

Selection for breeding value and variety release in multi-environment trials: exploiting multiplicative mixed models

Brian Cullis

`brian.cullis@industry.nsw.gov.au`

Biometrics

NSW Industry and Investment

Multiplicative mixed models for variety selection

Collaborations and Acknowledgements

- This presentation is joint work with Alison Smith (NSW Industry & Investment) and Cameron Beeck, Wallace Cowling (CBWA). Also acknowledge the work of Helena Oakey in this research.
- Thanks to Grains Research and Development Corporation for financial support.

The Curse of the Were-Breeders

Starring: The Breeders and the Statisticians

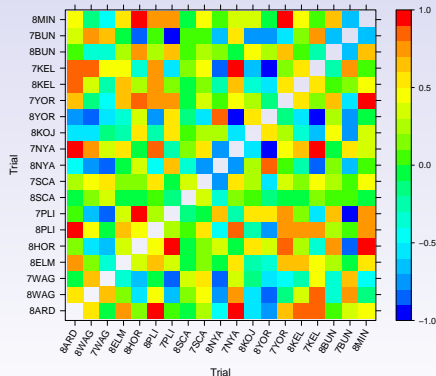
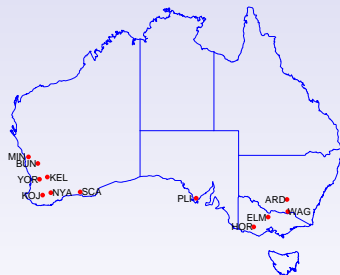


Multiplicative mixed model analysis of example data

A wealth of information: how to exploit it?

19 canola trials at 12
locations in 2007/8

Breeding values:
correlations between trials



Example: trial details

- 19 trials at 12 locations over 2 years. P-rep designs.
Rectangular (row by column) layout of plots.

Loc	State	Entries	2007		2008		
			%1rep	Mn Yld	Entries	%1rep	Mn Yld
ARD	NSW				154	36	1.31
BUN	WA	213	65	0.21	153	36	1.90
ELM	VIC				154	36	0.27
HOR	VIC				154	36	0.90
KEL	WA	232	68	1.32	153	36	1.14
KOJ	WA				154	36	1.53
MIN	WA				153	36	2.46
NYA	WA	245	73	0.84	153	36	1.38
PLI	SA	252	76	1.03	154	36	0.88
SCA	WA	254	75	0.90	154	36	0.94
WAG	NSW	220	66	1.06	154	36	0.59
YOR	WA	260	78	2.12	154	36	1.64

Example: entry details

- Total of $m_o = 332$ entries: 260 in 2007 and 154 in 2008 (82 in common between years). 2008 balanced; 2007 nearly so.
- Pedigree information on all entries in trials plus $m_p = 246$ parental lines
- Majority of entries fully inbred

The MET analysis

Without pedigree information

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{g_o}\mathbf{u}_{g_o} + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e}$$

- \mathbf{y} is grain yield data ordered as rows within columns within trials. Length $n = 5244$.
- $\boldsymbol{\tau}$ is vector of fixed effects including trial main effects and effects for modelling trend
- \mathbf{u}_p is vector of “peripheral” random effects for design factors and modelling trend

The MET analysis

Without pedigree information

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{g_o}\mathbf{u}_{g_o} + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e}$$

- \mathbf{u}_{g_o} is vector of entry effects for each trial (ordered as trials within entries).
- Assume $\text{var}(\mathbf{u}_{g_o}) = \mathbf{I}_{m_o} \otimes \mathbf{G}_g$
 - \mathbf{G}_g is $t \times t$ genetic variance matrix
 - diagonal elements are genetic variances for individual trials
 - off-diagonal elements are genetic covariances between pairs of trials
- \mathbf{e} is vector of plot errors (ordered as for \mathbf{y}).
- Assume $\text{var}(\mathbf{e}) = \text{diag}(\mathbf{R}_j)$ for $j = 1 \dots t$

The MET analysis

Non-genetic modelling

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{g_o}\mathbf{u}_{g_o} + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e}$$

- Use “simple” variance model for genetic effects analogous to separate analysis of each trial: $\mathbf{G}_g = \text{diag}(\sigma_{g_j}^2)$
- Fit major blocking factors as random effects (ie. in \mathbf{u}_p)
- Fit separable (row by column) AR1 \times AR1 covariance model for plot errors for each trial
- Use a range of diagnostics to assess adequacy of AR1 \times AR1

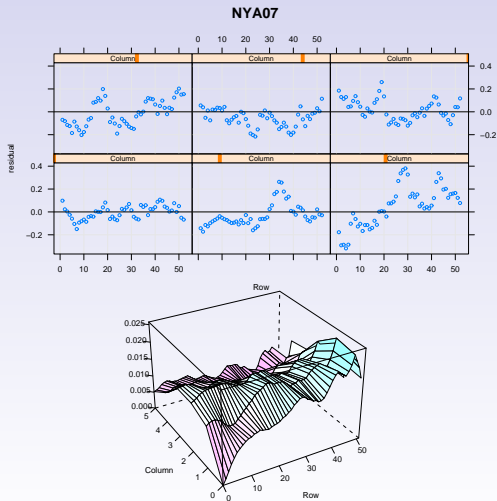
Example

Non-genetic modelling

- Extra peripheral effects added to model:
 - Fixed effect for linear regression over row number (7 trials)
 - Fixed effect for linear regression over column number (1 trial)
 - Random column effects (2 trials)
 - Random row effects (3 trials)
- Spatial autocorrelations:
 - Column correlations: ave=0.22
 - Row correlations: ave=0.62

Residual diagnostics for NYA in 2007

Spatial correlations = 0.21 (Columns) and 0.77 (Rows)



The MET analysis

With pedigree information

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e}$$

- \mathbf{u}_g is vector of entry effects for each trial expanded to include extra entries in pedigree but not grown in trials.
- We have extra $m_p = 246$ entries in pedigree so total of $m = 332 + 246 = 578$ entries
- So $\mathbf{u}_g = (\mathbf{u}_{g_p}', \mathbf{u}_{g_o}')'$
- $\mathbf{Z}_g = [\mathbf{0} \ \mathbf{Z}_{g_o}]$ where $\mathbf{0}$ is $n \times m_p$ matrix of zeros since no data for m_p entries

The MET analysis

With pedigree information

$$\begin{aligned}
 \mathbf{y} &= \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_p\mathbf{u}_p + \mathbf{e} \\
 \mathbf{u}_g &= \mathbf{u}_a + \mathbf{u}_{\bar{a}}
 \end{aligned}$$

- Total genetic effects have been partitioned into additive \mathbf{u}_a and non-additive $\mathbf{u}_{\bar{a}}$
- Assume $\text{var}(\mathbf{u}_a) = \mathbf{A} \otimes \mathbf{G}_a$
 - \mathbf{A} is $m \times m$ known relationship matrix ($m = 578$)
 - \mathbf{G}_a is $t \times t$ additive genetic variance matrix ($t = 19$)
- Assume $\text{var}(\mathbf{u}_{\bar{a}}) = \mathbf{I}_m \otimes \mathbf{G}_{\bar{a}}$
 - $\mathbf{G}_{\bar{a}}$ is $t \times t$ non-additive genetic variance matrix ($t = 19$)

The pedigree MET analysis

FA model for additive effects

$$\mathbf{u}_a = (\mathbf{I}_m \otimes \mathbf{\Lambda}_a) \mathbf{f}_a + \boldsymbol{\delta}_a$$

- Factor Analytic (FA) model for additive effects
 - $\mathbf{\Lambda}_a$ is $t \times k_a$ matrix of trial loadings
 - \mathbf{f}_a is vector (length mk_a) of entry scores
 - $\boldsymbol{\delta}_a$ is vector of lack of fit effects
 - k_a is number of factors in FA model
- A multiplicative model of trial loadings (covariates) and entry scores (slopes). Both loadings and scores to be estimated from data

The pedigree MET analysis

FA model for additive effects

- Variance assumptions for slopes and lack of fit effects:

$$\text{var}(\mathbf{f}_a) = \mathbf{A} \otimes \mathbf{I}_{k_a}$$

$$\text{var}(\boldsymbol{\delta}_a) = \mathbf{A} \otimes \boldsymbol{\Psi}_a$$

- $\boldsymbol{\Psi}_a$ is $t \times t$ diagonal matrix of “specific variances”
- Note that slopes (and lack of fit effects) are correlated between entries via \mathbf{A}
- These assumptions lead to

$$\text{var}(\mathbf{u}_a) = \mathbf{A} \otimes (\boldsymbol{\Lambda}_a \boldsymbol{\Lambda}_a' + \boldsymbol{\Psi}_a)$$

The pedigree MET analysis

FA model for non-additive effects

$$\mathbf{u}_{\bar{a}} = (\mathbf{I}_m \otimes \mathbf{\Lambda}_{\bar{a}}) \mathbf{f}_{\bar{a}} + \boldsymbol{\delta}_{\bar{a}}$$

- Factor Analytic (FA) model for non-additive effects. Number of factors $k_{\bar{a}}$. Now we assume

$$\text{var}(\mathbf{f}_{\bar{a}}) = \mathbf{I}_m \otimes \mathbf{I}_{k_{\bar{a}}}$$

$$\text{var}(\boldsymbol{\delta}_{\bar{a}}) = \mathbf{I}_m \otimes \boldsymbol{\Psi}_{\bar{a}}$$

- These variance assumptions lead to

$$\text{var}(\mathbf{u}_{\bar{a}}) = \mathbf{I}_m \otimes (\mathbf{\Lambda}_{\bar{a}} \mathbf{\Lambda}_{\bar{a}}' + \boldsymbol{\Psi}_{\bar{a}})$$

Example

Summary of variance modelling for the genetic effects

Models without pedigree

Model for G_g	residual logl	parameters in G_g	AIC
Diag	6491	19	-12943
FA1	6755	38	-13434
FA2	6935	56	-13757
FA3	7008	73	-13870
FA4	7072	89	-13966
FA5	7109	104	-14010

Models with pedigree

Model for		residual logl	parameters in $G_\alpha + G_{\bar{\alpha}}$	AIC
G_α	$G_{\bar{\alpha}}$			
Diag	Diag	6619	38	-13162
FA1	FA1	7043	76	-13933
FA2	FA2	7174	112	-14123
FA3	FA3	7230	146	-14168

Example

Genetic variance: additive as % total (additive + non-additive)

Loc	2007	2008
ARD		65
BUN	69	43
ELM		38
HOR		65
KEL	53	53
KOJ		84
MIN		13
NYA	62	56
PLI	71	54
SCA	17	48
WAG	40	45
YOR	37	26

Model interpretation

Genetic effects

- The additive effect (breeding value) of an entry is the component of the genetic effect that is heritable. The associated BLUP is the best prediction of the performance of the entry as a parent.
- The total genetic effect of an entry is the sum of the additive and non-additive effects. The associated sum of BLUPs is the best prediction of the commercial performance of the entry.
- The analysis provides a wealth of information to aid with these two types of selection
- Start with the additive effects . . .

FA model for additive effects

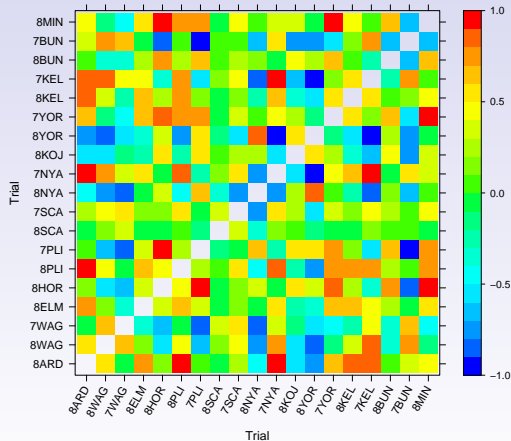
Genetic correlation matrix

$$\begin{aligned}
 G_a &= \Lambda_a \Lambda_a' + \Psi_a \\
 &= D_a C_a D_a \\
 &= D_a \left(\Lambda_a^{(c)} \Lambda_a^{(c)'} + \Psi_a^{(c)} \right) D_a
 \end{aligned}$$

- $D_a = \text{diag}(\sqrt{G_{a_{ii}}})$, $G_{a_{ii}}$ are the diagonal elements of G_a
- C_a is additive genetic correlation matrix
- $\Lambda_a^{(c)} = D_a \Lambda_a$ are loadings on correlation scale
- Additive genetic correlations can be displayed graphically
 - traditional approach uses bi-plots of $\Lambda_a^{(c)}$
 - new approach uses heatmap of C_a

FA model for additive effects

Heatmap of genetic correlations (Highway 1 order)



FA model for additive effects

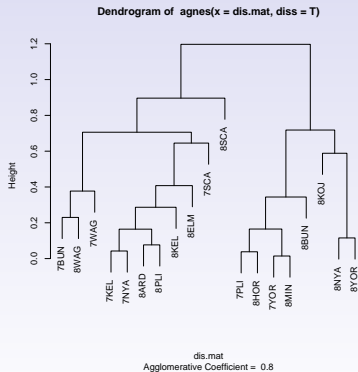
Genetic correlation matrix

- Heatmap can be ordered in sensible ways
- Prior information: geographic (eg. Highway 1) or
- Use data: eg. cluster trials on basis of C_a then order accordingly

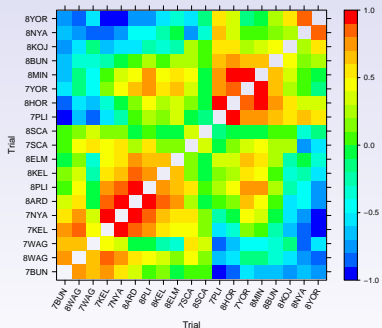
FA model for additive effects

Genetic correlation matrix

Clustering on C_a



Ordered heatmap



FA model for additive effects

Loadings

- When $k_a > 1$ loadings are not unique: can be rotated without changing G_a .
- To aid with interpretation (and do classical bi-plots) we rotate to PCA solution (so first loading accounts for maximum variance in G_a etc. etc.)
- For bi-plots it is crucial to know how much additive genetic variance is explained for each trial for each factor, so . . .

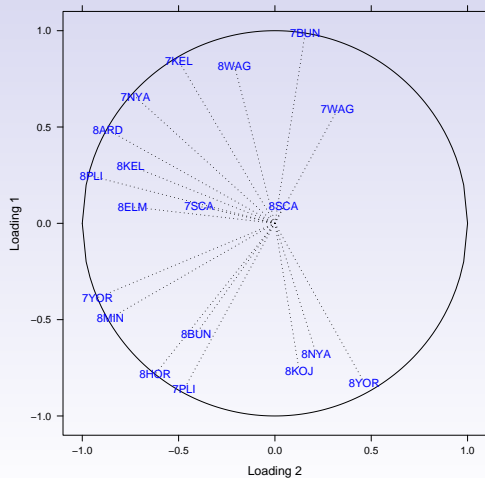
FA model for additive effects

Loadings and % additive variance accounted for

	Loadings $\times 1000$			% var			Variance $\times 10^4$	
	load 1	load 2	load 3	load 1	load 1+2	load 1+2+3	Additive	Non additive
7BUN	65	10	-1	97	99	100	84	37
7KEL	50	-30	13	71	96	100	70	61
7NYA	50	-55	16	43	95	100	111	69
7PLI	-108	-59	-25	74	96	100	306	127
7SCA	4	-18	43	1	16	100	42	204
7WAG	38	21	47	35	46	100	80	122
7YOR	-32	-75	5	15	100	100	130	225
8ARD	46	-83	-11	24	99	100	176	94
8BUN	-43	-31	15	33	50	53	109	147
8ELM	3	-28	-7	1	56	58	29	46
8HOR	-66	-53	2	61	100	100	140	76
8KEL	20	-51	-13	9	65	68	89	78
8KOJ	-136	23	72	59	61	76	617	114
8MIN	-26	-45	9	24	97	100	54	362
8NYA	-57	18	-60	46	51	100	139	109
8PLI	15	-60	11	6	97	100	77	67
8SCA	5	2	21	1	1	16	55	60
8WAG	65	-17	23	67	71	79	123	147
8YOR	-40	22	-15	68	89	100	45	131
Ave				39	73	87		

FA model for additive effects

Bi-plot for loadings 1 and 2 (correlation scale)



FA model for additive effects

BLUPs of effects

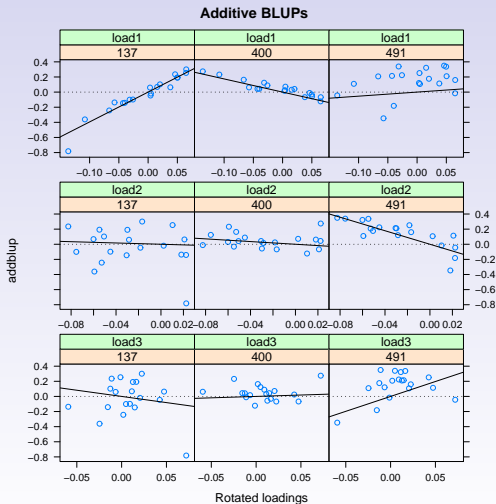
- Analysis provides BLUPs of u_a : a prediction for *each* entry for *each* trial
- Parental lines (not grown anywhere in this MET): BLUPs are obtained from data on lines that *were* grown via correlations in A
- FA model with 3 factors so additive effect for line i in trial j given by

$$u_{a_{ij}} = \lambda_{a_{1j}} f_{a_{1i}} + \lambda_{a_{2j}} f_{a_{2i}} + \lambda_{a_{3j}} f_{a_{3i}} + \delta_{a_{ij}}$$

- Slopes show entry response to individual covariates
- If covariates can be interpreted biologically then this helps explain genotype by trial interaction

FA model for additive effects

Three entries with differential response to loadings



FA model for additive effects

Reliability of BLUPs

- Measure reliability in terms of (squared) correlation between true and predicted effect
- Denote BLUP for entry i in trial j by $\tilde{u}_{a_{ij}}$. Reliability of this BLUP:

$$r_{ij} = \text{COR} (u_{a_{ij}}, \tilde{u}_{a_{ij}})^2$$

- A model based estimate can be obtained as

$$r_{ij} = 1 - C^{ij,ij} / (A_{ii} G_{a_{jj}})$$

- A_{ii} is diagonal element of relationship matrix corresponding to entry i
- $G_{a_{jj}}$ is diagonal element of estimated additive genetic variance matrix corresponding to trial j
- $C^{ij,ij}$ is diagonal element of inverse of mixed model coefficient matrix corresponding to entry i in trial j

FA model for non-additive effects

Information as for additive effects

- Estimated genetic correlation matrix (heatmaps)
- Estimated loadings
- BLUPs of $u_{\bar{a}}$: a prediction for each entry for each trial
- Reliability of BLUPs
- Regression plots of BLUPs vs loadings

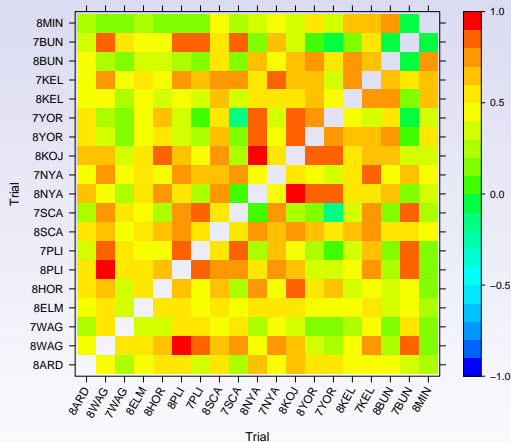
FA model for non-additive effects

Loadings and % additive variance accounted for

	Loadings \times 1000			% var		
	load 1	load 2	load 3	load 1	load 1+2	load 1+2+3
7BUN	33	50	-13	28	95	100
7KEL	72	15	27	84	88	100
7NYA	64	26	11	60	70	71
7PLI	71	76	-5	40	85	85
7SCA	81	112	36	33	94	100
7WAG	49	45	-8	20	37	37
7YOR	97	-77	-48	42	68	79
8ARD	63	-10	-22	42	43	48
8BUN	88	-50	40	53	70	80
8ELM	43	8	-10	41	42	44
8HOR	61	-5	-44	48	48	74
8KEL	69	-16	30	61	64	75
8KOJ	94	-25	-43	78	84	100
8MIN	110	-52	101	34	41	69
8NYA	85	-45	-41	66	85	100
8PLI	64	42	-21	61	88	94
8SCA	68	4	-5	76	76	77
8WAG	90	67	-28	55	85	91
8YOR	91	-50	-3	63	82	82
Ave				52	71	79

FA model for non-additive effects

Heatmap of genetic correlations (Highway 1 order)



Multiplicative mixed model with pedigree information

Selection of lines

- Analysis provides BLUPs of u_a and $u_{\bar{a}}$ thence
$$u_g = u_a + u_{\bar{a}}$$
- Use BLUPs of u_a to select entries for use as parents
- Use BLUPs of u_g to select entries for commercial use
- Use BLUPs of $u_{\bar{a}}$ in their own right?

Multiplicative mixed model with pedigree information

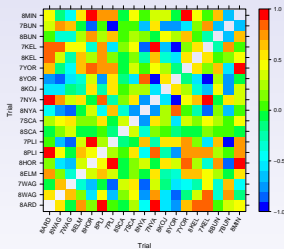
Selection of lines

- Analysis provides BLUPs for *each* entry for *each* trial
- How to sensibly combine across trials for selection?
- A difficult question: some suggestions
 - Do not blindly average BLUPs across all trials
 - Use estimated genetic correlation matrices to identify trials with similar performance (clustering) then average BLUPs across trials in each cluster

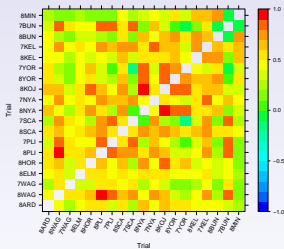
Multiplicative mixed model analysis of example

Genetic correlation matrices

Additive



Non-additive



Total

