2012

Multi-phase variety trials using both composite and individual replicate samples: A model-based design approach

Alison B. Smith
University of Wollongong, alismith@uow.edu.au

David G. Butler
Department of Agriculture

Colin R. Cavanagh
CSIRO

Brian R. Cullis
University of Wollongong, bcullis@uow.edu.au

Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

16-12

Multi-phase variety trials using both composite and individual replicate samples: A model-based design approach

Alison B. Smith, David G. Butler, Colin R. Cavanagh and Brian Cullis

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Multi-phase variety trials using both composite and individual replicate samples: A model-based design approach

Alison B. Smith

School of Mathematics and Applied Statistics, University Of Wollongong, Wollongong, NSW, Australia.

David G. Butler

Department of Agriculture, Fisheries & Forestry, Toowoomba, Qld, Australia.

Colin R. Cavanagh

CSIRO Plant Industry and Food Futures Flagship, Canberra, ACT, Australia.

Brian R. Cullis

School of Mathematics and Applied Statistics, University Of Wollongong, Wollongong, NSW, Australia and CSIRO, Canberra, ACT, Australia.

Summary. This paper provides an approach for the design and analysis of variety trials that are used to obtain quality trait data. These trials are multi-phase in nature, comprising a field phase followed by one or more laboratory phases. Typically the laboratory phases are costly relative to the field phase and this necessitates a limit on the number of samples that can be tested. Historically, this has been achieved by sacrificing field replication, either by testing a single replicate plot for each variety or a single composite sample, obtained by combining material from several field replicates. An efficient statistical analysis can not be applied to such data so that valid inference and accurate prediction of genetic effects is precluded. In this paper we propose an approach in which some varieties are tested using individual field replicate samples and others as composite samples. Replication in the laboratory is achieved by splitting a relatively small number of field samples into sub-samples for separate processing. We show that, if necessary, some of the composite samples may be split for this purpose. We also show that, given a choice of field compositing and laboratory replication strategy, an efficient design for a laboratory phase may be obtained using model-based techniques. The methods are illustrated using two examples.

Keywords: composite samples; mixed models; multi-phase trials; model-based design

1. Introduction

Accurate phenotypic information on quality traits is vital for successful variety selection in plant breeding programs and for genetic research including genomic

selection and the identification of quantitative trait loci. Many quality traits, such as milling yield, dough rheology and bread baking characteristics for wheat, are obtained from multi-phase experiments in which varieties are first grown in a field trial then further processed in a laboratory. Quality testing tends to be labour intensive and expensive so there is typically a limit to the total number of samples that can be tested. Historically this has led to the practice of testing a single field replicate for each variety (or a single composite sample formed by combining grain from the individual replicate plots for that variety) and no randomisation or replication of grain samples in the laboratory. Such an approach precludes efficient prediction of the genetic effects of interest.

Recent work, in particular Smith et al. (2006) and Brien et al. (2011) has shown the importance of using sound experimental design techniques (including randomisation and replication) in all phases of a multi-phase experiment. The testing of replicates from the field phase is vital since without this there is no valid estimate of error. Replication in a laboratory phase involves the splitting of the experimental units from the previous phase and testing as separate samples. Both Smith et al. (2006) and Brien et al. (2011) recommend laboratory replication although the latter suggested that it is only required when " ... uncontrolled variation in the laboratory is large relative to the first phase" or when "... the relative magnitudes of field and laboratory variation are to be assessed." In our experience with quality traits, one or both of these is generally the case.

The cost of quality testing and consequent restriction on number of samples was addressed in Smith et al. (2006) with the use of partial replication. In this approach and for a two-phase experiment, a subset of the varieties is tested using multiple field replicate samples (the remainder being tested using a single replicate only), then a subset of the selected field plots are split to produce replicate samples for the laboratory process. This has been found to work well in terms of minimising both selection errors and cost, but when the field trial is a fully replicated trial it can be wasteful in the sense that field plots of some varieties are completely ignored. Smith et al. (2011) considered a similar issue but in the context of grain quality traits that are derived from a single (field) phase alone. They proposed that some varieties be tested using individual replicate samples and others using composite samples. In this way all field plots are used to generate data on the trait/s of interest. Smith et al. (2011) demonstrated that with such data it is possible to fit mixed models that enable the efficient prediction of genetic effects. In the current paper we extend the approach of Smith et al. (2011) to suit the first phase in a multi-phase experiment. The concepts will be demonstrated using two motivating examples.

1.1. Motivating examples

1.1.1. Example 1: Wheat variety classification project

Prior to their commercial release, wheat varieties in Australia are classified according to their end-use capabilities. Classification is based on data submitted by private breeding companies to an expert advisory panel. Data on each candidate variety are obtained from a number of field trials and on a range of (multi-phase) traits including flour yield and dough and baking characteristics. Accurate classification of varieties is crucial since growers are paid differentially on this basis. Accuracy is heavily dependent on the use of appropriate data, both in terms of the type and nature, so that protocols regarding experimental design and data requirements are fundamentally important. Currently the data used for classification are based on fully composited data (that is, composites of all field replicates within a trial) from several designed field trials with no experimental designs for the laboratory phases. Such data do not allow a statistical examination of protocols since potential sources of variation (including variety by trial interaction, between plot variation in the field and between sample variation in the laboratory) are confounded so cannot be quantified. A recent project has been designed to enable estimation of all these sources so will ultimately allow examination of protocols for wheat variety classification. The full project spans 3 years with 24 field trials grown across Australia each year. In this paper we consider the experimental design for the measurement of flour yield for one of these field trials.

The field trial under study comprised 54 plots arranged in a rectangular array of 6 columns by 9 rows. There were 3 replicates each of 18 varieties with replicate blocks aligned with pairs of columns. After this trial is harvested grain samples will be taken from each plot, and provided the trial meets certain protein specifications, all 18 varieties will be milled to obtain flour yield. We will describe methods for designing the milling phase of this trial, noting that budgetary constraints have necessitated a restriction of 40 samples for milling.

1.1.2. Example 2: Genomic selection in wheat population

A wheat population with diverse genetic composition has been constructed in order to investigate marker-trait association and genomic selection for a range of complex traits. The population has been genotyped using SNIP and DArT markers and is being phenotyped through a number of field trials. The field trial under study comprised 1000 plots arranged in a rectangular array of 50 columns by 20 rows. There were 773 entries grown in the trial and these comprised 760 test lines and 13 commercial varieties. A resolvable p-replicate design (Cullis et al., 2006) was used in which 554 entries were sown as single replicates and 213 as two replicates. There were 6 entries that had additional replication (see Table 1). Replicate blocks were aligned with columns with the first replicate comprising columns 1-25 and the second columns 26-50. The trait considered here is flour yield. Unlike the field trial

Table 1. Example 2: Summary of replication in the field for all entries and subset of entries chosen for milling.

	All entries		Entries for milling	
Plots/entry	Entries	Plots	Entries	Plots
1	554	554	330	330
2	213	426	145	290
3	4	12	3	9
4	2	8	2	8
Total	773	1000	480	637

described in section 1.1.1 in which it is planned that all varieties will be milled, cost considerations necessitated a limit of less than 550 samples for milling so that not all entries could be milled. It was decided to use 480 entries. These were chosen from the full set of 773 both on the basis of their genetic diversity as identified using the markers and the fact that they had sufficient grain for milling. The latter was important for this trial which was unexpectedly low yielding. A summary of the field replication status of the chosen entries is given as the final columns in Table 1.

2. Design and analysis for multi-phase trials

2.1. Review of methods

In terms of the design of multi-phase trials Smith et al. (2006) and Brien et al. (2011) demonstrate the need for the application of valid experimental design techniques (in particular, replication and randomisation) in all phases. Experimental designs for field trials are well established and widely adopted. Thus in a multi-phase setting the design of the first phase is usually well constructed. The principles employed for this phase should also be applied to the second (and higher) phases. In terms of replication this means that for our two-phase milling examples there must be replication carried through from the field trial into the milling process, then replication applied in the milling process itself. The latter is achieved by taking grain samples from individual field plots and splitting into sub-samples (typically two) then milled separately.

In the context of quality testing fully replicated multi-phase designs are usually prohibitively expensive and not necessary from a statistical perspective. Typically there are restrictions on the total number of samples that can be tested in the laboratory. In the context of two-phase experiments Smith et al. (2006) achieve this using "p/q replicate" designs. In these designs individual field replicates are used for a proportion, p, of varieties with the remainder tested using a single replicate only. This defines the field plots to be tested, a proportion, q of which is then replicated in the laboratory. The approach is easily generalised for experiments

requiring more than two phases.

In terms of the analysis of multi-phase trials Smith et al. (2006) use a linear mixed model approach that accommodates the block structure for each phase as well as allowing for additional sources of variation and correlation. We let k denote the number of phases in the trial and let s denote the number of samples for which a measurement is obtained. The linear mixed model for the $s \times 1$ vector of data $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \boldsymbol{Z_g u_g} + \sum_{r=1}^{k} \boldsymbol{X}_r \boldsymbol{\tau}_r + \sum_{r=1}^{k} \boldsymbol{Z_{p}}_r \boldsymbol{u_{p}}_r + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{u_g}$ is the vector of random variety effects, $\boldsymbol{\tau}_r$ is the vector of fixed effects associated with phase $r (r = 1 \dots k)$, $\boldsymbol{u_{p}}_r$ is the vector of random non-genetic (peripheral) effects associated with phase r and $\boldsymbol{e}$ is the vector of residuals. The matrices $\boldsymbol{Z_g}$, $\boldsymbol{X}_r$ and $\boldsymbol{Z_{p}}_r$ are design matrices. Typically the vectors of random peripheral effects contain sub-vectors that will be denoted by $\boldsymbol{u_{p}}_{rs}$. Then the variance matrices for the random effects are given by

$$\begin{aligned} \text{var}\left(\boldsymbol{u_g}\right) &= \boldsymbol{G_g} \\ \text{var}\left(\boldsymbol{u_{p}}_{rs}\right) &= \boldsymbol{G_{p}}_{rs} \\ \text{var}\left(\boldsymbol{e}\right) &= \boldsymbol{R} \end{aligned}$$

In the simplest models all variance matrices are scaled identities with $\boldsymbol{G_g} = \sigma^2_g \boldsymbol{I}$, $\boldsymbol{G_{p}}_{rs} = \sigma^2_{p_{rs}} \boldsymbol{I}$ and $\boldsymbol{R} = \sigma^2 \boldsymbol{I}$ where the identity matrices have dimensions commensurate with the length of the associated vector of effects. However more complex forms, including separable auto-regressive processes of order 1 for the modelling of spatial correlation in the field plot effects (see Smith et al., 2006) can be used. All analyses in this paper were conducted using the mixed model software `ASReml-R` (Butler et al., 2009) within the R system for statistical computing (R Development Core Team, 2011). `ASReml-R` (Butler et al., 2009) provides residual maximum likelihood (REML) estimates of the variance parameters, empirical best linear unbiased estimates of the fixed effects and empirical best linear unbiased predictions of the random effects.

In terms of experimental design we first note that multi-phase experiments are typically designed sequentially rather than simultaneously with the second (and higher) phases often constructed after the field trial has been harvested. This may be neccessary for several reasons including the fact that in standard selection trials only those varieties that are selected on the basis of grain yield are then tested for quality traits. Thus the varieties to be tested are only known after conclusion of the field trial. Similarly the choice of plots to be replicated in the laboratory may be dependant on the field trial since quality testing requires a minimum amount of grain and there may be some plots with insufficient material to split into subsamples for separate processing.

Given a valid design for the first phase we seek a second phase design using the model-based techniques of Butler et al. (2013b). In this approach the criterion used is A-optimality so that the goal is to seek designs that minimise the average pairwise (prediction error) variance of the variety effects (or some subset there-of) given a pre-specified model of the form in equation (1). Bueno Filho and Gilmour (2007) discuss the fact that the use of this criterion minimises the probability of making incorrect selection decisions and also provides, in the random treatment effects setting, "... a sensible utility function for ranking treatments and for estimating treatment effects". All designs in this paper were generated using the package `od` (Butler et al., 2013a) which runs within the R system for statistical computing (R Development Core Team, 2011). The syntax of `od` (Butler et al., 2013a) is consistent with `ASReml-R` (Butler et al., 2009). The package produces designs given a specified model (and associated variance parameter values) and starting design.

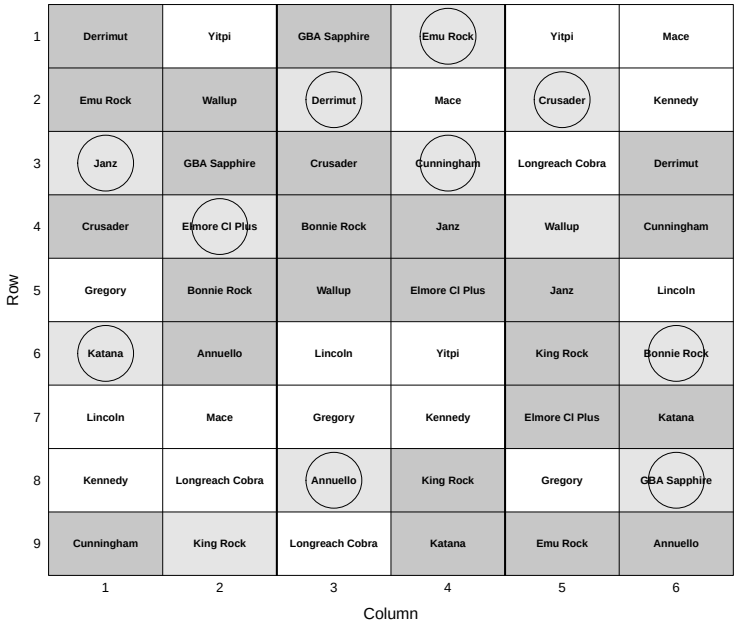
To illustrate these concepts we consider the first motivating example.

2.1.1. Example 1: Wheat variety classification project

A fully replicated design for this experiment (assuming two laboratory replicates) would require 108 samples to be milled (that is, 18 varieties $\times$ 3 field replicates $\times$ 2 laboratory replicates). It was noted in section 1.1.1 that cost considerations for the project necessitated a limit of 40 samples for milling. Using the p/q replicate ideas of Smith et al. (2006) one scheme that will achieve this uses values of p and q both equal to one third. Thus 6 out of the 18 varieties will be milled using all 3 field replicates (a total of 18 plots) and the remainder using a single replicate each (a total of 12 plots). Note that for the second group there needs to be a choice of which replicate plots to test and which to ignore. Then of the 30 field plots to be tested, 10 will be replicated in the laboratory. With these values of p and q, one possible selection of varieties and plots to be replicated is shown in Figure 1. The plots to be replicated in the milling process have been chosen to provide reasonable spatial coverage across the field. Additionally they were chosen from plots sown with the 12 varieties that are only being milled with single field replicates. In this way an attempt is made to balance the total number of samples for each variety (also see section 2.1.2). In our example this results in 6 varieties with 3 samples each, 10 with 2 and 2 with a single sample. Note that we may not always have the luxury of choosing the plots to be replicated (see section 2.2.2).

In terms of the laboratory design we first note that samples will be milled as 8 samples per day for each of five days. There is often a break in the middle of each day making a total of 10 milling sessions (half days). Experience has shown that there are often substantial effects associated with milling sessions. This is a natural blocking factor that should be accommodated in the experimental design (also see Brien et al. (2011)). We choose to enforce resolvability in the sense that samples to be replicated in the laboratory are positioned with one replicate in sessions 1-5

Fig. 1. Field layout for example 1 showing plots to be milled in p/q replicate design. Field trial comprises 18 varieties and 3 replicates (columns 1,2; columns 3,4 and columns 5,6). Plots coloured light grey and white are to be milled as individual replicates (varieties in light grey plots have a single replicate only; varieties in white plots have all 3 replicates). Plots coloured dark grey will not be milled. Plots to be replicated in the milling phase are circled.



(milling replicate 1) and the other in sessions 6-10 (milling replicate 2). This will be examined in more detail in section 2.1.2.

In terms of the analysis the base-line mixed model for the data is as in equation (1) with $k = 2$ and $s = 40$. The only fixed effects are τ_2 which is simply an overall mean (τ_1 is omitted). The peripheral effects for the first phase comprise $\boldsymbol{u}_{\boldsymbol{p}_{11}}$ which represents the 3×1 vector of field replicate effects and $\boldsymbol{u}_{\boldsymbol{p}_{12}}$ which represents the 54×1 vector of field plot effects. Note that the corresponding 40×54 design matrix $\boldsymbol{Z}_{\boldsymbol{p}_{12}}$ will contain zero columns for those field plots that are not milled. The peripheral effects for the second phase comprise $\boldsymbol{u}_{\boldsymbol{p}_{21}}$ which represents the 2×1 vector of milling replicate effects and $\boldsymbol{u}_{\boldsymbol{p}_{22}}$ which represents the 10×1 vector of milling session effects. In the base-line mixed model all variance matrices have the simple variance component form as described in section 2.1.

The second phase design can be constructed in `od` (Butler et al., 2013a) using a model that is the same as that described for the analysis except with the addition

of random row and column effects for the first phase (denoted $\boldsymbol{u}_{\boldsymbol{p}_{13}}$ and $\boldsymbol{u}_{\boldsymbol{p}_{14}}$ respectively). These were added to the design model as a precautionary measure. We choose not to confound field and milling replicates so in the starting design allocate approximately half of the samples from each field replicate to each milling replicate (also see section 2.1.2). This then allows estimation of both the field and milling replicate sources of variation. An example (optimised) design based on values of the variance parameters $\sigma^2_g = 1.0$, $\sigma^2_{p_{11}} = 0.1$, $\sigma^2_{p_{12}} = 0.2$, $\sigma^2_{p_{13}} = 0.1$, $\sigma^2_{p_{14}} = 0.1$, $\sigma^2_{p_{21}} = 0.3, \sigma^2_{p_{22}} = 0.2$ and $\sigma^2 = 1.0$ is shown in Figure 2. Note that we assume without loss of generality that the residual variance has a value of 1.0 so that all other variance values can be regarded as ratios relative to residual variance. The values were chosen on the basis of experience from analysing numerous milling trials.

Fig. 2. Milling layout for example 1 using p/q replicate design. Laboratory phase comprises 40 samples milled as 4 per session with 2 sessions per day (morning session = orders 1-4; afternoon = 5-8) and for 5 days. Samples are labelled according to their variety and field replicate number. A total of 30 field samples was milled and 10 of these were replicated in the milling process (samples coloured grey). Milling replicates were aligned with sessions (replicate 1 = days 1 and 2 and morning of day 3; replicate 2 = afternoon of day 3 and days 4 and 5).

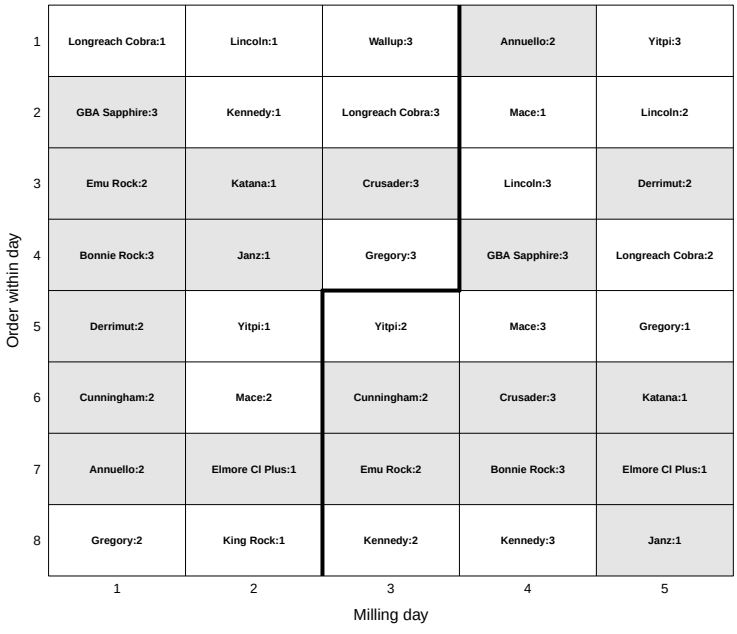


Table 2. Reduced example 1: data for field trial comprising 8 plots and 6 varieties.

Fplot	Frep	Variety
1	1	Derrimut
2	1	Emu Rock
3	1	Yitpi
4	1	Wallup
5	2	GBA Sapphire
6	2	Derrimut
7	2	Emu Rock
8	2	Mace

2.1.2. Reduced example 1: 8 field plots

In order to more clearly illustrate the principles of resolvability and orthogonality of field and milling replicates alluded to in section 2.1.1 we consider a very small subset of the field trial for that example. If we consider the 8 plots in the top left hand corner of Figure 1, namely rows 1 and 2 and columns 1-4 we see that this constitutes a valid resolvable $p-$replicate design (Cullis et al., 2006) for 6 varieties. The varieties Derrimut and Emu Rock are grown in 2 plots each (once in each replicate) and the varieties Wallup, Yitpi, GBA Sapphire and Mace in single plots. The data for this reduced field phase is shown in Table 2. Note that due to the small size of the trial we have ignored the two-dimensional (row by column) layout and indexed the plots simply as 1-8. In our example we assume that all 8 field plots will be milled with 4 of these being replicated, making a total of 12 samples to be milled (hence-forth called "milling samples"). In terms of the choice of field plots to be replicated we aim to equalise the (total) number of milling samples for each variety. Thus we replicate those varieties that did not have field replicates (varieties Yitpi, Wallup, GBA Sapphire and Mace). This provides 2 observations for each variety.

We assume that the 12 samples will be milled as 3 per day for each of 4 days. Days are natural blocking factors and we enforce resolvability so that the field samples to be replicated (field plots 3, 4, 5 and 8) will have one sub-sample milled in days 1 or 2 (milling replicate 1) and the other in days 3 or 4 (milling replicate 2). Finally we choose not to confound field and milling replicates so allocate the remaining field samples (field plots 1, 2, 6 and 7) so that for each field replicate the samples are balanced across both milling replicates. We therefore form a starting design that encompasses the two aspects of resolvability and orthogonality of field and milling replicates. One possibility is as given in Table 3 as the first 6 columns. Finally we use `od` (Butler et al., 2013a) to determine an optimum design. In each iteration of the design search, two rows of the data-frame are swapped subject to the constraint that swaps may only occur within milling replicates. If the swap

Table 3. Reduced example 1: data-frames (starting and optimised design) for milling trial comprising 12 samples milled as 3 per day for each of 4 days.

			Starting design			Optimised design		
Mrep	Mday	Mord	Fplot	Frep	Variety	Fplot	Frep	Variety
1	1	1	3	1	Yitpi	8	2	Mace
1	1	2	4	1	Wallup	7	2	Emu Rock
1	1	3	5	2	GBA Sapphire	5	2	GBA Sapphire
1	2	1	8	2	Mace	3	1	Yitpi
1	2	2	1	1	Derrimut	1	1	Derrimut
1	2	3	7	2	Emu Rock	4	1	Wallup
2	3	1	3	1	Yitpi	3	1	Yitpi
2	3	2	4	1	Wallup	4	1	Wallup
2	3	3	5	2	GBA Sapphire	2	1	Emu Rock
2	4	1	8	2	Mace	8	2	Mace
2	4	2	6	2	Derrimut	6	2	Derrimut
2	4	3	2	1	Emu Rock	5	2	GBA Sapphire

results in a reduction in the A-value the resultant data-frame is maintained for the next iteration. The optimum design given a mixed model with random variety effects, field replicate and plot effects and milling replicate and day effects (with variance component ratios of 1, 0.1, 0.2, 0.3 and 0.2 respectively) is shown as the final columns of Table 3. Note that the A-value for the starting design was 0.856 whereas for the optimum design it was 0.845.

2.2. Compositing strategies

In Section 2.1 it was shown how the p/q replicate ideas of Smith et al. (2006) provided an approach for limiting the cost of multi-phase testing whilst still enabling a valid statistical analysis. However there was substantial "waste" in the sense that for the varieties to be tested only a subset of the field plots was used with many plots ignored altogether. In the first example where all varieties grown in the field were subsequently milled only 30 out of a total of 54 plots were used in the milling process.

In the context of costly traits measured from single phase (field) trials, Smith et al. (2011) suggested the use of individual replicates for a proportion of varieties and composite samples for the remainder. For the latter a single sample is used for each variety but it represents a composite sample from all replicate plots for that variety rather than just a sample from a single replicate.

Here we extend the compositing ideas of Smith et al. (2011) for the multi-phase setting. For simplicity we focus on two-phase experiments and consider compositing for the first phase only. The model-based approaches for design and analysis

described in section 2.1 also apply here. The linear mixed model can be written as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{Z_g u_g} + \boldsymbol{X}_1\boldsymbol{\tau}_1 + \boldsymbol{Z_{p_1} u_{p_1}} + \boldsymbol{X}_2\boldsymbol{\tau}_2 + \boldsymbol{Z_{p_2} u_{p_2}} + \boldsymbol{e} \\ &= \boldsymbol{Z_g u_g} + \boldsymbol{C}_1\boldsymbol{X^*}_1\boldsymbol{\tau}_1 + \boldsymbol{C}_1\boldsymbol{Z^*_{p_1}}\boldsymbol{u_{p_1}} + \boldsymbol{X}_2\boldsymbol{\tau}_2 + \boldsymbol{Z_{p_2} u_{p_2}} + \boldsymbol{e} \end{aligned} \quad (2)$$

where all terms are as previously defined for equation (1). The difference now is that the design matrices for the first phase are non-standard with $\boldsymbol{X}_1 = \boldsymbol{C}_1\boldsymbol{X^*}_1$ and $\boldsymbol{Z_{p_1}} = \boldsymbol{C}_1\boldsymbol{Z^*_{p_1}}$ where $\boldsymbol{C}_1$ is an $s \times n$ averaging matrix that reflects the compositing of samples from the first phase (and n is the number of plots in the field trial). The concepts will be illustrated in the context of the two motivating examples.

2.2.1. Example 1: Wheat variety classification project

In section 2.1.1 a milling design was constructed in which 30 field samples were tested and 10 replicated in the laboratory. In that setting each field sample corresponded to a single field plot. We now consider a design with the same number of field samples but with some of these corresponding to composites of several plots. Replication in the laboratory will remain at 10 samples.

Smith et al. (2011) describe in detail compositing strategies for field trials that have two replicates and briefly allude to designs with more replicates. They suggest that in the latter there are numerous possibilities for compositing strategies. In terms of a three replicate design (as is the case in this example) Smith et al. (2011) suggest that the simplest approach would be to either test three individual plot samples for a variety or a single sample that is a composite of all three plots for the variety. In this strategy, each variety has either one or three field samples tested. An alternative strategy involves three possibile types of testing for a variety, namely one field sample (a composite of all three replicates), two field samples (a composite of two replicates and a separate sample for the remaining replicate) and three samples (individual replicate samples).

In the context of our two-phase example these provide strategies for replication from the first phase. Given that 30 field samples will be milled two possible compositing strategies are:

Strategy A: (one or three field samples for each variety)

- **C3:** 12 varieties with a single field sample (composite of all three replicates)
- **IR:** 6 varieties with three field samples (individual replicate samples)

Strategy B: (one, two or three field samples for each variety)

- **C3:** 10 varieties with a single field sample (composite of all three replicates)
- **C2:** 4 varieties with two field samples (composite of two replicates and a separate sample for remaining replicate)

Table 4. Example 1: replication groups for compositing strategies A and B. C3: variety is milled using single field sample (composite of 3 replicates); C2: variety is milled using two field samples (composite of 2 replicates and remaining replicate milled separately); IR: variety is milled using three field samples (individual replicates). Field samples that are replicated in the milling process correspond to varieties that are IR in both strategies (either all three or two field samples for the variety are replicated).

	Strategy		Field samples	
Group	A	B	rep'd in lab	Varieties
C3:C3	C3	C3	0	Annuello, Bonnie Rock, Crusader, Cunningham, Elmore Cl Plus, Emu Rock, GBA Sapphire, Janz, King Rock, Wallup
C3:C2	C3	C2	0	Derrimut, Katana
IR:C2	IR	C2	0	Longreach Cobra, Mace
IR:IR:2	IR	IR	2	Lincoln, Yitpi
IR:IR:3	IR	IR	3	Gregory, Kennedy

IR: 4 varieties with three field samples (individual replicate samples)

At this stage we will restrict laboratory replication to those field samples that correspond to individual plot samples. Alternatives will be explored in sections 2.2.2 and 2.2.3. In order to facilitate a statistical comparison of strategies A and B we have allocated varieties to replication groups in as similar a manner as possible. The details are given in Table 4. Note that the same 10 plots are replicated in the laboratory for both strategies. These correspond to all 3 replicates of the varieties Gregory and Kennedy and 2 out of the 3 replicates of the varieties Lincoln and Yitpi.

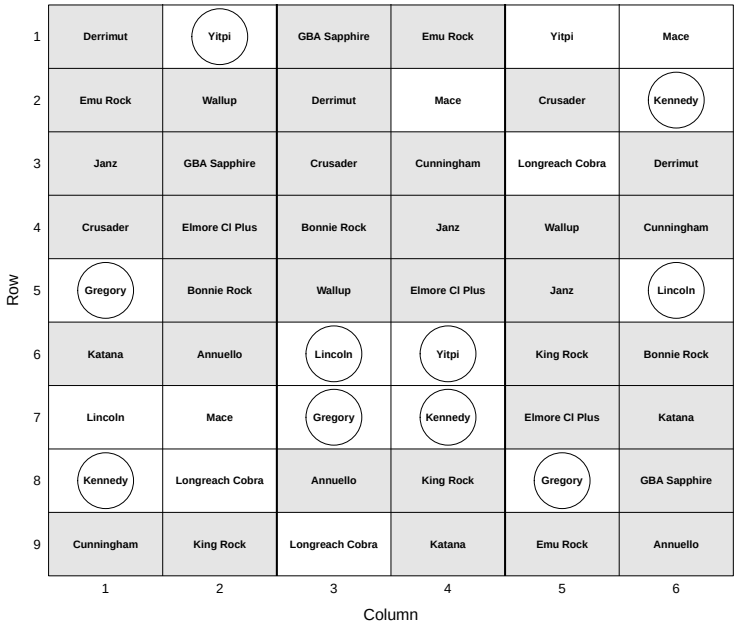
Possible selections of varietes and plots to be replicated and/or composited are shown in Figures 3 and 4 for strategies A and B respectively.

In terms of the laboratory design we use the same blocking factors as in section 2.1.1. The base-line mixed model for analysis is as in equation (2) with the same fixed and random effects as for the analysis model in section 2.1.1. The design matrices for the first phase involve $\boldsymbol{C}_1$ which is a 40×54 matrix that reflects the compositing strategy. For strategy A, the i^{th} row of $\boldsymbol{C}_1$ (which corresponds to the i^{th} sample) has elements given by:

- 1 in column j if this sample corresponds to field plot j alone
- $1/3$ in columns j, k and l if this sample corresponds to a composite of field plots j, k and l
- 0 otherwise

In the base-line mixed model all variance matrices have the simple variance component form as described in section 2.1.

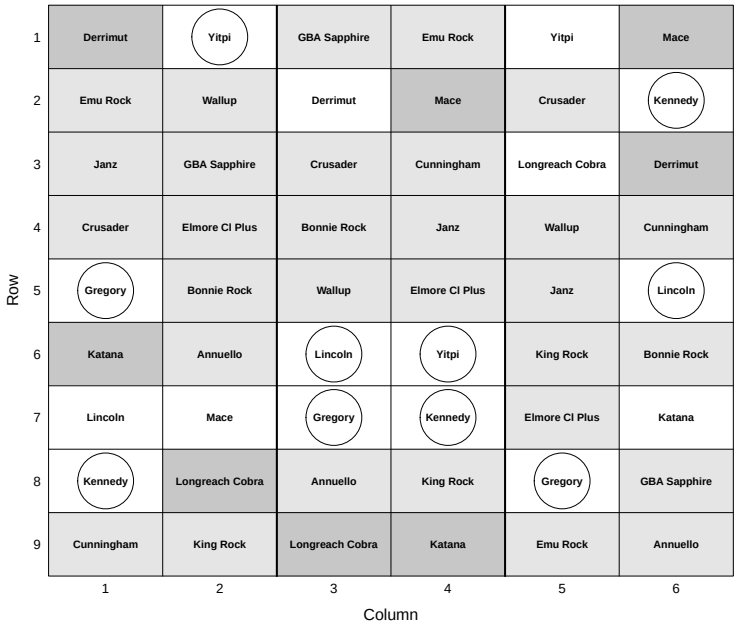
Fig. 3. Field layout for example 1 showing compositing strategy A. Field trial comprises 18 varieties and 3 replicates (columns 1,2; columns 3,4 and columns 5,6). Plots coloured grey are to be milled as composites of all three replicates of a variety; white plots are to be milled as individual replicates. Plots to be replicated in the milling phase are circled.



The second phase design can be constructed using the same approach as in section 2.1.1. Example randomisations for strategies A and B using the design model and variance parameters as in section 2.1.1 are shown in Figures 5 and 6. The A-values for these two designs are 0.919 and 0.9024 for strategies A and B respectively.

In the designs shown in Figures 5 and 6 there was a superiority of strategy B over A with a difference of 0.0166 in the A-values. Since compositing applies at the field plot level it was of interest to establish whether this superiority would be maintained over a range of values for the plot variance component ratio. Therefore, additional designs were generated with plot variance component ratios of 0.5 and 1.0. The associated A-values were lower for strategy B than A in all cases (see Table 5). The superiority of strategy B over A can be explained in terms of the effective replication for individual varieties. Since there is little variation in effective replication between varieties within a replication group (as defined in Table 4) the results are presented on a group basis (see Table 5). The only differences between

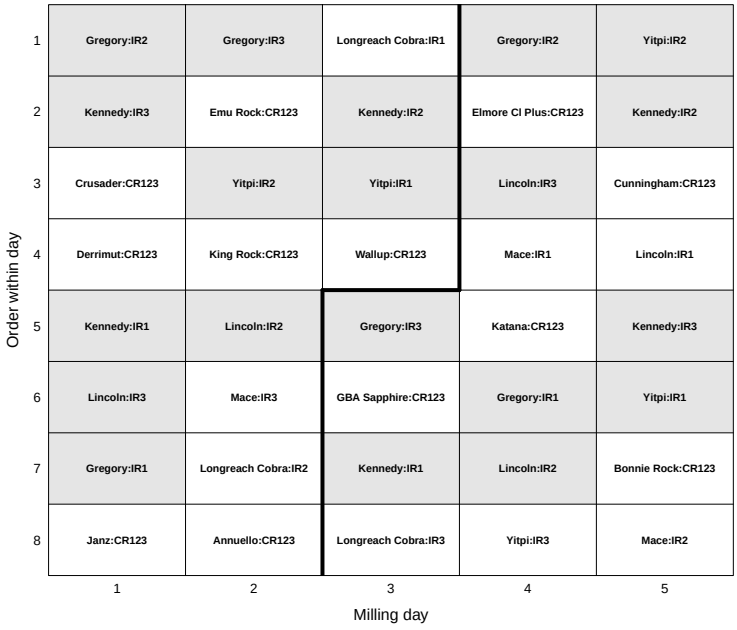
Fig. 4. Field layout for example 1 showing compositing strategy B. Field trial comprises 18 varieties and 3 replicates (columns 1,2; columns 3,4 and columns 5,6). Plots coloured light grey are to be milled as composites of all three replicates of a variety; plots coloured dark grey are to be milled as composites of two replicates of a variety; white plots are to be milled as individual replicates. Plots to be replicated in the milling phase are circled.



the strategies are in terms of the four varieties that were milled as C2 in strategy B (two of which were milled as C3 in strategy A and two as IR). The key finding was that the gain in effective replication for C2 milling compared with C3 milling was always greater than the loss for C2 milling compared with IR. This is an important result in terms of making recommendations about compositing strategies for field trials with more than two replicates.

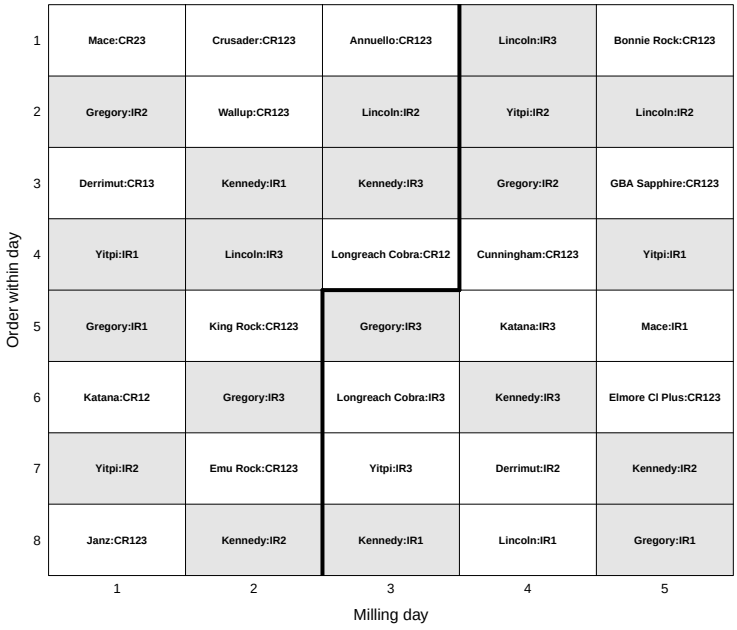
The superiority of strategy B over A was measured in terms of A-values from the design generation stage so reflect the average pairwise prediction error variance of varieties assuming known variance parameters. In order to confirm these findings in the context of data analysis (thence estimation of variance parameters) a simulation study was conducted. The data were generated on the basis of the designs in Figures 5 and 6. The model used for data generation was the same as that used for design construction, namely with random effects for varieties (with $\sigma^2_g = 1.0$), field replicates ($\sigma^2_{p_{11}} = 0.1$), field plots ($\sigma^2_{p_{12}} = 0.2$), field rows ($\sigma^2_{p_{13}} = 0.1$), field

Fig. 5. Milling layout for example 1 using compositing of some field samples according to strategy A. Laboratory phase comprises 40 samples milled as 4 per session with 2 sessions per day (morning session = orders 1-4; afternoon = 5-8) and for 5 days. Samples are labelled according to their variety and field replicate status. A total of 30 field samples was milled and 10 of these were replicated in the milling process (samples coloured grey). Milling replicates were aligned with sessions (replicate 1 = days 1 and 2 and morning of day 3; replicate 2 = afternoon of day 3 and days 4 and 5).



columns ($\sigma^2_{p_{14}} = 0.1$), milling replicates ($\sigma^2_{p_{21}} = 0.3$), milling sessions ($\sigma^2_{p_{22}} = 0.2$) and residuals ($\sigma^2 = 1.0$). In each simulation only one set of effects was generated for each term in the model and used for both strategies. The model fitted to the data matched that used for design generation. The results are presented in Table 6 as means over 700 simulations (where the fitted models converged for both strategies). Note that Table 6 contains an additional strategy, namely strategy "C" which will be discussed in Section 2.2.3. The estimated variance parameters in Table 6 show some bias for both methods in terms of the variance for the field replicate and row effects. The latter was to be expected given the small amount of information (small number of columns) for estimation of this component. The main finding, however, is that both strategies show similar performance. It is also important to note that the mean across simulations of the average pairwise prediction error variance for

Fig. 6. Milling layout for example 1 using compositing of some field samples according to strategy B. Laboratory phase comprises 40 samples milled as 4 per session with 2 sessions per day (morning session = orders 1-4; afternoon = 5-8) and for 5 days. Samples are labelled according to their variety and field replicate status. A total of 30 field samples was milled and 10 of these were replicated in the milling process (samples coloured grey). Milling replicates were aligned with sessions (replicate 1 = days 1 and 2 and morning of day 3; replicate 2 = afternoon of day 3 and days 4 and 5).



varieties was lower for strategy B than A (difference of 0.0318) thereby confirming the A-values from the design generation stage.

Finally we note that both designs are quite unbalanced in terms of the total numbers of samples for each variety. In strategy B there are 10 varieties (C3) with a single sample each, 4 varieties (C2) with 2 samples each, 2 varieties (IR:2) with 5 samples each and 2 varieties (IR:3) with 6 samples each. The range in effective replication in Table 5 mirrors this. Strategy A is similarly unbalanced. This was difficult to avoid given the restriction of only replicating in the laboratory those field samples that corresponded to individual plot samples. In section 2.2.3 we investigate an alternative that uses the ideas presented in the following section.

Table 5. Example 1: A-values and effective replication for laboratory designs using strategies A and B and a range of values for the field plot variance $\sigma^2_{p_{12}}$.

	$\sigma^2_{p_{12}} = 0.2$		$\sigma^2_{p_{12}} = 0.5$		$\sigma^2_{p_{12}} = 1.0$	
Group	A	B	A	B	A	B
C3:C2	0.74	1.28	0.68	1.12	0.61	0.92
C3:C3	0.74	0.74	0.68	0.68	0.61	0.61
IR:C2	1.73	1.28	1.46	1.12	1.16	0.92
IR:IR:2	2.26	2.25	1.80	1.80	1.35	1.35
IR:IR:3	2.60	2.59	2.03	2.02	1.49	1.49
A-value	0.919	0.9024	0.9797	0.9648	1.0655	1.0535

Table 6. Example 1 simulation study: REML estimates of variance parameters for strategy A, B and C (mean over 700 simulations). Estimates in bold face have bias of greater than 50 percent.

	True	REML estimates		
Source	Value	Strategy A	Strategy B	Strategy C
Variety	1.0	1.048	1.021	1.027
Field rep	0.1	**0.156**	**0.166**	**0.182**
Field column	0.1	0.140	0.132	0.145
Field row	0.1	**0.168**	**0.178**	**0.204**
Field plot	0.2	0.216	0.225	0.277
Mill rep	0.3	0.322	0.319	0.318
Mill session	0.2	0.239	0.229	0.243
Residual	1.0	0.841	0.852	0.800

Table 7. Example 2: Distribution of field and milling samples across testing regimes for 480 entries chosen for milling. C2 entries: tested as composite of two field replicates; R1 entries: tested as single field replicate; R2 entries: tested as two field replicates.

Testing Regime	Entries	Field Plots	Field Samples	Milling Samples
R1	337	337	337	349
C2	114	228	114	139
R2	29	58	58	58
Total	480	623	509	546

2.2.2. Example 2: Genomic selection in wheat population

The 480 entries selected for milling will be tested using either individual field replicate or composite samples as follows:

- Single field sample:

  C2: 114 entries with a composite sample (composite of two replicates)

  R1: 337 entries with a sample from a single replicate

- Two field samples:

  R2: 29 entries with two individual replicate samples

Thus there is a total of 509 field samples to be tested. Note that these involve 623 out of the 637 field plots sown with the 480 chosen entries (see Table 1). The field sampling and compositing is summarised in Table 7.

In section 2.2.1 the varieties to be composited, as opposed to using each replicate separately, were chosen at random. In the current example the decision was driven by the fact that there is a minimum sample size (amount of clean grain) required for milling and subsequent end-product testing. Upon harvesting of this field trial it was found that some plots had insufficient material to form a sample for quality testing. In this case the only mechanism to facilitate quality testing of the associated entry was to use a composite sample (provided the entry was sown with more than one plot). Many of the 114 entries in the C2 group were composited for this reason.

Laboratory replication will be used for 37 of the 509 field samples making a total of 546 laboratory samples for milling. These will be milled as 7 samples per day for each of 78 days. As with the choice of entries to composite, the choice of field samples to replicate in the milling process was influenced by the minimum grain requirements. In section 2.2.1 the samples replicated in the laboratory corresponded to individual field replicate samples. However in the current example there were very few individual replicate samples with sufficient meterial to allow replication in

the milling process. Therefore the majority (25 out of 37) of field samples replicated in the laboratory corresponded to composite samples (see Table 7).

In terms of the laboratory design we use similar blocking factors as in section 2.1.1, the only difference being the use of milling days rather than sessions (half-days). The base-line mixed model for analysis is as in equation (2) with the fixed effects τ_2 comprising an overall mean (and $\boldsymbol{\tau}_1$ omitted), the peripheral effects for the first phase comprising $\boldsymbol{u}_{\boldsymbol{p}11}$ (the 2×1 vector of field replicate effects) and $\boldsymbol{u}_{\boldsymbol{p}12}$ (the 623×1 vector of field plot effects) and the peripheral effects for the second phase comprising $\boldsymbol{u}_{\boldsymbol{p}21}$ (the 2×1 vector of milling replicate effects) and $\boldsymbol{u}_{\boldsymbol{p}22}$ (the 78×1 vector of milling day effects). The design matrices for the first phase involve $\boldsymbol{C}_1$ which is a 546×623 matrix that reflects the compositing strategy. The i^{th} row of $\boldsymbol{C}_1$ (which corresponds to the i^{th} sample) has elements given by:

- 1 in column j if this sample corresponds to field plot j alone
- $1/2$ in columns j and k if this sample corresponds to a composite of field plots j and k
- 0 otherwise

The second phase design can be constructed in `od` (Butler et al., 2013a) using a model that is the same as that described for the analysis except with the addition of random row and column effects for the first phase (denoted $\boldsymbol{u}_{\boldsymbol{p}13}$ and $\boldsymbol{u}_{\boldsymbol{p}14}$ respectively). The starting design is constructed as resolvable so that field samples to be replicated are allocated with one sub-sample in days 1-39 (milling replicate 1) and the other in days 40-78 (milling replicate 2) and the design search is subsequently restricted to swaps within milling replicates. Once again we choose not to confound field and milling replicates so in the starting design ensure that the samples from each field replicate are allocated approximately equally across the two milling replicates. The values chosen for the variance parameters were $\sigma^2_g = 1.0$, $\sigma^2_{p_{11}} = 0.1$, $\sigma^2_{p_{12}} = 0.2$, $\sigma^2_{p_{13}} = 0.1$, $\sigma^2_{p_{14}} = 0.1$, $\sigma^2_{p_{21}} = 0.3$, $\sigma^2_{p_{22}} = 0.2$ (and $\sigma^2 = 1.0$).

Since the concept of forming a field sample by compositing field replicates and then splitting this sample to form milling replicates is novel, a simulation study was conducted in order to examine the reliability of estimation. The model used for data generation was the same as that used for design construction. In each simulation the model fitted to the data matched that used for design generation. A total of 1000 simulations was conducted and the results presented as means (noting that the fitted models converged in all simulations). The estimated variance parameters in Table 8 show some positive bias for the variance of the field plot effects. Importantly, however, there is negligible bias for the genetic variance which is the main focus of the analysis.

Table 8. Example 2 simulation study: REML estimates of variance parameters (mean over 1000 simulations). Estimates in bold face have bias of greater than 10 percent.

Source	True Value	REML estimate
Entry	1.0	0.972
Field rep	0.1	0.093
Field column	0.1	0.097
Field row	0.1	0.105
Field plot	0.2	**0.286**
Mill rep	0.3	0.307
Mill day	0.2	0.199
Residual	1.0	0.956

2.2.3. Example 1: revisited

As noted in section 2.2.1 the restriction of only using individual field plot samples for replication in the laboratory resulted in very unequal numbers of samples across varieties. Given the ideas presented in section 2.2.2 we now re-consider Strategy B from section 2.2.1 and for the 10 field samples to be replicated in the laboratory we now use 6 composite samples (6 of the C3 samples) and 4 individual plot samples. With this strategy (which we shall label 'C') our choice of samples to be replicated in the laboratory led to a design in which there were 4 varieties with a single sample each, 10 varieties with 2 samples each and 4 varieties with 4 samples each. The A-value for an optimised design using this strategy was 0.8147 which is substantially lower than that for strategy B (0.9024). Inclusion of this strategy in the simulation study of section 2.2.1 showed comparable results with strategies A and B in terms of bias in the estimation of variance parameters.

3. Concluding remarks

In this paper we have shown how the use of both individual field replicate samples and composite samples can produce valid experimental designs for a multi-phase variety trial. Replication in the laboratory phase is achieved by splitting some field samples and processing separately. In contrast to the use of composite field samples alone (or a single field replicate), this approach enables the application of an appropriate mixed model analysis to the resultant data. It is superior to the p/q replicate approach of Smith et al. (2006) in the sense that all relevant field plots are used so that the accuracy of predictions of genetic effects will be maximised (also see Smith et al. (2011)). It also has practical advantages compared with the approach of Smith et al. (2006) since it offers a flexible means of dealing with the

problem of minimum grain requirements. The second example was a case where, due to low plot yields, a valid design could only have been achieved with the use of the proposed approach. It was then possible to select varieties to be composited or tested as individual replicates according to the amount of grain harvested from the associated plots. Additionally, sufficient laboratory replication was achieved by splitting some of the composite samples.

In this paper we have used model-based techniques in order to produce designs for the second (laboratory) phase given a design for the first phase. The laboratory phases in multi-phase trials are typically unbalanced and non-orthogonal (even more so with the advent of partial compositing) and have several potential sources of non-genetic variation. An interesting design issue is the manner in which major sources of variation should be accommodated across phases. In the case of orthogonal multi-phase designs Brien et al. (2011) recommend confounding "... big first phase unit sources ... with potentially big second phase unit sources". In our small p/q replicate example the model based design approach resulted in confounding of this nature, with field replicate effects being confounded with milling day effects. This may be a function of the type of model used (that is, a simple variance component model) and the optimality criterion. In terms of the former we note that the analysis of multi-phase traits may involve more complex variance models, in particular spatial correlation structures for field plot effects. At present in `od` (Butler et al., 2013a), correlated effects are only allowed at the residual level (that is, asssociated with the final phase) but this will be addressed in future versions of the software. The use of more complex models for design generation may break the type of confounding observed in the small example. The criterion used was A-optimality so that the goal was to minimise the average pairwise prediction error variance of variety effects. This is a commonly used criterion and likely to be applicable in most cases. However in the first example the aim was to estimate sources of variation rather than predict variety effects so that an alternative criterion may be more appopriate. This is the focus of current research.

Acknowledgements

We gratefully acknowledge the financial support of the Grains Research and Development Corporation of Australia and Wheat Quality Australia (WQA). We acknowledge Marcus Newberry (CSIRO Plant Industry and Food Futures Flagship) as a lead investigator on the WQA project.

References

Brien, C., B. Harch, R. Correll, and R. Bailey (2011). Multiphase experiments with at least one later laboratory phase. i. orthogonal designs. *Journal of Agricultural, Biological and Environmental Statistics 16*, 422–450.

Bueno Filho, J. S. and S. G. Gilmour (2007). Block designs for random treatment effects. *Journal of Statistical Planning and Inference 137*, 1446–1451.

Butler, D. G., B. R. Cullis, A. R. Gilmour, and B. J. Gogel (2009). Analysis of mixed models for S language environments, ASReml-R reference manual release 3. Training and development series, No QE02001, QLD Department of Primary Industries and Fisheries, Brisbane, QLD.

Butler, D. G., A. B. Smith, and B. R. Cullis (2013a). On model based design of comparative experiments. *In preparation*.

Butler, D. G., A. B. Smith, and B. R. Cullis (2013b). On the design of experiments where treatment effects are correlated. *In preparation*.

Cullis, B. R., A. B. Smith, and N. E. Coombes (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological and Environmental Statistics. 11*, 381–393.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Smith, A. B., P. Lim, and B. R. Cullis (2006). The design and analysis of multi-phase plant breeding experiments. *Journal of Agricultural Science, Cambridge 144*, 393–409.

Smith, A. B., R. Thompson, D. G. Butler, and B. R. Cullis (2011). The design and analysis of variety trials using mixtures of composite and individual plot samples. *Journal of the Royal Statistical Society, Series C. 60*, 437–455.