the *Extended* MET models (excluding Model 6) are superior to Model 4 which fits only additive effects. As discussed previously, the models (Table 3) are fitted in a hierarchial order so that the choice of models fitted further down the table depend on the results of the previous models. For example, Models 7 through 10 have structures for $\boldsymbol{G}_d$ and $\boldsymbol{G}_i$ that are fitted at a reduced set of sites, because having examined Model 3, the REML estimates of some of the site variances of $\boldsymbol{G}_d$ and $\boldsymbol{G}_i$ converged to zero. In particular, for $\boldsymbol{G}_d$, two site variances (MQN and MYB) converged to zero and for $\boldsymbol{G}_i$, three site variances (BIN2, FMD and ISS) converged to zero.

On comparing the AIC of the models fitted Models 8, 9 and 10 are the best performing models with little difference between Models 9 and 10 (Table 3). However, Model 10 has the lowest AIC and therefore it is chosen as the most appropriate and final model. The results of the final model are now examined. The REML estimates of the additive, dominance and residual non-additive genetic variance matrices for sites are summarized (Table 4).

The full genetic variance involves not only the $\boldsymbol{G}_a$, $\boldsymbol{G}_d$ and $\boldsymbol{G}_i$, but also $\boldsymbol{A}$ and $\boldsymbol{D}$. The REML estimates of the average variance of the additive, dominance and residual non-additive line effects of the final model is shown (Table 5). These are REML estimates of the diagonal elements of $\boldsymbol{G}_a$, $\boldsymbol{G}_d$ and $\boldsymbol{G}_i$ (Table 4), with $\boldsymbol{G}_a$ and $\boldsymbol{G}_d$ being multiplied by the average of the diagonal elements of the $\boldsymbol{A}$ and $\boldsymbol{D}$, respectively.

Table 3 Summary of models fitted showing the structure of the site genetic variance matrix for each of the genetic components

Model	Structure of site genetic variance matrix			q[a]	AIC[b]
	$\boldsymbol{G}_a$	$\boldsymbol{G}_d$	$\boldsymbol{G}_i$		
1	–	–	XFA2	59	3178.94
2	DIAG	–	DIAG	53	3126.10
3	DIAG	DIAG (1, 3, 4, 5)	DIAG (1, 2, 6)	54	3124.10
4	XFA1	–	–	53	3057.64
5	XFA1	–	XFA1	65	3028.62
6	CS	CS	CS	47	3076.80
7	DIAG/CS	DIAG/CS (1, 3, 4, 5)	DIAG/CS (1, 2, 6)	57	3051.82
8	XFA1	DIAG/CS (1, 3, 4, 5)	DIAG/CS (1, 2, 6)	62	1130.56
9	XFA1	XFA1 (1, 3, 4, 5)	XFA1 (1, 2, 6)	67	0.42
10[c]	XFA2	XFA1 (1, 3, 4, 5)	XFA1 (1, 2, 6)	72	0.00

[a] q number of parameters fitted

[b] AIC are relative the Model 10, so that positive values indicate the AIC is higher than Model 10

[c] Final model

Key: *CS* same site variance, same covariance between sites (Patterson 1977); *DIAG* different site variance, no covariance between sites, equivalent to fitting a single site analysis; *DIAG/CS* different site variance, same covariance between sites (Cullis et al. 1998); *XFAl* factor analytic with *l* factors (Smith et al. 2001); *(sites)* subset of sites fitted (*note: if not specified all sites fitted*); *AIC* Akaike Information Criteria (Akaike 1974)

Table 4 REML estimate of the components of the additive, dominance and residual non-additive genetic variance matrices[a] for sites of the final model (Model 10, Table 3)

G_a	BIN1	MQN	BIN2	FMD	ISS	MYB
BIN1	0.33	0.89	0.95	–0.05	0.73	0.65
MQN		0.52	0.89	0.15	0.72	0.72
BIN2			2.38	0.21	0.79	0.81
FMD				0.46	0.34	0.73
ISS					0.94	0.76
MYB						1.45

G_d	BIN1	BIN2	FMD	ISS
BIN1	0.67	0.74	0.44	1.00
BIN2		0.94	0.33	0.74
FMD			0.24	0.44
ISS				0.47

G_i	BIN1	MQN	MYB
BIN1	0.84	0.24	–0.19
MQN		0.51	–0.77
MYB			0.19

[a] These matrices are symmetric therefore only the upper triangle is shown the diagonal elements of these matrices are variance components and the off-diagonal elements are correlations between sites

At five sites, the non-additive component of variance was composed of *either* dominance *or* residual non-additive variance. At only one site was *both* dominance and residual non-additive variance estimable. If selection was solely on the basis of CCS, then it may be expected that the genetic variance in the FATs would be less than that observed in the CATs. However, selection to progress clones from the FATs to the CATs was based on Net Merit Grade which is only weakly associated with CCS. For the clonal trials (Site BIN1 and MQN) in particular, the non-additive variance comprised a greater proportion of the total variance than at the other sites. This would suggest that the FATs are the more appropriate trials to select the best parents from as these have a much higher proportion of estimated additive genetic variation than the CATs. Indeed, this is the current practice in BSES-CSIRO breeding programs.

Table 5 Summary of the REML estimate of the percent average variance[a] of the each of genetic components of the final model (Model 10, Table 3)

Site	Type	%var(***a***)	%var(***d***)	%var(***i***)	var(***g***)
BIN1	CAT	19.3	34.5	46.2	1.810
MQN	CAT	52.3	0	47.7	1.060
BIN2	FAT	74.3	25.7	0	3.407
FMD	FAT	68.7	31.3	0	0.711
ISS	FAT	69.3	30.7	0	1.442
MYB	FAT	89.1	0	10.9	1.725

FAT final assessment trial, *CAT* clonal assessment trial

[a] These are REML estimates of the diagonal elements of $\boldsymbol{G}_a$, $\boldsymbol{G}_d$ and $\boldsymbol{G}_i$, shown in Table 4, with $\boldsymbol{G}_a$ and $\boldsymbol{G}_d$ being multiplied by the average of the diagonal elements of $\boldsymbol{A}$ and $\boldsymbol{D}$, respectively

To complete the view of the genetic models the genetic correlations were examined (Table 4). Firstly, the genetic variance components differ between sites for all types of genetic effects. The variance for sites differ in their relative sizes across components (additive, dominance and residual non-additive). For the additive component, a strong positive estimated correlation exists between five of the six sites (Table 4). FMD was the exception and shows reduced correlations with all other sites except MYB. For the dominance component, a strong positive estimated correlation exists between three of the four sites; again FMD was the exception showing reduced correlations. For the residual non-additive component, the correlation between MYB and the other sites is negative. In summary, where genetic variation existed at the additive, dominance and residual non-additive levels, the site FMD appears to be different while the other sites tend to perform similarly. This site appeared to have a much lower total genetic variance (var(***g***), Table 5) than other sites.

The main aim of this analysis was to provide line selection. Predictions of genetic line effects for individual sites can be used to form an appropriately weighted selection index for each of the genetic components. Cooper and Podlich (1999) and Podlich et al. (1999) show through computer simulation that weighted selection strategies perform as well or better than the traditional unweighted strategies. Particularly, performance of weighted strategies is better when only a few sites are sampled in a MET or when there is a lack of genetic correlation between sites.

The weights may be chosen in a number of ways. Cooper et al. (1996) suggest giving bigger weights to sites that are more representative of target sites and Kelly et al. (2007) consider merit in equal weights across all sites. Let w_s be the weight for site s and $\tilde{\boldsymbol{a}}_s, \tilde{\boldsymbol{d}}_s (= \tilde{\boldsymbol{d}}_{bs} + \tilde{\boldsymbol{d}}_{ws})$ and $\tilde{\boldsymbol{i}}_s$ be the vectors of genetic line E-BLUPs for the additive, dominance and residual non-additive effects at site s, respectively. The selection index $\boldsymbol{M}_a$ for additive genetic line effects is

$$\boldsymbol{M}_a = w_1\tilde{\boldsymbol{a}}_1 + \cdots + w_6\tilde{\boldsymbol{a}}_6,$$

and the selection index $\boldsymbol{M}_d$ for dominance genetic line effects is

$$\boldsymbol{M}_d = w_1\tilde{\boldsymbol{d}}_1 + \cdots + w_6\tilde{\boldsymbol{d}}_6,$$

For total genetic line effects under the *Extended* model,the selection index $\boldsymbol{M}_g$ for $\boldsymbol{g}$ is

$$\boldsymbol{M}_g = w_1(\tilde{\boldsymbol{a}}_1 + \tilde{\boldsymbol{d}}_1 + \tilde{\boldsymbol{i}}_1) + \cdots + w_6(\tilde{\boldsymbol{a}}_6 + \tilde{\boldsymbol{d}}_6 + \tilde{\boldsymbol{i}}_6)$$

As the selection of FAT lines from the CATs has taken place, the Figs. 1, 2 and 3 show only those lines in the FATs. The six sites were given equal weights in each selection index.

There was little evidence of a relationship between the predicted additive selection index and the (overall) predicted dominance selection index of the lines (Fig. 1). This implied that lines with the highest additive selection indices did not necessarily have the highest dominance selection indices. So in the selection of lines, breeders must trade off between these two values.

A relatively high correlation (0.92) between the predicted selection indices for the total genetic effects of the *Standard* model (Model 1, Table 3) and the final model (Model 10, Table 3) was apparent (Fig. 2). However, in comparison to the final model the *Standard* model generally under-estimates the total selection indices values. There were also important differences in the ranking of some of the lines between the two models. For example, the top ranking line under the *Standard* model is ranked 5th under the final model and the top ranking line under the final model was ranked 12th under the *Standard* model. In addition, when we consider the ranking of the top 20 lines, 4 of the selections are different under the two models.

A positive correlation (0.87) was found between the predicted total genetic selection index of the *Standard* model (Model 1, Table 3) and the additive genetic predicted selection index of the final model (Fig. 3). However, again, there were important differences in the ranking of some of the lines between the two models. For example, the top ranking line under the *Standard* model is ranked 5th under the final model and the top ranking line under the final model was ranked 6th under the *Standard* model. In addition, when we consider the ranking of the top 20 lines, 6 of the selections are different under the two models.

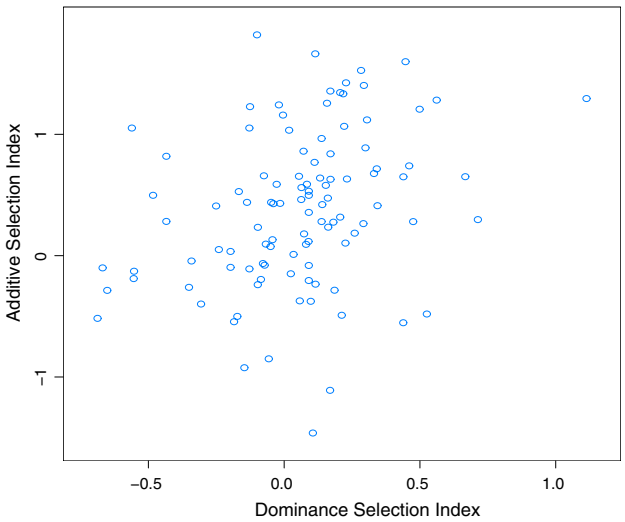

Fig. 1 The predicted *additive* selection index (breeding value index) plotted against the predicted *dominance* selection index of CCS for the final model (Model 10, Table 3)

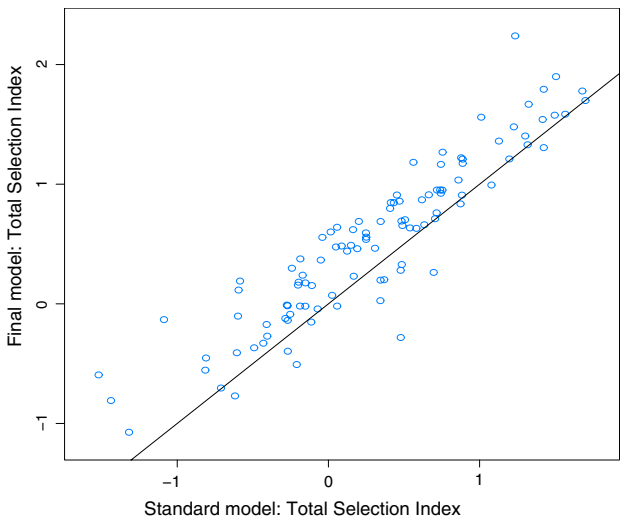

Fig. 2 The predicted *total* selection index of the *Standard* model (Model 1, Table 3) plotted against the predicted *total* selection index of CCS for the final model (Model 10, Table 3)

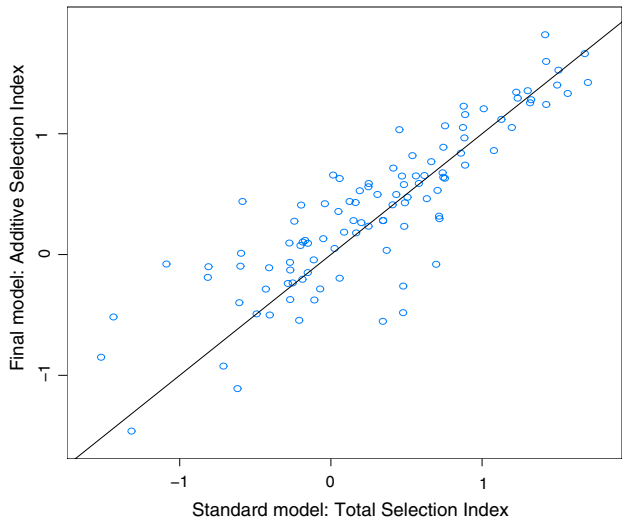

Fig. 3 The predicted *total* selection index of the *Standard* model (Model 1, Table 3) plotted against the predicted *additive* genetic effects (breeding values) of CCS for the final model (Model 10, Table 3)

Discussion

This paper develops a statistical approach that can be used in crop breeding trials with pedigree information and replication of lines. It involves fitting a model that predicts additive and non-additive (dominance and residual non-additive) genetic effects of test lines, simultaneously models spatial variation, and allows for heterogeneity of environmental variance and correlations between environments to be accommodated. It offers advantages over current approaches in that it enables the selection of the best performing line for commercial release, the selection of best parents and best combinations of parents for further crosses in a single analysis and from standard crop breeding trials.

The additive line effects of this model are estimated breeding values and as such are the preferable means of determining potential parents for breeding programs. The dominance line effects give an indication of how well the genes from an individual's parents combined. The residual non-additive line effects may include inbreeding depression effects, homozygous dominance effects, the covariance between additive and dominance effects and epistatic effects which could account for enhanced or reduced performance of a particular line. The overall or total genetic value of a line is obtained from the sum of additive and non-additive effects and is used to determine the commercial worth of a line, as it is the overall performance and therefore overall genetic value that is important.

In trials with only completely inbred lines (eg. wheat and barley) the approach presented in this paper is still applicable although somewhat simplified. Completely inbred lines are assumed homozygous due to inbreeding and therefore the dominance effect of a line is assumed to be zero. As a result the non-additive effects consist only of epistatic effects. This is in fact a multi-environment extension of the ''*Pedigree*'' model seen in Oakey et al. (2006) and was fitted here as an illustration. Recently, including the pedigree information in the form of the additive relationship matrix has been used to predict additive effects or breeding values in plants (Panter and Allen 1995; Durel et al. 1998; Dutkowski et al. 2002; Davik and Honne 2005; Crossa et al. 2006). However, all of these papers fail to account for non-additive effects. Many authors (van der Werf and de Boer 1989; Hoeschele and VanRaden 1991; Lu et al. 1999) have indicated that accounting for non-additive effects might improve the estimation of additive effects. For this example, it was shown that almost all of the *Extended* MET models fitted which included non-additive effects were superior to the model which excluded non-additive effects. Therefore, from these results, we suggest that in data sets with completely inbred lines, it will be important to estimate the non-additive effect in the form of a *Pedigree* model extended for METs. In the case of data sets with F1-hybrid lines the partitioning of the non-additive effect into dominance and residual non-additive effects should be equally important. In animals and outcrossing species such as trees, additive and dominance effects could be obtained using methods given here if a well-structured half-sib design was available.

The hybrid example explored here is sugarcane. Sugarcane is a polyploid, showing more than two copies of the basic set of chromosomes having been derived from interspecific hybridization. It also exhibits aneuploidy, where the chromosome number of a particular individual commonly varies between 100 and 130 chromosomes (Jannoo et al. 2004). A recent study by Jannoo et al. (2004) has shown that pairing in sugarcane at meiosis is predominately bivalent (in pairs), with some non-preferential pairing. The same study shows however that sugarcane shows a combination of disomic and polysomic inheritance. The theoretical developments presented here are derived for disomic inheritance. Therefore, for this specific data set, results from this method will be approximate. Thus, this data set is not an ideal example, but it does provide a practical illustration of the general method presented. Any interactions that are present between chromosomal sets are allowed for by including the non-additive residual component. Given advances in molecular technology, sugarcane and other

polyploids present a good advocate for developing an ***A*** matrix and subsequently a ***D*** matrix from information on the molecular markers of individual lines. This would perhaps provide a more accurate indication of the relationship between individuals and their parents rather than using relationships based only on pedigree information.

Hoeschele and Van Raden (1991) suggested that a computationally feasible way of including dominance effects under no inbreeding is by fitting sire by dam subclass effects (or between family effects) and back solving for the within subclass effects (or within family line effects). The method presented here extends their approach in two ways. Firstly, results are presented under varying levels of inbreeding and, secondly, the within family line effects are included in the model (with the appropriate constraints). This means that by partitioning the dominance effects into the two terms both of which are included in the model we obtain a computationally more feasible approach that is *equivalent* to fitting the complete dominance effect.

It should be noted however, that fitting the dominance relationship matrix by partitioning it into two components as proposed still requires the two dominance relationship matrices to ultimately be inverted, as it is the inverses that are required in the mixed model equations (Henderson 1950). For large data sets, with few full-sib relationships, the ability to invert the between family dominance matrix may still be a limiting factor to using this method as the between family dominance matrix may not be much smaller than the full dominance matrix. For the within family line dominance matrix, the size of this matrix should not be an issue for inversion, even for large data sets since this is a diagonal matrix.

Thus in conclusion, although the model presented is an approximation of the 'true' genetic model, we believe it is a good first practical step and an improvement on current practices.

Acknowledgements Thanks to the Grain Research Development Corporation (GRDC) who fund the research of Helena Oakey through a Grains Industry Research Scholarship. Arūnas Verbyla and Brian Cullis also thank GRDC for support through the National Statistics Program, key program 3. We would also like to thank the Associate Editor and two anonymous referees for their constructive comments and suggestions.

Appendix 1: ASReml code for fitting the final model

The following is the code for the *.as* ASReml file used for fitting the final model.

```
!WORK 500
MET
 subtrial 12 !A
 Site 6 !A
 row 58
 column 30
 replicate 2
 ccs
 Clone !P !LL 26
 lrow
 lcol
 familyB 187
 familyW 2267 !A !SORT
 iSite 3 !A  # Sites BIN1 MYB MQN
 dSite  4 !A # Sites BIN1 BIN2 FMD ISS
data.ped  !skip 1 !ALPHA #pedigree file
DB.grm !skip 1 !GIV #Dominance matrix BETWEEN family
DW.grm !skip 1 !GIV #Dominance matrix WITHIN family
data.asd !skip 1 !mvinclude !maxit 60 !extra 5
ccs~ -1 Site,
at(subtrial,4).lcol at(subtrial,4).lrow at(subtrial,5).lrow,
!r subtrial   xfa(Site,2).Clone  xfa(iSite,1).ide(Clone),
xfa(dSite,1).giv(familyB,1)  xfa(dSite,1).giv(familyW,2) ,
at(subtrial,1).replicate at(subtrial,4).replicate ,
at(subtrial,6).replicate at(subtrial,7).replicate,
at(subtrial,1).row   at(subtrial,11).row,
at(subtrial,3).column  at(subtrial,1).column,
!f mv
11 2 4 !NODISPLAY # number of sites # number of R-str # G-str
14 column   AR 0.59    !S2=2.86   # subtrial 1
8 row       AR 0.50
14 column   AR 0.168   !S2=1.445  # 2
8 row       ID
30 column   AR 0.07125 !S2=1.36   # 3
46 row      AR 0.0819
16 column   AR 0.439   !S2=0.421  # 4
7 row       AR 0.246
16 column   AR 0.104   !S2=0.474  # 5
7 row       AR 0.201
14 column   ID         !S2=0.4205 # 6
8 row       AR 0.01
14 column   AR 0.0814  !S2=0.311  # 7
8 row       ID
16 column   AR 0.29    !S2=0.22   # 8
58 row      AR 0.25
8 column    AR 0.16    !S2=1.04   # 9
27 row      AR 0.103
16 column   AR 0.24    !S2=0.55   # 10
7 row       ID
16 column   AR 0.02    !S2=0.51   # 11
7 row       ID
xfa(Site,2).Clone 2
8  0 XFA2  !+18  !G6P8UZ3U #FOR FA2=3*6
0.13     0.001 0.244  0.265  0.824   0.001
1.29    0.51  0.50   0.83   0.652   1.19
0.1     0.1   0      0.1    0.1     0.1
Clone 0 AINV
xfa(dSite,1).giv(familyB,1) 2
5 0 XFA1  !+8  !GP !=%ABCDEFGH #FOR FA1=2*4
0.7  0.001 0.2   0.001
0.87 0.84  0.04  0.72
familyB 0 GIV1
xfa(dSite,1).giv(familyW,2) 2
5 0 XFA1  !+8  !GP !=%ABCDEFGH #FOR FA1=2*4
0.7  0.001 0.2  0.001
0.87 0.84  0.04 0.72
familyW 0 GIV2
xfa(iSite,1).ide(Clone) 2
4 0 XFA1  !+6  !GP #FOR FA1=2*3
0.85 0.54 0 .1
0.4  0.4  0.4
ide(Clone) 0 ID
```

The *data.ped* is a file containing the pedigree file, from which ASReml calculates the inverse of the relationship matrix $\boldsymbol{A}^{-1}$. ASReml requires a file which has three columns: clone parent1 parent2. The file must be ordered with founding individuals first. DB.grm and DW.grm are the dominance between family and dominance within family line matrices respectively. DW.grm is a scale identity. The *.grm* indicates that these are not inverse matrices (ie. DB.grm is $\boldsymbol{D}_b$ not $\boldsymbol{D}_b^{-1}$) and ASReml will invert them. (A *.giv* ending would indicate that these were inverse matrices). ASReml requires just the lower triangle of this matrices. It is important to ensure that the numbering of lines in the corresponding factors *familyB* and *familyW* corresponds directly to the ordering of rows and columns in the .grm file. Row one and column one of the $\boldsymbol{D_b}$ matrix contain the dominance between relationships of family 1, and this should correspondingly be labeled as 1 in the *familyB* factor, similarly for the *familyW*.

The *data.asd* is a text file containing the data.

The additive genetic effect with a factor analytic structure of order two for $\boldsymbol{G}_a$ is fitted by including the term *xfa(Site,2).Clone* in the random part of the model specification. A factor analytic structure of order one for $\boldsymbol{G}_d$ at four sites is fitted by including the term *xfa(dSite,1).giv(familyB,1)* and *xfa(dSite,1).giv(familyW,2)*, *dSite* has four levels instead of six (the other sites are set to 'NA') and so ensures that a dominance effect is just fitted at these sites and the *.giv(,)* indicates which .grm file to associate with each effect. In addition, these two dominance genetic effects must be constrained to be equal. This is achieved most simply by the *!=%ABCDEFG* command in the G-structure line of both these terms. The residual non-additive genetic effect has a factor analytic structure of order one for $\boldsymbol{G}_i$ at three sites.

Appendix 2: Dominance covariance under varying levels of inbreeding

Harris (1964) derives the covariance between individuals j and k, with parents Y, Z and U, V, respectively, based on whether the alleles of j and k are identical by descent (IBD). He states the coefficient of the dominance covariance between individuals j and k with parents Y, Z and U, V, respectively, to be given by the probability t_{jk}. We now define t_{jk} as in Harris (1964) and then derive this term explicitly below.

$$\begin{aligned} t_{jk} &= p(\alpha_{j_Y} \equiv \alpha_{k_U} \not\equiv \alpha_{j_Z} \equiv \alpha_{k_V}) + p(\alpha_{j_Y} \equiv \alpha_{k_V} \not\equiv \alpha_{j_Z} \equiv \alpha_{k_U}) \\ &= p(\alpha_{j_Y} \not\equiv \alpha_{j_Z}) p(\alpha_{k_U} \not\equiv \alpha_{k_V}) [p(\alpha_{j_Y} \equiv \alpha_{k_U}) p(\alpha_{j_Z} \equiv \alpha_{k_V}) \\ &\quad + p(\alpha_{j_Y} \equiv \alpha_{k_V}) p(\alpha_{j_Z} \equiv \alpha_{k_U})] \\ &= (1 - F_j)(1 - F_k)(f_{YU} f_{ZV} + f_{YV} f_{ZU}) \end{aligned} \tag{16}$$

where α_{j_Y} represents the allele of j derived from parent Y, $\equiv$ indicates identity by descent and the following equivalences given by Cockerham and Weir (1984) are used.

$$\begin{aligned} p(\alpha_{j_Y} \equiv \alpha_{j_Z}) &= F_j \\ p(\alpha_{j_Y} \not\equiv \alpha_{j_Z}) &= 1 - F_j \\ p(\alpha_{j_Y} \equiv \alpha_{k_U}) &= f_{YU} \\ p(\alpha_{j_Y} \not\equiv \alpha_{k_U}) &= 1 - f_{YU} \end{aligned}$$

References

Akaike H (1974) A new look at statistical model identification. IEEE Trans Autom Control AU-19:716–722

Bernardo R (1994) Prediction of maize single-cross performance using rflps and information from related hybrids. Crop Sci 34:20–25

Bernardo R (1996) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

de Boer IJM, Hoeschele I (1993) Genetic evaluation methods for populations with dominance and inbreeding. Theoret Appl Genet 86:245–258

BSES (1984) The Standard laboratory manual for Australian Sugar Mills, vol 1. Bureau of Sugar Experiment Stations, Indooroopilly, QLD, Australia, principles and practices edition

Cockerham CC (1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859–882

Cockerham CC (1983) Covariances of relatives from self-fertilization. Crop Sci 23:1177–1180

Cockerham CC, Weir BS (1984) Covariances of relatives stemming from a population undergoing mixed self and random mating. Biometrics 40:157–164

Cooper M, Hammer GL (2005) Preface to special issue: Complex traits and plant breeding-can we understand the complexities of gene-to-phenotype relationships and use such knowledge to enhance plant breeding outcomes? Aust J Agric Res 56:869–872

Cooper M, Podlich DW (1999) Genotype × environment interactions, selection response and heterosis. In: Coors JG, Pandey S (eds) Genetics and exploitation of heterosis in crops, chap 8, pp 81–92

Cooper M, Brennan P, Sheppard J (1996) A strategy for yield improvement of wheat which accomodates large genotype by environment interactions. In: Cooper M, Hammer GL (eds) Plant adaption and crop improvement, pp 487–512

Costa e Silva J, Borralho NMG, Potts BM (2004) Additive and non-additive genetic parameters from clonally replicated and seedling progenies of *Eucalyptus globulus*. Theoret Appl Genet 108:1113–1119

Crossa J, Burgueno J, Cornelius PL, McLaren G, Trethowan R, Krishnamachari A (2006) Modelling genotype X environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci 46:1722–1733

Cullis BR, Lill W, Fisher J, Read B, Gleeson A (1989) A new procedure for the analysis of early generation variety trials. Appl Stat 38:361–375

Cullis B, Gogel B, Verbyla A, Thompson R (1998) Spatial analysis of Multi-Environment early generation trials. Biometrics 54:1–18

Cullis B, Smith A, Coombes N (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11:381–393

Davik J, Honne B (2005) Genetic variance and breeding values for resistance to wind-borne disease [*Sphaeotheca Macularis* (wallr. ex fr.)] in strawberry (*Fragaria x ananassa* duch.) estimated by exploring mixed models and spatial models and pedigree information. Theoret Appl Genet 111:256–264

Durel CE, Laurens F, Fouillet A, Lespinasse Y (1998) Utilization of pedigree information to estimate genetic parameters from large unbalanced data sets in apple. Theoret Appl Genet 96:1077–1085

Dutkowski GW, Costa e Silva J, Gilmour AR, Lopez GA (2002) Spatial analysis methods for forest genetic trials. Can J For Res 32:2201–2214

Edwards JW, Lamkey KR (2002) Quantitative genetics of inbreeding in a synthetic maize population. Crop Sci 42:1094–1104

Falconer DS, Mackay T (1996) Introduction to quantitative genetics, 4th edn. Longman Group Ltd, London

Gilmour AR, Gogel B, Cullis BR, Thompson R (2006) ASReml, user Guide. Release 2.0. Hemel Hempstead. VSN International Ltd, UK

Gilmour AR, Cullis B, Verbyla AP (1997) Accounting for natural and extraneous variation in the analysis of field experiments. J Agric Biol Environ Stat 2:269–293

Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing systems. Aust J Biol Sci 9:463–493

Harris DL (1964) Genotypic covariances between inbred relatives. Genetics 50:1319–1348

Henderson CR (1950) Estimation of genetic parameters (abstract). The Ann Math Stat 21:309–310

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. Biometrics 32:69–83

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph

Hoeschele I, VanRaden PM (1991) Rapid inversion of dominance relationship matrices for noninbred populations by including sire by dam subclass effects. J Diary Sci 74:557–569

Jannoo N, Grivet L, David L, Glaszmann J-C (2004) Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers. Heredity 93:460–467

John J, Ruggiero K, Williams E (2002) ALPHA(n)-designs. Aust N Z J Stat 44:457–465

Kelly AM, Smith AB, Eccleston JA, Cullis BR (2007) The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Sci (in press)

Lo LL, Fernando RL, Cantet RJC, Grossman M (1995) Theory for modelling means and covariances in a two-breed population with dominance inheritance. Theoret Appl Genet 90:49–62

Lu PX, Huber DA, White TL (1999) Potential biases of incomplete linear models in heritability estimation and breeding value prediction. Can J For Res 29:724–736

Meuwissen THE, Luo Z (1992) Computing inbreeding coefficients in large populations. Genet Select Evol 24:305–313

Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H (2006) Joint modelling of additive and non-additive genetic line effects in single field trials. Theoret Appl Genet 113:809–819

Panter DM, Allen FL (1995) Using best linear unbiased predictions to enhance breeding for yield in soybean: I. choosing parents. Crop Sci 35:397–405

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

Patterson HD, Silvey V, Talbot M, Weatherup STC (1977) Variability of yields of cereal varieties in U.K. trials. J Agric Sci Camb 89:238–245

Podlich DW, Cooper M, Basford KE (1999) Computer simulation of a selection strategy to accommoodate genotype-environment interactions in a wheat recurrent selection programme. Plant Breed 118:17–28

Quaas RL (1976) Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32:949–953

R Development Core Team (2005) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0

Smith A, Cullis B, Thompson R (2001) Analyzing variety by environmental data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57:1138–1147

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci 143:1–14

Stuber CW, Cockerham CC (1966) Gene effects and variances in hybrid populations. Genetics 54:1279–1286

Verbyla AP, Oakey H (2006) The variance-covariance matrix for relatives undergoing mendelian sampling and inbreeding (submitted)

van der Werf JHJ, de Boer IJM (1989) Influence of non-additive effects on estimation of genetic parematers in diarly cattle. J Diary Sci 72:2606–2614

Whitaker D, Williams ER, John JA (2006) Cycdesign 3.0: A package for the computer generation of experimental designs. CycSoftware Ltd, Hamilton, New Zealand

Wright S (1922) Coefficients of inbreeding and relationship. Am Nat 56:330–338