

# Predicting Additive and Non-additive Genetic Effects from Trials Where Traits Are Affected by Interplot Competition

Colleen H. HUNT, Alison B. SMITH, David R. JORDAN, and  
Brian R. CULLIS

There are two key types of selection in a plant breeding program, namely selection of hybrids for potential commercial use and the selection of parents for use in future breeding. Oakey et al. (in *Theoretical and Applied Genetics* 113, 809–819, 2006) showed how both of these aims could be achieved using pedigree information in a mixed model analysis in order to partition genetic effects into additive and non-additive effects. Their approach was developed for field trial data subject to spatial variation. In this paper we extend the approach for data from trials subject to interplot competition. We show how the approach may be used to obtain predictions of pure stand additive and non-additive effects. We develop the methodology in the context of a single field trial using an example from an Australian sorghum breeding program.

**Key Words:** Mixed models; Statistical analysis; Variogram; Spatial trends; Competition; Pure stand; Pedigree information; Parental effects.

## 1. INTRODUCTION

It is widely recognized that data from plant breeding trials often exhibit spatial variation due to the spatial location of plots in the field. Model-based analyses that focus on controlling spatial variation have been shown to result in substantial gains in response to selection. Most of the current spatial approaches involve a direct modeling of trend using a

---

Colleen H. Hunt (✉) is a Senior Biometrician, Queensland Department of Agriculture, Fisheries and Forestry, Hermitage Research Station, 604 Yangan Rd, Warwick, Qld 4370, Australia, and The University of Queensland School of Agriculture and Food Science, Brisbane, Australia (E-mail: [Colleen.Hunt@daff.qld.gov.au](mailto:Colleen.Hunt@daff.qld.gov.au)). Alison B. Smith is a Principal Research Fellow and Brian R. Cullis is a Professor, School of Mathematics and Applied Statistics, Faculty of Informatics, University of Wollongong, Wollongong, Australia. David R. Jordan is a Principal Research Fellow, Queensland Alliance for Agriculture and Food Innovation, Hermitage Research Station, 604 Yangan Rd, Warwick, Qld 4370, Australia. Brian R. Cullis is a Professor, Mathematics Informatics and Statistics, CSIRO, Clayton, Australia.

© 2012 *International Biometric Society*  
*Journal of Agricultural, Biological, and Environmental Statistics*, Volume 18, Number 1, Pages 53–63  
DOI: [10.1007/s13253-012-0117-7](https://doi.org/10.1007/s13253-012-0117-7)

correlation model, the basic premise being that plots that are closer together are more similar (more highly correlated) than plots that are further apart. One such approach is that of Gilmour, Cullis, and Verbyla (1997) which has been used successfully for the analysis of grain yield data from Australian cereal breeding programs for many years. However, these models are not appropriate for traits measured in trials that exhibit interplot competition. An important example of this type in Australia is yield from sorghum breeding trials. Hunt and Jordan (2009) examined sorghum yield (in tonnes per hectare t/ha) for 36 such trials and found evidence of interplot competition in one third of those trials. They suggested that for this type of data a joint modeling approach that can accommodate both interplot competition and spatial trend is desirable.

Stringer (2006) discussed a number of approaches for the joint modeling of interplot competition and fertility trends. In one of these approaches interplot competition was modeled using a random effects analogue of the Besag and Kempton (1986) treatment interference model (TIM). In this model an individual variety is assumed to have both a direct effect (on the plots in which it is grown) and a neighbor effect (on adjacent plots). In the random effects setting these are regarded as (correlated) genetic effects so that competition is modeled at the genetic level. Stringer (2006) analyzed a number of early stage sugarcane trials and found that the random effects treatment interference model (R-TIM) (or a reduced rank version there-of) provided a good fit to the data in many cases. In terms of hybrid yield performance the key trait of interest is yield in a pure stand, that is, the yield from a hybrid that is not surrounded by hybrids of differing genetic background. Predictions of hybrid effects for this trait are easily obtained from the R-TIM as a simple linear combination of the predictions for direct and neighbor effects.

In Australian sorghum breeding programs the aim is primarily to develop new (fully in-bred) parental lines for commercial companies to use within their hybrid breeding programs. Oakey et al. (2006) demonstrated that this aim is best met using a statistical analysis in which pedigree information is incorporated. Oakey et al. (2006) proposed a mixed model for field trial data in which genetic effects are partitioned into additive and non-additive components using an additive relationship matrix while error variation is simultaneously modeled using the spatial techniques of Gilmour, Cullis, and Verbyla (1997). In this paper we propose an extension of the approach in Oakey et al. (2006) that incorporates an R-TIM to accommodate interplot competition. The resultant model enables the partitioning of pure stand genetic effects into additive and non-additive components. Here we consider the analysis of a single trial. Extensions for the analysis of multiple trials will be considered elsewhere.

The paper is arranged as follows. First we introduce a motivating example (Section 2). In Section 3 we present a sequence of statistical models for the analysis of a single field trial. We commence with a base-line analysis then build to an analysis that incorporates pedigree information and accommodates both spatial variation and interplot competition. Results of the application of these methods to the example are given in Section 4.

## 2. MOTIVATING EXAMPLE

Our motivating example is taken from the Queensland Department of Agriculture, Fisheries and Forestry sorghum breeding program. This program runs what is analogous to two separate pedigree breeding programs, one for female parents and one for male parents. All field evaluation of lines within these programs is undertaken using F1 hybrids of combinations between the two pedigree programs. We consider a trial grown in 2008 at the Hermitage Research Station in Warwick Queensland. The trial is a preliminary yield trial for males (PYTM).

The trial contained 791 F1 hybrids, comprising 783 test hybrids, being the result of a cross between a single unreleased female and 783 male parental lines, six commercial F1 hybrids two checks, being F1 hybrids close to release. The experimental design for the trial was a resolvable p-rep design (Cullis, Smith, and Coombes 2006). Test and check F1 hybrids were sown in either one or two plots in the trial, while most commercial F1 hybrids had additional replication.

The sorghum breeding program plants trials in a rectangular array of plots in which we notionally index plots by two factors, namely Rows and Columns. Plots are  $1.5 \times 10$  m comprising two plot-rows of plants. Plots which are row-neighbors (i.e., within the same column) share the longest plot boundary. The prevalence of midge necessitates the inclusion of spray-out rows to allow for access of spraying machinery. These spray-out rows occur at regular intervals across the trial, in pairs every 10 rows. Thus rows (11, 12), (23, 24), ..., (12n - 1, 12n), where  $n$  depends on the total number of rows in the trial, will be spray-out rows. The PYTM trial we consider as an example consisted of 64 rows and 20 columns. The resolvable blocks were aligned so that block 1 occupied rows 1 to 31 and block 2 occupied rows 32 to 64, for all columns. The sizes of the blocks differed due to the occurrence of extra spray-out rows in block two.

The genetic design was determined by the aim. The aim of the PYTM trial is to select roughly 10 % of the F4 male parental lines for promotion to Advanced trials. The PYTM trial represents the first opportunity for selection on yield and therefore the breeder is interested in both family and individual performance. A total of 783 F4 males were crossed with one female. The 783 F4 males were distributed across 48 full-sib families. The number of lines per family varied from 1 to 70 with an average of 17.4. In addition to the phenotypic data from the trial there was pedigree information on 1778 lines, including 61 founder lines (i.e. lines with unknown parents). The average inbreeding coefficient of the F1 hybrids was 0.07, ranging from 0 to 0.24, while the average inbreeding of the ancestral lines was 0.985. The genetic connectivity in the design was high with an average additive correlation of 0.499 between the F1 hybrids. The availability of pedigree information is fundamental to the analysis that follows.

### 3. STATISTICAL METHODS

#### 3.1. EXCLUDING INFORMATION ON PEDIGREES

We begin by considering the analysis which does not use information on pedigrees. Our approach uses the enhanced spatial modeling ideas found in Stefanova, Smith, and Cullis (2009). They describe an approach to the analysis of individual trials using a “hybrid” approach which includes terms in the linear mixed model to account for spatial variation and randomization processes used in the design.

The model for data vector  $\mathbf{y}^{n \times 1} = \text{vec}(\mathbf{Y}^{r \times c})$  can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{gdo}\mathbf{u}_{gdo} + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e} \quad (3.1)$$

where the vectors  $\boldsymbol{\tau}$ ,  $\mathbf{u}_{gdo}$ ,  $\mathbf{u}_p$  represent fixed effects, random variety direct effects and random non-genetic (or peripheral, i.e. design and additional) effects, respectively. The  $\text{vec}()$  operator stacks the columns (here 1, 2, ...,  $c$ ) of its matrix argument into a vector of length  $n = rc$ , where  $r$  is the number of rows in the trial and  $c$  is the number of columns. The additional subscript  $o$  and  $d$ , for the vector of direct effects  $\mathbf{u}_{gdo}$  has been used to distinguish that this vector contains direct effects for entries which are in the data-set, as opposed to entries which are in the pedigree but are not in the data-set. We shall denote the vector of the latter direct effects by  $\mathbf{u}_{gdp}$  (see Section 3.2) and we also introduce neighbor effects in Section 3.3.

All random effects are assumed to follow a Gaussian distribution, with mean zero and each of the three random effect vectors are assumed pair-wise independent. Variance models used for the random and residual effects are given by

$$\begin{aligned} \text{var}(\mathbf{u}_{gdo}) &= \sigma_{gdd}\mathbf{I}_{m_o}, \\ \text{var}(\mathbf{u}_p) &= \bigoplus_{l=1}^b \sigma_{p_l}^2 \mathbf{I}_{q_l}, \\ \text{var}(\mathbf{e}) &= \mathbf{R} = \sigma^2 \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r \end{aligned} \quad (3.2)$$

where we use  $\mathbf{I}_n$  to denote an identity matrix of order  $n$ .  $m_o$  represents the number of hybrids present in the data-set. The symbol  $\otimes$  denotes the Kronecker product and is defined for example in the appendix in Smith, Cullis, and Thompson (2005). The symbol  $\oplus$  denotes the direct sum and is a shorthand method for expressing a block diagonal matrix. For example,

$$\bigoplus_{l=1}^b \sigma_{p_l}^2 \mathbf{I}_{q_l} = \begin{pmatrix} \sigma_{p_1}^2 \mathbf{I}_{q_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{p_2}^2 \mathbf{I}_{q_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \sigma_{p_b}^2 \mathbf{I}_{q_b} \end{pmatrix}. \quad (3.3)$$

The variance models in (3.2) allow for a genetic variance component ( $\sigma_{gdd}$ ), a maximum of  $b$  random non-genetic terms, with the  $l$ th term ( $l = 1 \dots b$ ) having  $q_l$  effects and an associated variance component ( $\sigma_{p_l}^2$ ), a residual variance parameter ( $\sigma^2$ ) and residual

(scaled) covariance structure expressed as a Kronecker product of two (scaled) covariance matrices for the factors which enumerate the two dimensions of the field layout (typically called rows and columns; the factor rows is assigned, by default to the largest dimension of the array). The correlation structure is modeled using a first order separable autoregressive process (AR1) in each direction. The submatrices  $\Sigma_r$  and  $\Sigma_c$  are the scaled covariance matrices for columns and rows, respectively, and are functions of vectors of unknown parameters denoted by  $\phi_r$  and  $\phi_c$ , respectively.

### 3.2. INCLUDING INFORMATION ON PEDIGREES

The extension of (3.1) to include pedigree information has been described in Oakey et al. (2006) for single trials and Oakey et al. (2007), Beeck et al. (2010) and Cullis et al. (2010) for multi-environment trials. These papers fully describe the partitioning of the genetic variance into additive and non-additive parts. This method has been used extensively, see for example Burgueno et al. (2007), de los Crossa et al. (2010), Piepho et al. (2008) to name a few.

In the following we present a brief overview of the models described by Oakey et al. (2006) and Oakey et al. (2007) but extend their models to explicitly account for the partitioning of the vector of (total) genetic effects into two sub-vectors. That is, if we let  $\mathbf{u}_{gd}$  be the vector of (total) genetic direct effects, then we assume that  $\mathbf{u}_{gd} = (\mathbf{u}_{gdp}^T, \mathbf{u}_{gdo}^T)^T$ . The vector  $\mathbf{u}_{gdp}$  is the vector of genetic direct effects of entries in the pedigree but not present in the data-set and as before the vector  $\mathbf{u}_{gdo}$  is the vector of genetic effects for entries in the pedigree and present in the data-set. These vectors are of length  $m_p$  and  $m_o$ , respectively, and  $m = m_p + m_o$ .

We use the genetic model for  $\mathbf{u}_{gd}$  which assumes that

$$\mathbf{u}_{gd} = \mathbf{u}_{ad} + \mathbf{u}_{ed} \tag{3.4}$$

where  $\mathbf{u}_{ad}$  represents the vector of entry additive genetic direct effects and  $\mathbf{u}_{ed}$  represents the vector of residual genetic direct effects. Each of these vectors are partitioned conformably with  $\mathbf{u}_{gd}$  with respect to the present/not present in the current data-set (the third suffix, viz  $p$  for “parent” and  $o$  for “offspring” present in the data-set). Our model including pedigree information is derived by replacing  $\mathbf{u}_{gdo}$  in (3.1) with  $\mathbf{u}_{gd}$  of (3.4) so is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{gd}(\mathbf{u}_{ad} + \mathbf{u}_{ed}) + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e} \tag{3.5}$$

where  $\mathbf{Z}_{gd} = [\mathbf{0} \ \mathbf{Z}_{gdo}]$ .

We assume that each of the vectors of genetic direct effects namely  $\mathbf{u}_{ad}$  and  $\mathbf{u}_{ed}$  are (pair-wise) independent and are Gaussian with zero mean, with variance matrices  $\sigma_{add}\mathbf{A}$ , and  $\sigma_{edd}\mathbf{I}_m$ .

The matrix  $\mathbf{A} = \{a_{ij}\}$  is the relationship matrix and its elements are given by  $a_{ii} = 1 + F_i$  and  $a_{ij} = 2f_{ij}$  where  $F_i$  is the inbreeding coefficient of entry  $i$  and  $f_{ij}$  is the coefficient of parentage between entries  $i$  and  $j$ . The inbreeding coefficient is the percentage of similarity a genotype will have when crossed with itself. The coefficient of parentage is the genetic

distance between two genotypes calculated as the sum of all the coefficients for all common ancestors between the genotypes.

All computations including the matrix  $\mathbf{A}^{-1}$  are computed in the R (Development Core Team 2008) package ASReml-R (Butler et al. 2009). The matrix  $\mathbf{A}^{-1}$  is calculated using the `asreml.Ainverse` function which uses the algorithms of Meuwissen and Luo (1992) and Henderson (1976) with modifications to adjust for selfing. Details are given in an unpublished report (Gilmour, pers comm.).

### 3.3. INCLUDING INFORMATION ON PEDIGREES AND COMPETITION

To allow for interplot competition in the row direction we incorporate the random effects treatment interference model (R-TIM) of Stringer, Cullis, and Thompson (2011). Each entry is assumed to have a direct genetic effect (for each of the components) on the plot into which it was sown and a neighbor effect on the adjacent row-neighbor plots. Hence (3.5) can be extended as follows

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g(\mathbf{u}_a + \mathbf{u}_e) + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e} \quad (3.6)$$

where  $\mathbf{u}_a = (\mathbf{u}_{a_d}^T, \mathbf{u}_{a_n}^T)^T$ , and  $\mathbf{u}_e = (\mathbf{u}_{e_d}^T, \mathbf{u}_{e_n}^T)^T$ , where the subscripts  $d$  and  $n$  represent the direct and neighbor effects, respectively. The associated genetic design matrix is given by  $\mathbf{Z}_g = [\mathbf{Z}_{g_d} \ \mathbf{N}_g \ \mathbf{Z}_{g_d}]$ , where  $\mathbf{N}_g = \mathbf{I}_c \otimes \mathbf{N}_r$  and  $\mathbf{N}_r$  is the within-row first order neighbor incidence matrix.

Stringer, Cullis, and Thompson (2011) proposed two variance models for the R-TIM. In the first, more general model, the variance matrices for the vectors of genetic effects are given by

$$\begin{aligned} \text{var}(\mathbf{u}_a) &= \begin{pmatrix} \sigma_{a_{dd}} & \sigma_{a_{dn}} \\ \sigma_{a_{dn}} & \sigma_{a_{nn}} \end{pmatrix} \otimes \mathbf{A} = \mathbf{G}_a \otimes \mathbf{A}, \\ \text{var}(\mathbf{u}_e) &= \begin{pmatrix} \sigma_{e_{dd}} & \sigma_{e_{dn}} \\ \sigma_{e_{dn}} & \sigma_{e_{nn}} \end{pmatrix} \otimes \mathbf{I}_m = \mathbf{G}_e \otimes \mathbf{I}_m. \end{aligned}$$

The second form for the R-TIM corresponds to the model of Draper and Guttman (1980) in which the neighbor effects are assumed to be a scalar multiple of the direct effects. In terms of our notation this leads to reduced rank forms (with rank 1) for the variance matrices  $\mathbf{G}_a$  and  $\mathbf{G}_e$ . This model can be fitted in ASReml-R using the algorithm described in Thompson et al. (2003).

All models in this paper were fitted using the ASReml-R package (Butler et al. 2009). This provides residual maximum likelihood (REML) estimates of the variance parameters, empirical best linear unbiased estimates (E-BLUES) of the fixed effects and empirical best linear unbiased predictions (E-BLUPs) of the random effects.

It is important to note that the design of this trial did not allow for the genetic relationships and therefore there may be a chance that neighboring plots contain hybrids that are genetically related. In this case the yields may display similarities that are not due to interplot competition but rather to the genetic relationship. This demonstrates that it is vital that the pedigree relationships be allowed for in order to assess competition effects and appropriate pure stand yields.

Table 1. Summary of the models fitted to the PYTM trial. The notation RR() denotes the Draper and Guttman variance model for the terms in brackets, D—direct effects, N—neighbor effects. All models also include a random Block term.

Model	Add	Nonadd	Other	logl	Test	P-value
1	D	D		-453.02		
2	D	D	Row, Column	-430.80		
2a	D	D	Column	-450.83	M2a v M2	0.000
2b	D	D	Row	-434.93	M2b v M2	0.002
3	RR(D,N)	RR(D,N)	Row, Column	-420.48		
3a	RR(D,N)	D	Row, Column	-422.41	M3a v M3	0.049
3b	D	RR(D,N)	Row, Column	-426.67	M3b v M3	0.000

Table 2. REML estimates of variance parameters from three models fitted to PYTM data. Model 1: base-line spatial with pedigree information; Model 2: base-line spatial with pedigree information plus random row and column effects; Model 3: joint spatial and competition with pedigree information. Genetic parameters are above the line and non-genetic below.  $\sigma_{p1}^2$ ,  $\sigma_{p2}^2$  and  $\sigma_{p3}^2$  are the variance components for blocks, columns and rows, respectively.

Variance parameter	Model 1 estimate	Model 2 estimate	Model 3 estimate
$\sigma_{add}$	0.281	0.239	0.137
$\sigma_{ann}$			0.020
$\sigma_{adn}$			-0.052
$\sigma_{edd}$	0.241	0.248	0.120
$\sigma_{enn}$			0.006
$\sigma_{edn}$			-0.027
$\sigma^2$	0.630	0.525	0.449
$\sigma_{p1}^2$	0.312	0.307	0.304
$\sigma_{p2}^2$		0.017	0.016
$\sigma_{p3}^2$		0.092	0.087
$\phi_c$	0.17	-0.01	0.05
$\phi_r$	-0.11	-0.18	0.12

### 4. RESULTS AND DISCUSSION

Table 1 presents the summary of the sequence of models fitted to the PYTM trial. Our analysis commenced by fitting a base-line model following the approaches recommended by Gilmour, Cullis, and Verbyla (1997) and recently modified by Stefanova, Smith, and Cullis (2009). This model included direct (D) effects for both additive and non-additive genetic effects, as well as a Block term to respect the resolvability of the design, and lastly used a separable first order autoregressive variance model for the residuals.

The base-line spatial analysis for the PYTM trial resulted in the estimated variance parameters as given for Model 1 in Table 2. The negative auto-correlation (-0.11) for the row dimension is indicative of the existence of interplot competition (Stringer and Cullis 2002). A standard tool for examining the adequacy of an assumed spatial model is the graph of the sample variogram Gilmour, Cullis, and Verbyla (1997). In order to focus on the effect

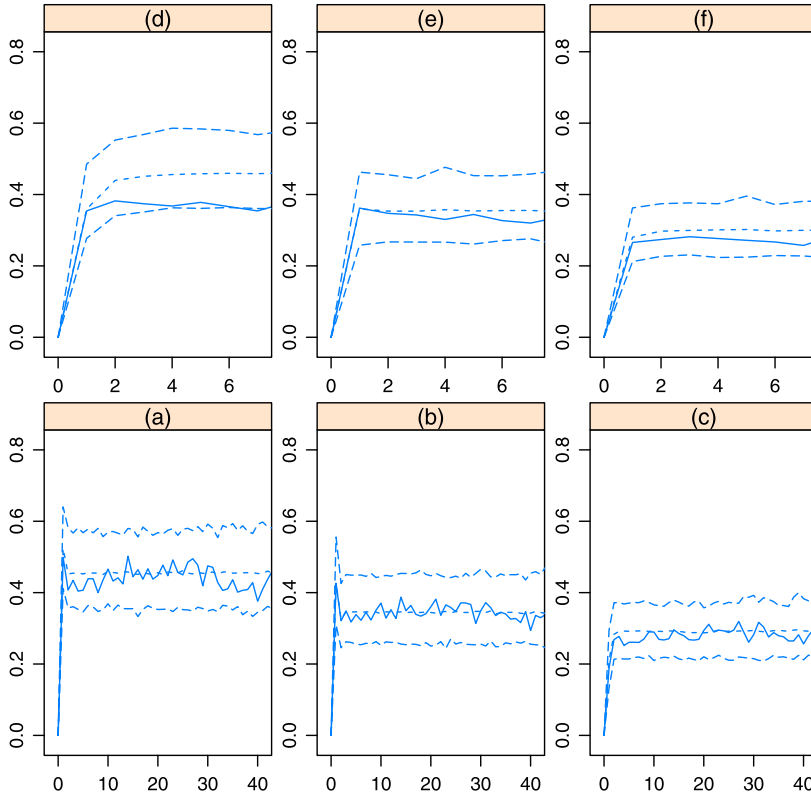


Figure 1. Plots of the row and column faces of the empirical semi-variogram for the residuals (solid line) for the PYTM trial from model 1 (panels (a) and (d)), model 2 (panels (b) and (e)) and model 3 (panels (c) and (f)). These plots are augmented with the mean and 95 % point-wise coverage intervals of the row and column faces of the empirical semi-variogram from a parametric bootstrap sample of size 100.

of competition we restrict our attention to the slice of the variogram corresponding to zero column separation. This is given in Figure 1 for row separations up to 15. In the case of spatial trend (that is, with a positive auto-correlation) this graph should increase smoothly to a plateau. However the large spike at a row separation of one in Figure 1(a) means that adjacent plots (one row apart) have a higher semi-variance than those that are further apart. This suggests that adjacent plots have a negative effect on each other.

Figure 1 (panels (a) and (d)) present the diagnostic plots suggested by Stefanova, Smith, and Cullis (2009). These are the row and column faces of the sample values of the empirical semi-variogram of the residuals from model 1 in Table 1. These plots are augmented with the mean and 95 % point-wise coverage intervals of the faces of the empirical semi-variogram from a parametric bootstrap sample of size 100. This procedure is fully described in Stefanova, Smith, and Cullis (2009), essentially the current model is simulated 100 times using the current variance components and the sample variogram is calculated for each simulation. The 2.5 % and 97.5 % percentiles are obtained and included in Figure 1. There are clear and systematic discrepancies between the mean row and column faces of the parametric bootstrap sample and the residuals from model 1. In both figures



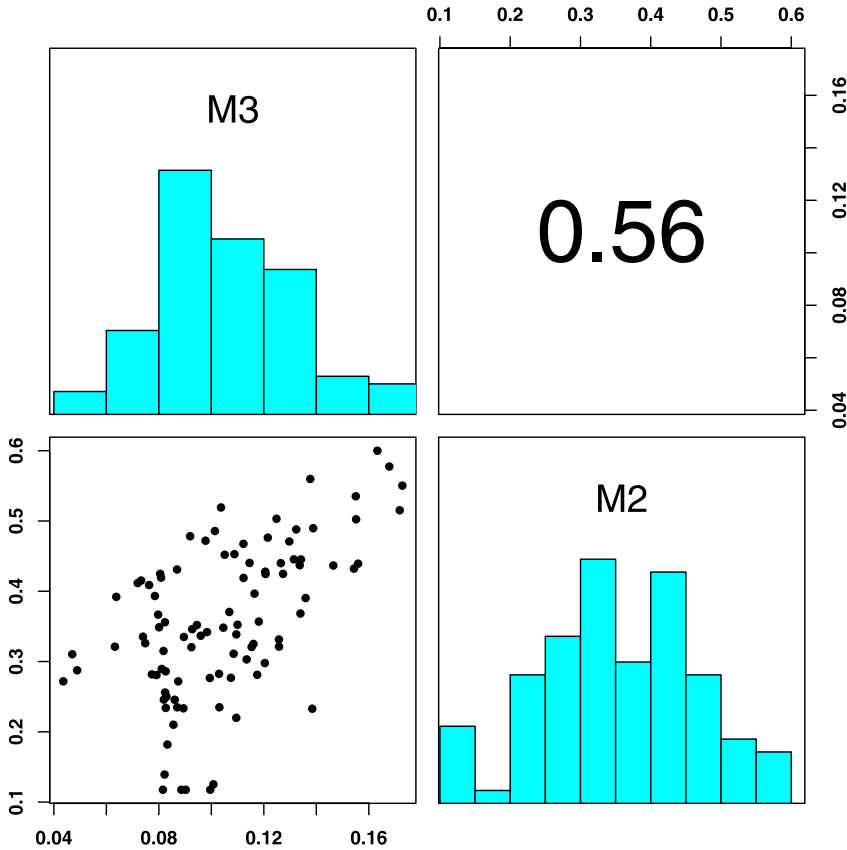


Figure 2. Pair-wise scatter plot (lower left), simple correlation coefficient (upper right) and histograms (diagonals) of the E-BLUPs of the pure stand effects from model 3, and the E-BLUPs of the direct effects from model 2 for the top 10 % of additive effects for the F4 male parents in the PYTM trial.

the mean is generally higher for all lags. This indicates the presence of both row and column effects.

Model 2 investigates this possibility by including random effects for both rows and columns. There is a substantial increase in the residual log-likelihood for model 2 over model 1. Models 2(a) and 2(b) drop the Row and Column terms, respectively, to formally test the need for these terms. Both terms are deemed significant ( $p < 0.05$ ) using Residual Maximum Likelihood Ratio Tests (REMLRTs). Figure 1 (panels (b) and (e)) present the diagnostic plots for the residuals from model 2. There is generally good agreement between the row and column faces of the empirical semi-variogram with the mean of the parametric bootstrap sample.

The noteworthy feature of these plots is the presence of a sharp “spike” at lag one for the row-face of the empirical semi-variogram. The REML estimate of the row auto-correlation parameter for model 2 was  $-0.18$ . This suggests that there is competition present in this direction (i.e. between neighboring plots within the same column, sharing a common long boundary).

Our approach for modeling this (apparent) competition is to fit the reduced rank version of the R-TIM to both terms, denoting this by RR(D,N).

Model 3 provided a substantial improvement in fit over model 2, with both non-additive and additive competition deemed significant (using a REMLRT for models 3a and 3b vs. model 3, respectively).

The diagnostic plots of the empirical semi-variogram of the residuals from model 3 are satisfactory (panels (c) and (f) in Figure 1). Note that the level of these plots has dropped quite appreciably from the previous model (panels (b) and (e)), reflecting the amount of variation explained by the competition effects, this is also reflective in the reduction of  $\sigma^2$  (see Table 2). Also note that the large spike in the row-face of the empirical semi-variogram has been removed. The REML estimate of the row auto-correlation parameter for this model was 0.12, compared with  $-0.18$  for model 2 (Table 2).

Figure 2 presents a plot of the top 10 % of the empirical BLUPs (E-BLUPs) of the pure stand yield ( $\tilde{\mathbf{u}}_{a_d} + 2\tilde{\mathbf{u}}_{a_n}$ ) from model 3 versus the E-BLUPs of the direct effects for model 2 for the F4 male parents. The simple correlation coefficient, displayed in the top right panel shows a correlation of 0.56 for the top 10 % of the E-BLUPs from these two models. This suggests that the selection of male parents from each model is noticeably different, in fact the top 10 % of the E-BLUPs from both models only have 77 % of the male parents in common. Additionally, the E-BLUPs of the pure stand effects are substantially smaller in magnitude than the E-BLUPs of the direct effects. This is due to the negative relationship between the direct and neighbor effects.

[Received March 2012. Accepted October 2012. Published Online November 2012.]

## REFERENCES

- Beeck, C., Cowling, W., Smith, A., and Cullis, B. (2010), "Analysis of Yield and Oil From a Series of Canola Breeding Trials. Part I: Fitting Factor Analytic Models With Pedigree Information," *Genome*, 53, 992–1001.
- Besag, J., and Kempton, R. A. (1986), "Statistical Analysis of Field Experiments Using Neighbouring Plots," *Biometrics*, 42, 231–251.
- Burgueno, J., Crossa, J., Cornelius, P. L., McLaren, G., Trethowan, R., and Krishnamachari, A. (2007), "Modeling Additive  $\times$  Environment and Additive  $\times$  Additive  $\times$  Environment Using Genetic Covariances of Relatives of Wheat Genotypes," *Crop Science*, 47, 311–320.
- Butler, D. G., Cullis, B. R., Gilmour, A. R., and Gogel, B. J. (2009), "ASReml-R Reference Manual Release 3," Technical report, QLD Department of Primary Industries and Fisheries, Brisbane, QLD.
- Crossa, J., de los Campos, G., Perez, P., Gianola, D., Burgueno, J., Araus, J. L., Makumbi, D., Singh, R. P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., and Braun, H.-J. (2010), "Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers," *Genetics*, 186, 713–724.
- Cullis, B., Smith, A., Beeck, C., and Cowling, W. (2010), "Analysis of Yield and Oil From a Series of Canola Breeding Trials. Part II: Exploring VxE Using Factor Analysis," *Genome*, 53, 1002–1016.
- Cullis, B. R., Smith, A. B., and Coombes, N. E. (2006), "On the Design of Early Generation Variety Trials With Correlated Data," *Journal of Agricultural, Biological, and Environmental Statistics*, 11, 381–393.
- Draper, N. R., and Guttman, I. (1980), "Incorporating Overlap Effects From Neighbouring Units Into Response Surface Models," *Applied Statistics*, 29, 128–134.
- Gilmour, A. R., Cullis, B. R., and Verbyla, A. P. (1997), "Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments," *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 269–293.

- Henderson, C. (1976), "A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in the Prediction of Breeding Values," *Biometrics*, 32, 69–83.
- Hunt, C., and Jordan, D. (2009), "Competition Effects in Sorghum Breeding Trials," in *14th Australasian Plant Breeding Conference, SABRAO Journal of Breeding and Genetics*, Cairns, Australia.
- Meuwissen, T. H. E., and Luo, Z. (1992), "Computing Inbreeding Coefficients in Large Populations," *Genetics, Selection and Evolution*, 24, 269–293.
- Oakey, H., Verbyla, A., Cullis, B., Wei, X., and Pitchford, W. (2007), "Joint Modelling of Additive and Non-additive (Genetic Line) Effects in Multi-environment Trials," *Theoretical and Applied Genetics*, 114, 1319–1332.
- Oakey, H., Verbyla, A., Pitchford, W., Cullis, B., and Kuchel, H. (2006), "Joint Modelling of Additive and Non-additive Genetic Line Effects in Single Field Trials," *Theoretical and Applied Genetics*, 113, 809–819.
- Piepho, H. P., Möhring, J., Melchinger, A. E., and Büchse, A. (2008), "BLUP for Phenotypic Selection in Plant Breeding and Variety Testing," *Euphytica*, 161, 209–228.
- R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna. ISBN:3-900051-07-0, <http://www.R-project.org>.
- Smith, A. B., Cullis, B. R., and Thompson, R. (2005), "The Analysis of Crop Cultivar Breeding and Evaluation Trials: An Overview of Current Mixed Model Approaches," *Journal of Agricultural Science, Cambridge*, 143, 449–462.
- Stefanova, K. T., Smith, A. B., and Cullis, B. R. (2009), "Enhanced Diagnostics for the Spatial Analysis of Field Trials," *Journal of Agricultural, Biological, and Environmental Statistics*, 14, 392–410.
- Stringer, J. (2006), "Joint Modelling of Spatial Variability and Interplot Competition to Improve the Efficiency of Plant Improvement," PhD thesis, The University of Qld, Brisbane.
- Stringer, J., Cullis, B., and Thompson, R. (2011), "Joint Modeling of Spatial Variability and Within-Row Interplot Competition to Increase the Efficiency of Plant Improvement," *Journal of Agricultural, Biological, and Environmental Statistics*, 16, 269–281.
- Stringer, J. K., and Cullis, B. R. (2002), "Application of Spatial Analysis Techniques to Adjust for Fertility Trends and Identify Interplot Competition in Early Stage Sugarcane Selection Trials," *Australian Journal of Agricultural Research*, 53, 911–918.
- Thompson, R., Cullis, B. R., Smith, A. B., and Gilmour, A. R. (2003), "A Sparse Implementation of the Average Information Algorithm for Factor Analytic and Reduced Rank Variance Models," *Australian and New Zealand Journal of Statistics*, 45, 445–460.