

NIASRA

NATIONAL INSTITUTE FOR APPLIED
STATISTICS RESEARCH AUSTRALIA



***National Institute for Applied Statistics Research
Australia***

University of Wollongong, Australia

Working Paper

11-17

Sparse Designs for Multi-Environment Trials

Brian Cullis, Nicole Cocks, Alison Smith and David Butler

*Copyright © 2021 by the National Institute for Applied Statistics Research Australia, UOW.
Work in progress, no part of this paper may be reproduced without permission from the Institute.*

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia Phone +61 2 4221 5076, Fax +61 2 4221 4998.

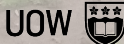
Email: karink@uow.edu.au

Sparse designs for multi-environment trials


Brian Cullis, Nicole Cocks, Alison Smith and David Butler

Centre for Bioinformatics and Biometrics
NIASRA, University of Wollongong, Australia

CEN-ISBS Conference, August 2017
Vienna



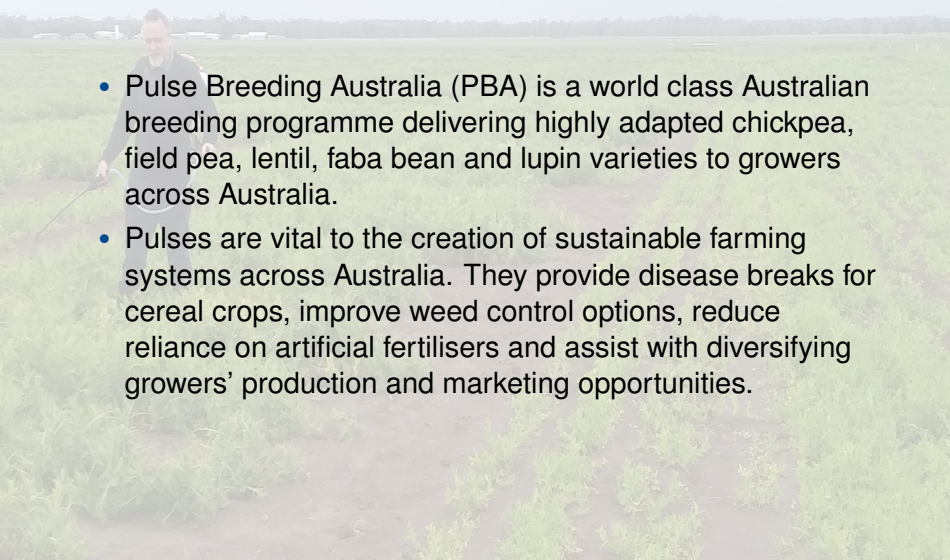
Acknowledgements

A person wearing a dark jacket and pants is standing in a field of green crops. They are holding a white hose that loops around their waist. The background shows a line of trees under a clear sky.

CBB gratefully acknowledges the Grains Research and Development Corporation (GRDC) for their funding. We would also like to thank Garry Rosewarne for the motivating example, his willingness to implement the designs and many useful discussions.

Background and motivation

Field peas breeding programme: Pulse Breeding Australia

- 
- A man in a dark jacket and light-colored shirt is standing in a field of green crops, possibly a pulse crop. He is holding a long-handled tool, possibly a hoe or a similar agricultural implement. The field is vast and green, with a line of trees and buildings in the background under a clear sky.
- Pulse Breeding Australia (PBA) is a world class Australian breeding programme delivering highly adapted chickpea, field pea, lentil, faba bean and lupin varieties to growers across Australia.
 - Pulses are vital to the creation of sustainable farming systems across Australia. They provide disease breaks for cereal crops, improve weed control options, reduce reliance on artificial fertilisers and assist with diversifying growers' production and marketing opportunities.

Overview of PBA field pea breeding programme

Stages of testing

Stage	Varieties (Lines)	Sites	Reps/design
S0	400-800	1-2	grid-plot, p -rep or 2 rep RCB
S1	300-500	3	p -rep or 2 rep RCB
S2	200-300	8	2 rep RCB
S3	150-250	14	3 rep RCB
S4	20-40	14	3 or 4 rep RCB

- p -rep refers to the partial replicate designs of Cullis *et al.* (2006)
- RCB - randomised complete block design
- grid-plot designs are being phased out

PBA field pea programme

Selection procedures

- Annually, MET analyses are conducted to determine which lines progress through the stages of the testing system
- Progression is contingent on so-called selection decisions (SD); SD1: S0 to S1, SD2: S1 to S2, SD3: S2 to S3, SD4: S3 to S4 and SD5: S4 to release
- SDs are (mainly) applied to sets of lines within so-called contemporary groups (CGs)
- A CG is a set of lines which enter the testing system in the same calendar year

PBA field pea programme

Selection procedures

- The structure of testing system and progression of lines within CGs can be displayed in a CG progression array:

Stage	Year					SD
	2012	2013	2014	2015	2016	
S0	E	D	C	B	A	SD1
S1	-	E	D	C	B	SD2
S2	-	-	E	D	C	SD3
S3	-	-	-	E	D	SD4
S4	-	-	-	-	E	SD5

- where in this example the five CGs are denoted by Roman letters A, B, C, D and E
- recall that the number of lines which are in a CG at S0 is around 700-900
- there are multiple trials within each cell, and the combination of trials within year and years as *environment*

PBA field pea programme

Analysis procedures

The MET analysis is an extension of the Smith *et al.*, (2001) approach which uses a one-stage linear mixed model and

- partitions the total genetic variety by environment (VE) effects into additive and non-additive VE effects (VEa and VEe respectively)
- uses factor analytic models for the variance structure of the VEa and VEe
- incorporates the plot structure and accommodates modelling other sources of non-genetic variation such as spatial dependence
- accommodates variance heterogeneity for all random terms which are environment specific

PBA field pea programme

Analysis procedures: key results

Results of MET analyses for the previous 5 years indicate:

- Variety by environment interaction (VEI) is a *major* source of variation at all stages of the testing system
- VEI_a is much larger than VEI_e particularly for S0, S1 and S2 (of the order of 4:1, but varies according to the environment)
- moderate levels of non-genetic variation, dominated by spatial variation and other sources of non-genetic variation associated with extraneous variation due to sowing, harvesting and so on

PBA field pea programme

Experimental Design: Current limitations and challenges

- Irrespective of the trait used for selection (either overall yield performance or yield stability),
- genetic gain within a CG is a function of the accuracy of the predictions for the VEa and VEe effects
- Accuracy is lowest for the most critical stages of testing (viz S0, S1 and S2) due to limited testing across environments (years and locations) and low replication within an environment
- MET designs used in the MET analyses for CG A and CG B were:
 - CG A: 404 lines tested at one location in a p-rep design
 - CG B (2015): 666 lines tested at one location in a grid plot design
 - CG B (2016): 373 lines tested at three locations in a p-rep design

The S0 MET design for 2017

- Garry suggested a MET design with a total of 870 lines tested in two environments, with two replicates of each line in each environment
- This design has complete fill-in of the $V \times E$ table and a total of 3480 plots
- We consider this MET design to be a sub-optimal use of resources for various reasons including
 - Given the presence of substantial VEI use of two environments and two replicates per environment will provide less accurate predictions of VEa and VEe effects, except for pathological cases where there is zero genetic correlation between all pairs of environments,
 - Use of two replicates of each line for both environments is unnecessary to adequately model the non-genetic variation, and

The S0 MET design for 2017

- The distribution of the number of packets for each line precludes use of *all* lines, viz

	1	2	3	4	5
Lines	9	24	51	513	273

- Perhaps we could consider a MET design with say three environments and use a p-rep design for each environment with $p \simeq 0.30$ (for the majority of lines)
- but we can go even further . . .

Sparse MET designs

Definition

A sparse MET design is a MET design which involves increasing the number of environments by judiciously decreasing the fill-in of the $V \times E$ table along with the amount of replication of lines within environments subject to

- maintenance of a minimum level of partial replication of lines within environments
- maintenance of high fill-in for the pedigree-based or marker based relationships between environments
- maintenance of moderate fill-in of the $V \times E$ table

Sparse MET designs

Towards a unified model based approach

- Designs with spatially dependent plot errors have been recently considered (Butler *et al.* 2008; Cullis *et al.* 2006; Williams *et al.* 2006; Coombes 2002; Chan 1999; Martin and Eccleston 1997; Martin 1986)
- Bueno Filho and Gilmour (2003) (BFG) consider (random) correlated treatment effects for non-resolvable incomplete block designs.
- Williams *et al.*, 2011 and Moring *et al.*, 2014 propose augmented p-rep designs for METs using an extension of the α -design concepts.
- Butler *et al.* (2014) extended the work of Coombes (2002) to include random treatment effects in a model-based framework

Sparse MET designs

Towards a unified model based approach

In order to develop a unified approach for generating a (near) optimal design function (Bailey, 2008) for a Sparse MET design we must (potentially) account for

- spatially dependent plot errors
- correlated treatment effects
- partitioning of the total VE effects into additive (either marker or pedigree based) and non-additive VE effects
- a MET framework (i.e. VEIa and VEIe)
- so-called swap blocks

Sparse MET designs

A unified model based approach

- We aim to search a design space for a configuration which is optimal in some sense under a pre-specified linear mixed model
- BFG have shown that A -optimality is a useful optimality criterion in plant selection experiments
- Choose a design function which minimises A by searching the design space \mathcal{D} , under the supervision of an optimisation strategy such as TABU search (Glover 1989)
- Coombes (2002) implements TABU in the DiGGeR framework
- Butler (2013) developed `od` which uses TABU, allows for correlated treatment effects and swap blocks ...

Sparse MET designs

Prediction as a basis for optimal design

- From Gilmour *et al.*, (2004), let

$$\tilde{\pi} = D\tilde{\beta}$$

be a vector of (estimable) predictions for some known $D = [D_{\tau} \ D_u]$, where $\tilde{\beta}$ is any non-unique solution to the mixed model equations (MMEs) with coefficient matrix C

- The prediction error variance of $\tilde{\beta}$ is given by

$$\Lambda = \text{var}(\tilde{\pi} - \pi) = DC^{-1}D^{\top}$$

- Our objective function (i.e. an optimality criteria) is then simply a function of Λ
- For Sparse MET designs we use A -optimality (see Butler *et al.*, 2014))

Sparse MET designs

Example: PBA S0 2017 design

- Genetic material:
 - six check varieties
 - 870 (test) lines from 466 families with between 1 to 10 F5/F2 derived full-sibs per family
 - pedigree file contained 1382 individuals
- Constraints:
 - So-called home site - all test lines present
 - two plots of each check variety in all environments
 - Minimum of 5% p-replication at all environments
- Resources: *site and cost related*

Site	nCol	nRow	nBlock	nRowperBlk	nPlot
1	12	88	4	22	1056
2	12	50	2	25	600
3	12	50	2	25	600
4	12	50	2	25	600

Sparse MET designs

Example: PBA S0 2017 design

- Resources: Packets of seed, 2856 plots but 3675 packets available

Packets	Lines	
	Available	Final
1	9	28
2	24	305
3	51	151
4	513	213
5	273	173
8	6	6
Sum	3675	2856

- Reduction algorithm focussed on lines which had (multiple) full-sibs to ensure maintenance of genetic diversity

PBA S0 2017 design

Design process

- Swap blocks: lines per swap block

Site	Swap Block						Total
	A	B	C	D	E	F	
1	870	174	12				1056
2		588		12			600
3		588			12		600
4		588				12	600

- Swap block A: one plot of each test line at site 1; B: all other plots containing test lines; C, D, E F: plots containing two plots of each check variety for sites 1,2,3 and 4 respectively
- Hence design generation requires two steps:
 - allocation of lines to environments (aka sites)
 - allocation of lines to plots within sites

PBA S0 2017 design

Stage 1 allocation of lines to site: Desirable call

```
s0.od1 <- od(fixed=~ Site, random = ~ ped(Line) + ide(Line)
+ ped(Line):Site + ide(Line):Site,permute=~ ped(Line) +
ide(Line)|ped(Line):Site + ide(Line):Site, swap=~Swapblk,
geneticeffects='total', search='tabu', maxit=1, ...)
```

- **search='tabu'**: specifies TABU search with 1 cooling loop,
- **permute=~ ped(Line) + ide(Line) + ...**: specifies the terms which define the objective function along with additional model terms to be moved with each swap (placed after the "|")
- **geneticeffects='total'**: optimality criteria computed for total genetic effects
- od requires both an initial design specified in the **data** argument

Time constraints and computational load necessitated a simplified model and search algorithm . . .

PBA S0 2017 design

Stage 1 allocation of lines to site: Actual call

```
s0.od1 <- od(fixed=~ Site, random = ~ ped(Line),  
swap=~Swapblk, geneticeffects='additive',  
search='random', maxit=1500, data= ...)
```

- **search='random'**: specifies random search with 1500 interchanges rather than using **'tabu'**,
- **permute=~ ped(Line) ...**: reduced the terms in this to only include **ped(Line)**
- **geneticeffects='additive'**: optimality criteria computed for additive genetic effects instead of **'total'**
- od requires both an initial design specified in the **data** argument

PBA S0 2017 design

Stage 2 allocation of lines to plots within sites: Desirable call

```
s0.od2 <- od(fixed=~ Site, random = ~ ped(Line) +  
ide(Line) + ped(Line):Site + ide(Line):Site + Site:Cblk  
+Site:Column+Site:Row,permute=~ ped(Line) +  
ide(Line)|ped(Line):Site + ide(Line):Site, swap=~Site,  
geneticeffects='total', search='tabu', maxit=1,  
data=s0.od1$design, ...)
```

- Changed the **swap** argument from `=~ Swapblk` to `=~Site`
- added terms to the **random** model formula for columns and rows within site
- specified the start design to be the design found in stage 1
- Spatial dependence not included to avoid computational load

PBA S0 2017 design

Stage 2 allocation of lines to plots within sites: Actual call

```
s0.od2 <- od(fixed=~ Site, random = ~ ped(Line) +  
Site:Cblk + Site:Column + Site:Row, permute=~ ped(Line),  
swap=~Site, geneticeffects='additive', search='random',  
maxit=4500, ...)
```

- **search='random'**: specifies random search with 4500 interchanges rather than **'tabu'**,
- **permute=~ ped(Line) ...**: reduced the terms in this to only include **ped(Line)**
- **geneticeffects='additive'**: optimality criteria computed for additive genetic effects instead of **'total'**
- increased the number of random interchanges to improve efficiency of the design with respect to additional plot structures (rows and columns) which are likely to be present

PBA S0 2017 design

Properties of final design

- Produced a sparse MET design with the desired properties
- Maintained a minimum partial replication of lines within sites of 5%
- Maintained a moderate degree of fill-in in the $V \times E$ table

Site	Initial					Final				
	1	2	3	4	p-rep	1	2	3	4	p-rep
1	876	594	545	395	21.4	876	568	562	554	21.4
2		594	292	244	1.0		568	386	374	5.6
3			545	395	10.9			562	387	6.8
4				395	51.9				554	8.3

PBA S0 2017 design

Properties of final design

- Produced a sparse MET design with the desired properties
- Maintained a high degree of fill-in of the additive relationships of lines within and between sites: measured by the site mean co-ancestry matrix, expressed as a percentage of the mean co-ancestry for site 1

Site	Initial				Final			
	1	2	3	4	1	2	3	4
1	100	102	98	96	100	99	98	100
2		105	99	97		99	98	99
3			97	96			98	99
4				97				100

Sparse MET designs

Concluding remarks

- Introduced a new class of designs for METs
- These designs have been implemented in several large breeding programmes in Australia for several years
- **Overwhelmingly** positive reaction thus far
- Computational and theoretical problems remain
- New β -version of od has implemented a modification of the updating formulae in Coombes (2002): early results suggest up to 10 fold reduction in computing time for moderate to large problems
- Need to extend arguments and options in od to facilitate more flexible syntax for specification of objective function using the prediction approach