Mixed Models and Experimental Design (MMaED) Lab
NIASRA, University of Wollongong, Australia

A model-based design approach for the design of selection experiments using ODW.

Brian Cullis, David Butler & Alison Smith

Biometrics in the Bay November 2023

Acknowledgements/thanks

- To all collaborating breeders for many helpful discussions and being brave enough to implement the ideas
- Particular thanks to AGT breeders, Intergrain breeders, all pulse breeders, Chris Proud from Rice Breeders Australia
- Special thanks to Kristy Hobson for the use of her MET design in today's talk & to Chris Proud for the use of the rice MET design

- There have been major advances in the analysis of multi-environment trial (MET) data-sets which encompass all stages of a plant improvement program
- Have known for some time that factor analytic linear mixed models (FALMM) consistently provide a good fit to MET data
- Most uses of FALMMs in plant breeding programs incorporate genetic relatedness either using the Numerator Relationship Matrix (NRM) or the Genomic Relationship Matrix (GRM)
- Recently we have developed tools which enhance breeders' confidence in selection by identifying varieties which are high yielding, stable or have similar patterns of Variety by Environment Interaction (VEI)

- Maximal gains from the use of FALMMs will occur if the MET data-set has been constructed and designed in an appropriate manner
- There has been very little attention given to the design of MET data-sets, and even less which utilise genetic relatedness
- Classical approaches to design are incapable of constructing optimal designs which include genetic relatedness and cater for the constraints which exist in the design of METs for early stage selection, such as with seed supply issues, resource allocation constraints and optimal allocation of genotypes across sites . . .

- Model-based designs provide the only sensible framework for construction of designs with the desirable/required properties
- The paradigm is to use a computer-intensive search of the *design space* to generate an optimal design with respect to a pre-specified (analysis) model
- **odw** is an R package which constructs optimal designs under the LMM framework & can generate designs for a wide range of problems:
  1. classical designs such as latinised row column designs
  2. p-rep designs (Cullis *et.* al, 2006)
  3. designs for multi-phase experiments
  4. designs for additive effects - eg diallel experiments
  5. (incomplete) MET designs
  6. selective phenotyping (Huang *et.* al, 2013)

For this talk, we illustrate some of the applications of **odw** which use genetic relatedness, including

- Part One
  - a new approach for single site *p*-rep design - in detail
  - Selective phenotyping within a *p*-rep design - in brief
  - Reduced animal model to construct a *p*-rep design with excessively large numbers of genotypes - in brief
- Part Two
  - design of multi-phase experiments
  - design of METs

Motivating example
Stage one Desi Chickpea experiment

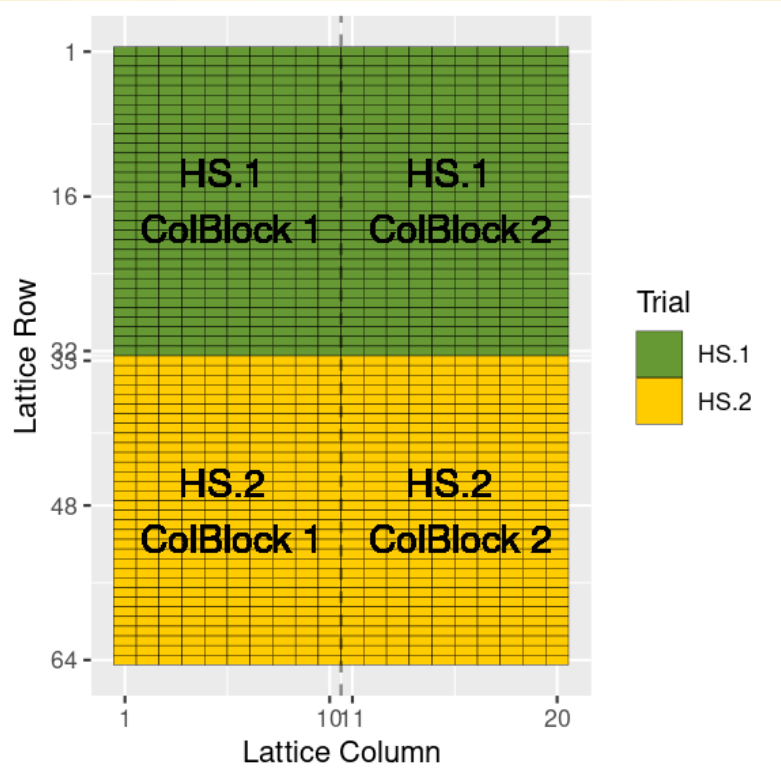
- We consider the multi-environment trial design for stage one (S1) in 2022.
- The full design included two home sites, both at Narrabri, but on two different soil types.
- One home site experiment which was grown on a heavy soil type was assigned to “northern adapted” genotypes & the other on a lighter soil type was assigned to “southern adapted” genotypes.
- There were five satellite sites: two located in the northern (chickpea growing) region and three located in the southern region.

Motivating example
Stage one Desi Chickpea experiment

- Home sites include all genotypes, satellite sites include subsets of genotypes, actual numbers are constrained by seed supply, land availability and cost.
- A total of 4240 genotypes, including test and check varieties.
- We constructed an efficient MET design for all sites, but here . . .
- To illustrate simple principles of model-based design in **odw** we begin with the design for the *south* home site.
- This experiment had 1280 plots and 1139 genotypes, hence the only sensible option is to use a p-rep design (Cullis et al, 2006).

- At the request of the breeder, the experiment comprised two trials each having a manageable number of plots so that sowing and harvesting of plots within a trial could be undertaken on one day
- Each trial was then sub-divided into two large blocks which were aligned with columns within a trial.
- Trials comprised a contiguous rectangular array of plots with 32 rows and 20 cols; column blocks (across trials) comprised 64 rows by 10 columns, separated by ghost plots to facilitate agronomic and cultural operations

Motivating example
S1 Desi Chickpea experiment



Design construction for single site p-rep design
S1 Desi Chickpea experiment

- Our new approach for design construction of a p-rep design almost always involves two stages
  - Stage One: Allocation of packet choice status (i.e. one packet or two packets) to genotypes. Packet is synonymous with plots in the experiment
  - Stage Two: Allocation of plots to genotypes, given packet choice status
- Each stage uses a different call to **odw**

Paradigm for model-based design
Each use of (call to) odw requires:

- An initial configuration (a data-frame in R)
- A linear mixed model to set up the design model
- A *permute* factor - like a GENSTAT 5 *Treatment* factor
- A set of non-permute or so-called *static* factors - a bit like GENSTAT 5 *Block* factors
- A design quality measure - the $\mathcal{A} - criterion$
- A set of *swap* factors - these inform **odw** as to the legal interchanges (see below) which can be made during the design search

1. Initialise the iteration number $N = 1$, calculate $\mathcal{A} - \textit{criterion}$ for the initial design - set as the current design
2. Undertake a legal interchange of the Permute factors between any two plots, subject to the interchange being a legal swap
3. Calculate the $\mathcal{A} - \textit{criterion}$ for the new design obtained from this interchange
4. Accept the new design as the current design if the $\mathcal{A} - \textit{criterion}$ of the new design is less than the $\mathcal{A} - \textit{criterion}$ of the current design
5. $N = N + 1$
6. If $N < N_{max}$ return to 2, else terminate the search

where N_{max} is set by the user in the call to **odw**.

- The linear mixed model for $\bar{\boldsymbol{y}}$, the "pseudo" data vector of $m = 1139$ genotype BLUEs is, :

$$\begin{aligned} \bar{\boldsymbol{y}} &= \mu + \boldsymbol{u_g} + \boldsymbol{e} \\ &= \mu + \boldsymbol{u_a} + \boldsymbol{u_e} + \boldsymbol{e} \end{aligned}$$

  where the vector of total genetic effects $\boldsymbol{u_g} = \boldsymbol{u_a} + \boldsymbol{u_e}$, is the sum of additive and residual genetic effects respectively.
- The total genetic variance is then $\sigma_a^2 \boldsymbol{A} + \sigma_e^2 \boldsymbol{I}_m$, where $\boldsymbol{A}$ is the NRM, $\boldsymbol{I}_m$ is the identity matrix (of size m), and σ_a^2 and σ_e^2 are the additive and residual variances respectively.

- From the baseline LMM, the variance of $\boldsymbol{e}$ depends on the packet choice (i.e. the # plots) and is given by:

$$\text{var}(\boldsymbol{e}) = \begin{cases} \sigma^2 & : \text{ pC1 - one packet} \\ \sigma^2/2 & : \text{ pC2 - two packets} \end{cases}$$

- Computation efficiency is achieved in the design search, by forming a composite error term:

$$\text{var}(\boldsymbol{u_e} + \boldsymbol{e}) = \begin{cases} \sigma_e^2 + \sigma^2 & : \text{pC1 - one packet} \\ \sigma_e^2 + \sigma^2/2 & : \text{pC2 - two packets} \end{cases}$$

The initial data frame contains the following key fields:

- *Genotype* a factor with $m = 1139$ levels - to be used as the Permute factor
- *swp* a factor with three levels to be used as the Swap factor

    - The breeder required the commercial check variety, CBA CAPTAIN, to have two packets (aka plots)
    - 677 test lines only have enough seed for one packet
    - 461 test lines are free to have one or two packets

The initial configuration must comply with:

- Having a factor to define error variance sections, say, `pC` with two levels and a swap factor, say `swp` with three levels which sets up the legal interchanges
- Be ordered by `pC`
- The two-way contingency table for an appropriate initial configuration between `pC` and `swp` is:

swp	pC1	pC2
capt	0	1
one	677	0
two	321	140

Treatment structure `vm(Genotype, NRM)`

Plot structure 1/units

The **odw** call:

```
phi <- c(sigma.ideg,sigma)
sv$Value <- c(sigma.vmg*abar,
 phi[1]+phi[2]/1,phi[1]+ phi[2]/2)
Step1.odw <- odw(fixed=~1, random=~ vm(Genotype, NRM),
 residual = ~dsum(~units|pC), permute=~ vm(Genotype, NRM),
 swap=~swp, R.param = sv, G.param = sv, search = 'tabu+rw',
 maxit=20, data=step1init.df)
```

- Given the packet choice allocation from step 1 we use this to construct a *p*-rep design but do this in two stages.
- Step 2.1 involves construction of a near-resolvable block design with respect to large blocking factors
- Initial data frame has $n = 1280$ plots and an allocation of plots to genotypes which respects the packet choice allocation from stage 1.
- Permute factor is `Genotype`, use the total genetic effects for the calculation of $\mathcal{A} - criterion$.
- Static factors are `Trial` and `ColBlk`

The **odw** call:

Treatment structure `ric(Genotype, NRM)`

Plot structure 1/(Trial + ColBlk)

The **odw** call:

```
sv$Value[1:2] <- c(sigma.vmg*abar,sigma.ideg)
# leave others as defaults
Step2.1.odw <- odw(fixed=~ 1,
 random=~ ric(Genotype, NRM) +
  Trial + ColBlk,
  residual = ~units, permute=~ ric(Genotype, NRM),
 R.param = sv, G.param = sv, search = 'tabu+rw',
  maxit=10, data=init.df)
```

- Given the design from step 2.1 we use now introduce a more appropriate plot structure, viz a near resolvable two-way blocked row-column design.
- Initial data frame is the data frame from the previous call to **odw**
- Permute factor is `Genotype`, use the total genetic effects for the calculation of $\mathcal{A}$ – *criterion*.
- Static factors are `Trial, ColBlk, Column` and `Row`

The **odw** call:

Treatment structure `ric(Genotype, NRM)`

Plot structure 1/(Trial + ColBlk + Column + Trial:Row)

The **odw** call:

```
sv$Value[1:2] <- c(sigma.vmg*abar,sigma.ideg)
# leave others as defaults
Step2.2.odw <- odw(fixed=~ 1,
 random=~ ric(Genotype, NRM) +
 Trial + ColBlk + Column + Trial:Row,
 swap=~Trial:ColBlk,
 residual = ~units, permute=~ ric(Genotype, NRM),
 R.param = sv, G.param = sv, search = 'tabu+rw',
 maxit=10, data=step2.1.odw$design)
```

- Step 2.1 ensures that the design will be resolvable with respect to replicated test lines (and CBA CAPTAIN) having only one plot in each `Trial` and `ColBlk`.
- Step 2.2 finds a design which is optimal with respect to the Rows and Columns of the experiment, but maintains two-way blocking achieved in Step 2.1.
- Note we have not used AR1 $\otimes$ AR1 for the errors in either step. See later in this talk and Cullis *et. al* (2020) for reasons why and later in this talk.

Impact of using genetic relatedness on genetic gain
A small study

- To assess the impact of using genetic relatedness in the design we conducted a small study
- Used the S1 Desi Chickpea experiment, and generated four designs.
- These designs were the factorial combinations of using (+) or not using (-) genetic relatedness in stages one and two of the design construction. So

SG+/+ Uses genetic relatedness in both stages

SG+/- Uses genetic relatedness in stage one only

SG-/+ Uses genetic relatedness for stage 2 & random allocation of those test lines with enough seed to packet choice

SG-/- Does not use genetic relatedness in stages one and two - default *p*-rep designs from Cullis *et.* al (2006) & as generated in DiGGer (Coombes, 2009)

- The quality of each design was assessed by calculating the $\mathcal{A} - criterion$ of the design against the "correct" linear mixed model, using all static terms and the appropriate variance model for the total genetic effects.
- The design with the smallest $\mathcal{A} - criterion$ will result in a higher probability of selecting the best subset of genotypes for progression (Bueno Filho and Gilmour, 2007)
- The $\mathcal{A} - criteria$ for each design, expressed as the difference from design SG+/+, and multiplied by $1e4$ were

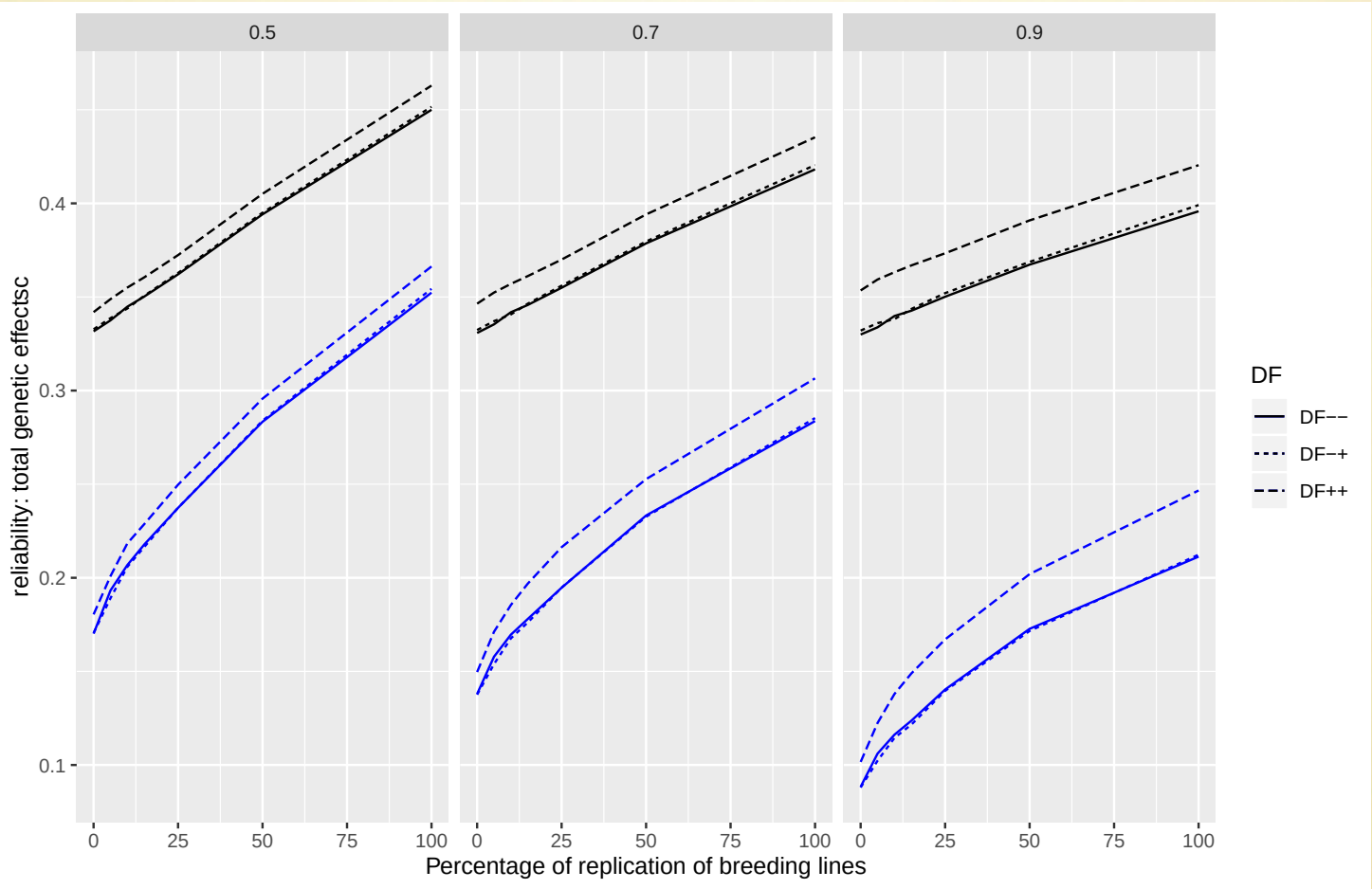
$$\text{SG+/+} = 0 \qquad \text{SG-/+} = 27$$
$$\text{SG+/-} = 15 \qquad \text{SG-/-} = 43$$

Simulation study from Cullis et al (2020)
In silico study aims and set up

- In silico experiment conducted to assess the impact of
  - Matching the design and data models exactly, viz matching the genetic and non-genetic models
  - Partial matching the design and data models, viz matching the non-genetic model alone, using fixed treatment effects
  - No matching of the design and data models, using fixed treatment effects and augmented *p*-rep designs after Williams et al (2011)
- This was achieved using three treatments
  - DF+/+: Matching the design and data models exactly
  - DF-/+: Matching the non-genetic model alone
  - DF-/-: Not matching either the genetic or non-genetic model
- Data genetic model included additive and non-additive effects
- Data non-genetic model included ColBlk, Rows, Columns and first order separable AR1$\times$AR1 process for the errors

- The impact of treatments was assessed using pedigree information for 260 fababean varieties laid out in a field trial as a rectangular array, across 63 ancillary treatments consisting of the factorial combinations of
  - Seven levels of p-replication: $(0, 5, 10, 15, 25, 50, 100)$
  - Three levels of proportion additive to total genetic variance: $(0.5, 0.7, 0.9)$, and
  - Three levels of reliability: $(1/3, 1/2, 2/3)$
- 4000 simulations were conducted for each of the 63 combinations using the three DF's on each of the 63×4000 simulated data-sets
- We present one set of results here:

Simulation study from Cullis et al (2020)
Results rsq=1/3



- Let's assume that there are only enough resources to sow 640 (not 1280) plots in the S1 chickpea experiment,
- How do we choose an optimal subset of the available genotypes to phenotype?
- Huang *et.* al (2013) referred to this a *selective phenotyping* & proposed a simple method based on forming $m < 640$ clusters, and selecting one genotype from each cluster, where m is chosen sensibly to maintain approximately 10% partial replication.
- We propose an alternate approach using **odw**, which can be used for any design, and with either ancestral or marker based relationship matrices.
- Our approach is a simple extension to the allocation of packet choice status to genotypes.

- All fields in the initial data-frame are the same as stage 1 in previous *p*-rep example, except the packet choice factor pC has an additional level:

$$\text{var}(\boldsymbol{u_e}) + \text{var}(\boldsymbol{e}) = \begin{cases} \sigma_e^2 + \sigma^2 : \text{pC1 - one packet} \\ \sigma_e^2 + \sigma^2/2 : \text{pC2 - two packets} \\ \sigma_e^2 + \alpha\sigma^2 : \text{pC3 - no packets} \end{cases}$$

- where α is set to a large number - so that if a genotype is allocated to pC3 then it will not be phenotyped (the weighting for these genotypes is effectively zero - hence it is dropped from the data)

Our approach can incorporate many constraints. For example the two-way contingency table for swp and pC is

swp	pC1	pC2	pC3
capt	0	1	0
ch(slash/str)	1	1	0
one	338	0	339
two	163	67	229

- Only 363/571 genotypes were in common between the two methods.
- Huang's method creates 568 (571-3 checks) clusters and selects one genotype from each cluster. Clearly piecemeal and hence inefficient subset selection.

Reduced animal model to reduce computing time
Overview

- Idea was motivated by an Intergrain barley MET design for stage one selection which had greater than 15,000 genotypes and more than 6 environments.
- Details too complex for this talk.
- Basic concept is to split genotypes in the full pedigree into parental and non-parental genotypes
- Exploit the recursive model for the additive genetic effects of non-parental genotypes as the mean of the parental additive genetic effects plus a Mendelian sampling term.
- Use only the mean of parental additive genetic effects for the design search(es).

- Compute times are reduced as there are much fewer parental genotypes (43 vs 1139).
- For step 2.2 with 200 TABU loops times were
  - RAM: 40 seconds
  - FULL: 4200 seconds
- Difference in $\mathcal{A} - \textit{criteria}$ relative to FULL $\times 1e4$ were (including earlier designs)

$$\begin{aligned} \text{SG+/+} &= 0 & \text{SG-/+} &= 27 \\ \text{SG+/-} &= 15 & \text{SG-/-} &= 43 \\ \text{RAM} &= 9 \end{aligned}$$

Concluding remarks: Part One

- Have demonstrated the potential gains in accuracy of selection by using genetic relatedness for simpler design problems
- Butler, Smith and Cullis (submitted). “On Model Based Design of Comparative Experiments in R.”
- Part Two - another talk
  - Using (genetic) relatedness and the features of **odw** for the design of multi-phase experiments offers substantial benefits
  - The potential increases in genetic gain from use of MET designs would most likely exceed those obtained from the design of single experiments

- The enzyme, α-amylase, is responsible for the degradation of starch into sugars in wheat grains.
- If it occurs at high levels, it significantly reduces the end-product quality of the grain.
- The falling number (FN) test is an internationally accepted standard as the field-based surrogate for assessing the suitability of grain for human consumption. Grain samples which have a FN of less than 300 seconds are downgraded at receival.
- FN does not directly measure α-amylase content in the grain, but measures changes in the physical properties of the starch portion of the wheat kernel caused by α-amylase (Perten, 1964).

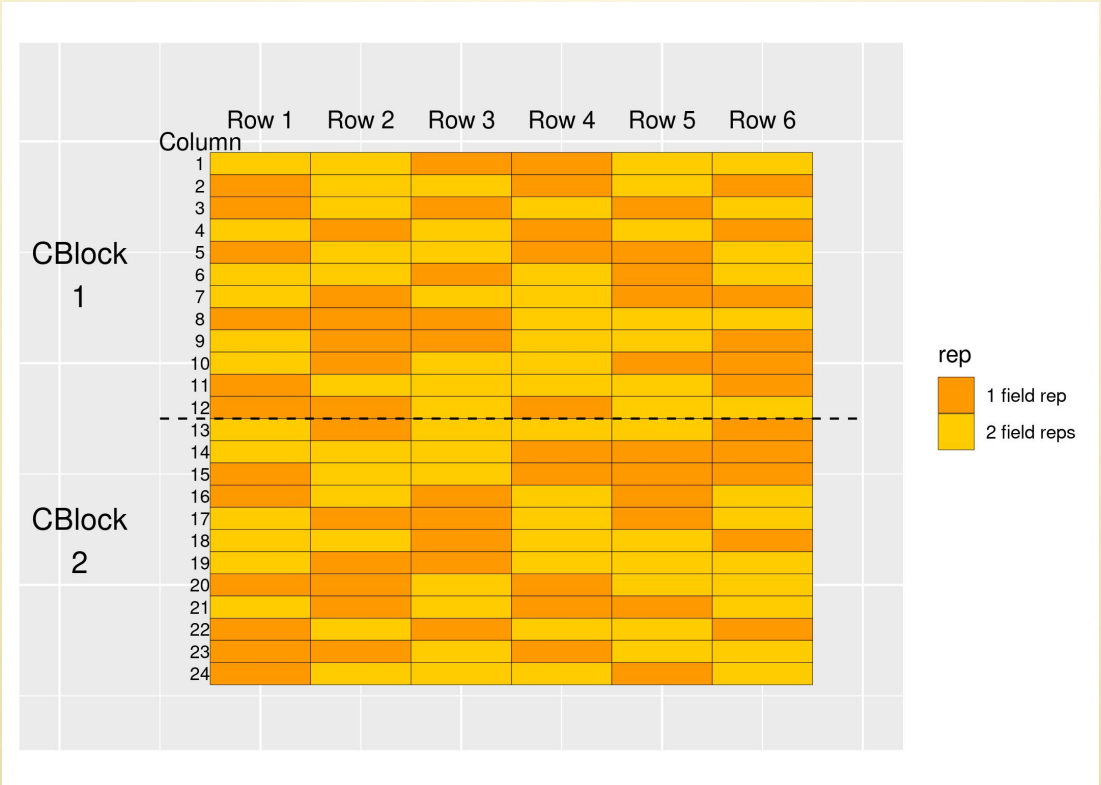
Phenotyping complex expensive quality traits
Multi-phase experiment for wheat quality

- The measurement of FN requires a multi-phase experiment (Brien, 1983; Smith et al., 2006) with two phases.
- The varieties are first grown in a field experiment (Phase I) and after this has been harvested, grain samples from individual plots are processed in a laboratory experiment (Phase II) to obtain the trait of interest.
- Several authors, including Brien (1983) and Smith et al. (2006), have stressed the need for the use of valid experimental designs for all phases of a multi-phase experiment.
- However, this rarely occurs in practice and one of the major impediments has been the lack of suitable software to generate optimal designs.
- Here we use **odw** to generate an efficient design for the seemingly complex scenario of multi-phase experiments.

- Phase I involved a field experiment and comprises 144 plots arranged in a rectangular array of 24 columns by 6 rows.
- A total of 105 varieties will be grown using a partially replicated (*p*-rep) design in which 39 varieties will be planted in 2 plots each while the remaining 66 varieties will be planted in single plots.
- In this example, there is no information available on the genetic relatedness of the varieties so that the varieties to be replicated will be chosen at random.

Phase I: Field experiment
Plot dimensions 10 × 2; not to scale

- *p*-rep design with replicated varieties resolvable with respect to column blocks
- Expect extraneous variation aligned with rows and columns due to agronomic practices.



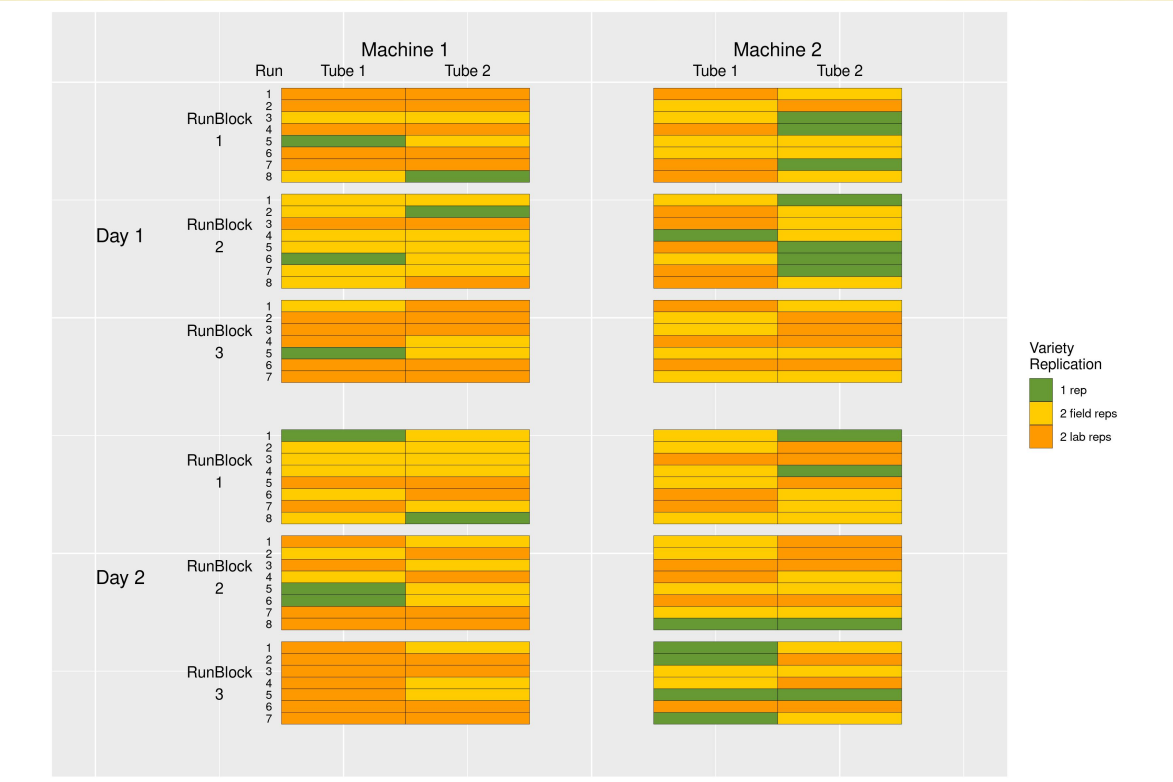
Multi-phase experiment for wheat quality
Phase II: Laboratory phase

- The laboratory experiment in Phase II involves the production of a slurry from each grain sample taken from the field.
- Each slurry is then placed in a tube on a FN machine to measure the trait, which is the time taken (in seconds) for a rod to travel through the slurry.
- In this experiment, samples from all 144 field plots will be processed, and replication in Phase II will be achieved by producing two slurries (from two separate grain samples) for a subset of the plots.
- As with the field experiment, a partially replicated (q-rep) design will be used to reduce cost and time.
- In Phase II, 40 plots will be tested using two slurries while the remaining 104 plots will be tested as single slurries, making a total of 184 slurries to be processed.

- The choice of plots to be replicated in Phase II can be made in an informed manner, using **odw** and an appropriate model.
- The slurries will be processed using two FN machines, each of which comprises two tubes.
- This allows four slurries to be processed simultaneously, and these will be referred to as a *run*.
- Thus the full Phase II design will require 46 runs which are processed sequentially.
- Practical considerations necessitate the grouping of runs into blocks with (no more than) 8 runs in each, and with 3 blocks per day.
- The full design spans 2 days, and the final block on each day will have 7 rather than 8 runs.

Phase II: Schematic diagram for laboratory phase

- *q*-rep design with replicated field plots (and varieties) resolvable with respect to days.
- Also expect variation from machine, tube within machine, run-blocks and runs.



- Recall . . .
  - 144 plots arranged in a rectangular array of 24 columns by 6 rows.
  - 105 varieties in which 39 varieties planted in 2 plots each, remaining 66 varieties planted in single plots.
- Design construction commences with defining the plot and treatment structures, after Bailey (2008) and Smith and Cullis (2020).
- Smith and Cullis (2020) developed the so-called Design Tableau approach for this process.

Treatment structure `Variety`

Plot structure `ColBlock + Column + Row + Column:Row`

- Note that in this example, the requirement for resolvable blocks was met using a single **odw** call:

```
Phase1.od <- odw(fixed=~ 1,
 random=~ Variety + ColBlock + Column + Row,
 residual = ~units, permute=~ Variety,
 search = 'tabu+rw',maxit=10, data=start.design)
 Phase1.od <- update(Phase1.od,maxit=10)
```

Application of the design tableau approach gives:

Treatment structure `Variety + ColBlock + Column + Row + Column:Row`

Plot structure `(Day/RunBlock/Run) * (Machine/Tube)`

Construction of an optimal design involves two stages

- The first stage is similar to determining the replication status of genotypes which was used in the chickpea example.
- Here, although there is no information on genetic relatedness the presence of linked factors in the permute set (i.e. treatments) from Phase I provides a convenient mechanism for finding an optimal set of field plots to replicate in Phase II.

The interim LMM used to determine an optimal subset of 40 field plots from Phase I to be replicated in Phase II is given by

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z_g u_g} + \boldsymbol{Z_c u_c} + \boldsymbol{Z_r u_r} + \boldsymbol{Z_b u_b} + \boldsymbol{u_p} + \eta$$

- $\boldsymbol{y}$ is the pseudo data vector representing field plot means (across laboratory replicates) of length $n = 144$
- μ is an overall mean parameter,
- $\boldsymbol{u_g}$, $\boldsymbol{u_c}$, $\boldsymbol{u_r}$ and $\boldsymbol{u_b}$ are the vectors of random *Variety*, *Column*, *Row* and *ColBlock* effects.
- $\boldsymbol{u_p}$ and η, both have length n and represent the field plot effects and Phase II errors, respectively.

- We assume the variance matrix for $\boldsymbol{\eta}$ is given by $\oplus_{i=1}^{2} \frac{\sigma^2}{r_i} \boldsymbol{I}_{n_i}$ where $n_1 = 104,\ n_2 = 40,\ r_1 = 1$ and $r_2 = 2$.
- As before, we take advantage of the identity design matrix for $\boldsymbol{u_p}$ to reduce computations and define a combined vector of errors,

$$\boldsymbol{\eta}^* = \boldsymbol{u_p} + \boldsymbol{\eta}$$

- so that $\text{var}(\boldsymbol{\eta}^*) = \oplus_{i=1}^{2} \left(\sigma_p^2 + \frac{\sigma^2}{r_i}\right) \boldsymbol{I}_{n_i}$.
- In order to define this heterogeneous error variance structure in **odw** we require the initial data-frame to include a two level factor (called *pC*) to be consistent with the first example, which has the value 1 for the first 104 records and 2 for the remainder.

The **odw** call to allocate levels of laboratory replication (associated with the factor *pC*) to varieties is:

```
PhaseIIrep.out <- odw(fixed=~1,
  random=~Variety + Column + Row + ColBlock,
  residual=~dsum(~units|pC),
  permute=~Variety|ColBlock + Column + Row,
  R.param=sv, G.param=sv, search='tabu+rw',
  data=dup.df, reorder=c('FieldPlot'), maxit=10)
```

- `dup.df` is the initial data-frame with 144 rows, indexed by field plots.
- We use bespoke variance parameter values provided in *sv*, namely $\sigma^2_g = 1.0$, $\sigma^2_c = 0.1$, $\sigma^2_r = 0.1$, $\sigma^2_b = 0.1$, $\sigma^2_p = 0.5$ and $\sigma^2 = 1.0$.
- The last two values in *sv* which relate to the combined vector of errors were 1.5 and 1.0 for levels 1 and 2 of *pC*, respectively.

Reiterating:

```
PhaseIIrep.out <- odw(fixed=~1,
  random=~Variety + Column + Row + ColBlock,
  residual=~dsum(~units|pC),
  permute=~Variety|ColBlock + Column + Row,
  R.param=sv, G.param=sv, search='tabu+rw',
  data=dup.df, reorder=c('FieldPlot'), maxit=10)
```

- This call demonstrates the '|' and `reorder` arguments for managing the permute (P), objective (O) and linked (L) sets of effects where $P = O \cup L$
- The O set appears in the `permute` argument before the '|'
- The effects in the L set occur after the '|', but additional effects which are associated with the permute set, but not in the LMM are specified in the `reorder` argument.

Treatment structure `Variety + ColBlock + Column + Row + Column:Row`

Plot structure `(Day/RunBlock/Run) * (Machine/Tube)`

Construction of an optimal design involves two steps. Given the optimal set of field plots to duplicate we

- Use a simplified LMM with major blocking factors in S to find a design which is resolvable for varieties
- Use this design as the starting design for a search with the full set of effects in S.

frep	pC1	pC2
Vars with 1 plot	26	40
Vars with 2 plots	39	0

```
PhaseIIbin.od <- odw(fixed=~1,
  random=~Variety + Day + Machine,
  residual=~units, permute=~Variety,
  reorder= c('FieldPlot','Column','Row','ColBlock'),
  search='tabu+rw', maxit=10, data=init.lab.df)
  PhaseIIbin.od <- update(PhaseIIbin.od, maxit=10)
```

where

- `init.lab.df` is the initial data-frame containing 184 records, indexed by slurries.
- the resultant design was resolvable for `machine` and `day`, and was used as the initial design in step 2.2 below

- Three designs were created, each using a different model for the swap formula, in order to investigate the impact on $\mathcal{A}-$optimality of the restrictions imposed by resolvability.
- The formulae were:

1. swap=~Day : Machine
2. swap=~Day
3. swap=~NULL

- First restricts interchanges within the intersection of `Day` and `Machine`, the second to `Day` and the third has no restrictions
- Apart from the `swap` formula, the **odw** call to construct all three designs was identical. The call for the first design is given by:

Phase II: Laboratory, resolvable q-rep design
Step 2.2


```
PhaseII.od <- odw(fixed=~ 1, random=~Variety + FieldPlot +
  Column + Row + ColBlock + Day +
  Day:RunBlock+Day:RunBlock:Run + Machine +
  Machine:Tube + Day:Machine + Day:RunBlock:Machine +
  Day:RunBlock:Run:Machine +
  Day:Machine:Tube + Day:RunBlock:Machine:Tube,
  swap=~Day:Machine, residual= ~units,
  permute=~Variety|FieldPlot + Column + Row + ColBlock,
  search='tabu+rw', maxit=10, data=PhaseIIbin.od$design)
PhaseII.od <- update(PhaseII.od, maxit=50)
```

- This call again demonstrates the '|' for managing the O and L sets of effects.
- The O set appears in the `permute` argument before the '|'
- The $\mathcal{A}$- values for the three designs were (1) 0.179415, (2) 0.179284 and (3) 0.179202; demonstrating the penalty associated with the restrictions for resolvability.
- In the second design, the removal of the resolvability restriction associated with *Machine* led to a non-binary design for this factor.
- Similarly, the third design, with the lowest $\mathcal{A}-$value, was non-binary for both *Day* and *Machine*.
- This was regarded as undesirable by the researcher, so, as a compromise, the design with resolvability for days alone was adopted.

- To examine the benefit of choosing field plots to replicate using our method we generated three designs.
  - **D1** Using the plot structure for Phase I: our method
  - **D2** Using **odw** but without the plot factors: replicate those field plots from varieties with one plot
  - **D3** At random
- The quality of each design was assessed by computing the $\mathcal{A}$ of the design against the full LMM.
- The $\mathcal{A}$-values for each design, were .179284, .179308 and .179467 respectively

- Another talk . . .
- Principles are similar to those presented in this talk.
- Genetic relatedness is used to:
  - Allocate packets to genotypes subject to constraints in terms of seed supply, home sites, numbers of sites and plots within sites and so on
  - Allocate sites to genotypes subject to home sites and regional adaptation of genotypes to site types
  - Allocate plots to genotypes - achieved in two steps, allowing only interchanges within sites

- A total of 419 stage one and 43 stage three early stage, advanced breeding lines and released varieties are to be tested across 14 environments in southern NSW and northern NSW and Southern Queensland in the current season
- Ancestral information for all genotypes is available
- Breeder has specified a range of constraints in terms of seed supply and non-negotiable site allocations for each genotype

Rice breeding MET example

Summary of sites: H sites are for S1 and S4; S sites are for S4

Site		Number of		
No.	Name	Plots	Rows	Columns
1	H1	630	30	21
2	H2	600	30	20
3	H3	600	3	20
4	Sa	90	10	9
5	Sb	90	10	9
6	Sc	90	10	9
7	Sd	90	10	9
8	Se	90	10	9
9	Sf	90	10	9
10	Sg	90	10	9
11	Sh	90	10	9
12	Si	99	9	11
13	Sj	99	9	11
14	Sk	90	10	9

Rice MET example
Design Requirements

Packets Start	Packets Used	Stage	Number of Target Sites	Genotype Frequency
1	1	S1	1	73
2	2	S1	2	118
3	3	S1	3	98
4	4	S1	3	49
5	5	S1	3	3
6	6	S1	3	9
7	7	S1	3	20
8	8	S1	3	1
9	9	S1	3	5
10	9	S1	3	34
10	10	S1	3	9
4	4	S4	3	1
9	9	S4	3	5
9	9	S4	5	1
15	15	S4	5	4
18	18	S4	6	2
33	33	S4	11	3
36	36	S4	12	1
42	42	S4	14	26

- Principles are similar to those presented in this talk but too complex to present here.
- Genetic relatedness is used to:
  - Allocate packets to genotypes subject to constraints in terms of seed supply, home sites, numbers of sites and plots within sites and so on
  - Allocate sites to genotypes subject to home sites and regional adaptation of genotypes to site types
  - Allocate plots to genotypes - achieved in two steps, allowing only interchanges within sites

Rice MET example
Anatomy of the final MET design: 79% for H sites

Site			Number of		S1 replications					S4 replications			
No.	Name	Plots	Rows	Columns	0	1	2	3	4	0	1	2	3
1	H1	630	30	21	0	341	71	7	0	0	1	1	41
2	H2	600	30	20	137	181	40	29	32	1	1	0	41
3	H3	600	3	20	127	185	49	38	20	2	0	0	41
4	Sa	90	10	9	0	0	0	0	0	13	0	0	30
5	Sb	90	10	9	0	0	0	0	0	13	0	0	30
6	Sc	90	10	9	0	0	0	0	0	13	0	0	30
7	Sd	90	10	9	0	0	0	0	0	13	0	0	30
8	Se	90	10	9	0	0	0	0	0	13	0	0	30
9	Sf	90	10	9	0	0	0	0	0	13	0	0	30
10	Sg	90	10	9	0	0	0	0	0	13	0	0	30
11	Sh	90	10	9	0	0	0	0	0	13	0	0	30
12	Si	99	9	11	0	0	0	0	0	9	1	1	32
13	Sj	99	9	11	0	0	0	0	0	9	1	1	32
14	Sk	90	10	9	0	0	0	0	0	13	0	0	30

Concluding remarks

- Have demonstrated the potential gains in accuracy of selection by using genetic relatedness for simpler design problems
- The potential increases in genetic gain from use of MET designs would most likely exceed those obtained from the design of single experiments
- Butler, Smith and Cullis (submitted). “On Model Based Design of Comparative Experiments in R.”