

Divide and Conquer: partitioning genetic variety effects in Multi-Environment-Trials (METs)

Helena Oakey, Ari Verbyla,
Brian Cullis, Xianming Wei
and Wayne Pitchford



Outline

- Background
 - Current approach in METs analysis
 - classical quantitative genetics model
 - $\mathbf{P} = \mathbf{G} + \mathbf{E}$
 - Phenotype is result of genetic and environmental influences
- Partitioning genetic component
 - $\mathbf{G} = \mathbf{A} + (\mathbf{D} + \mathbf{I})$
 - WHY?**
 - HOW?**
- Fitting D
- New approach incorporating partitioning genetic component
- Example

Classical Quantitative Genetics Model

$$\underbrace{y}_{P} \text{ }_{jkr} = \underbrace{\gamma}_{G} \text{ }_{jk} + \underbrace{\epsilon}_{E} \text{ }_{jkr}$$

Phenotype Genotype Environment



PEG

y_{jkr} is the response (yield) of the j th variety in k th environment with r th replicate

γ_{jk} is the variety mean of the j th variety in k th environment

ϵ_{jkr} is the non-genetic influences (see Gilmour et al 1997, Smith et al 2005)

Modelling of variety/genetic means γ

$$\underbrace{\gamma}_{\text{Variety}}_{\text{Mean}}_{jk} = \underbrace{\sigma}_k_{\text{Site}}_{\text{Mean}} + \underbrace{g}_{\text{Variety}}_{\text{by Site}}_{\text{Effect}}_{jk}$$

If aim is selection of varieties (Smith et. al. , 2005)

- variety effects j th variety in each of k th environment **RANDOM**
- Site/Environment effects **FIXED**

Summary Current Approach

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{g} + \boldsymbol{\varepsilon}$$

$\boldsymbol{\tau}$ is the (p x 1) fixed effect containing site means

\mathbf{X} is the (n x p) design matrix

\mathbf{Z}_g is the (n x mp) design matrix

\mathbf{g} is the (mp x 1) vector of m random variety effects in each of p environments

$\boldsymbol{\varepsilon}$ is the non-genetic effects

Variance of \mathbf{g} has a separable form

$$\text{var}(\mathbf{g}) = \mathbf{G}_e \otimes \mathbf{G}_v$$

\mathbf{G}_e is the $(p \times p)$ genetic variance matrix for environments/sites

diagonal terms are site variances and off-diagonal terms are covariances between sites

\mathbf{G}_v is the $(m \times m)$ genetic variance matrix for varieties

Current Approach

- predicts Total genetic effect \mathbf{g} or overall performance
- Assumes independence of lines

$$\text{var}(\mathbf{g}) = \mathbf{G}_e \otimes \mathbf{I}_m$$

$\mathbf{G}_v = \mathbf{I}_m$
the genetic
variance matrix
for varieties

Partitioning genetic effect



GAnDI

$$\mathbf{g} = \mathbf{a} + (\mathbf{d} + \mathbf{i})$$

$\mathbf{a}^{(mp \times 1)}$ \Leftrightarrow additive genetic effects

$\mathbf{d}^{(mp \times 1)}$ \Leftrightarrow dominance genetic effect

$\mathbf{i}^{(mp \times 1)}$ \Leftrightarrow residual non-additive
genetic effects

Why might we want to partition g ?

Aim of plant breeding trial select varieties with best performance and also varieties that can be used in future crosses

Currently plant breeders use total genetic effect from the Standard or current approach to assess BOTH aims

Why might we want to partition g ?

- Total genetic effects (g) include non-additive effects which are not necessarily passed on to offspring, however are important in assessing variety performance for commercial release.
- Additive genetic effects are passed on to offspring and are referred to as “breeding values” so give an indication of the potential of variety as a parent (in future cross)

Why might we want to partition \mathbf{g} ?

- Partitioning \mathbf{g}
 - Achieving both aims more accurately
 - \mathbf{g} reflects underlying genetic mechanism
 - If \mathbf{g} modelled correctly less biased estimates with lower prediction error variance

How do we partition \mathbf{g} ?

- Can't quantify additive and non-additive effects directly as the allelic information is unknown
- However if the pedigree of lines is available then we can determine the additive and dominance relationships between lines relative to a “base population (ancestors of lines)” and form
- Additive relationship matrix \mathbf{A} (Henderson, 1976)
- Dominance relationship matrix \mathbf{D}
- Use \mathbf{A} and \mathbf{D} to determine \mathbf{a} and \mathbf{d}

Additive Relationship matrix \mathbf{A}

- Rules to create \mathbf{A} and its **inverse** were developed by Henderson (1976)

$$\text{var}(\mathbf{a}) = \mathbf{G}_a \otimes \mathbf{A} \quad \text{Partitioned}$$

$$\text{var}(\mathbf{g}) = \mathbf{G}_e \otimes \mathbf{I}_m \quad \text{Standard}$$

Dominance Relationship matrix

- **D** has been derived for animal breeding programs which assume no inbreeding (Cockerham, 1954)
- Rules need to be extended to include varying levels of inbreeding (Verbyla and Oakey, 2006)

$$\text{var}(d) = \mathbf{G}_d \otimes \mathbf{D} \quad \text{Partitioned}$$

$$\text{var}(g) = \mathbf{G}_e \otimes \mathbf{I}_m \quad \text{Standard}$$

Modelling \mathbf{d}

- \mathbf{D} can be derived from the terms of \mathbf{A}
- \mathbf{A}^{-1} and \mathbf{D}^{-1} are needed for the mixed model equations
- \mathbf{A}^{-1} can be determined directly
- However no obvious way of developing \mathbf{D}^{-1} from \mathbf{A}^{-1} directly
- If \mathbf{D} large computational limitations fitting dominance effect

Family Dominance Relationships

- Use family dominance relationships to reduce the computational limitations
- dominance relationship between two individuals is defined by the relationships between their parents.
- Individuals from the same family (i.e. same parents) therefore share the same dominance relationships.
- If a pedigree contains many individuals from the same family, the dominance relationship between these individuals can be summarized in a reduced form (Hoeschele and VanRaden, 1991)

Partitioning \mathbf{d}

- Hoescele and VanRaden (1991) considers partitioning dominance effect $\mathbf{d}^{(mp \times 1)}$ into two components;
 - between family dominance effects $\mathbf{d}_b^{(vp \times 1)}$
v=no. family \ll m=no. individuals
 - within family dominance effects $\mathbf{d}_w^{(mp \times 1)}$
 - With corresponding dominance matrices \mathbf{D}_b and \mathbf{D}_w
 - Again rules needed to be derived for varying levels of inbreeding (Oakey et al, 2006)

Partitioning \mathbf{g}

$$\mathbf{g} = \mathbf{a} + (\mathbf{Z}_b \mathbf{d}_b + \mathbf{d}_w + \mathbf{i})$$

$$\text{var}(\mathbf{g}) = \mathbf{G}_a \otimes \mathbf{A} +$$

$$\mathbf{G}_d \otimes (\mathbf{Z}_b \mathbf{D}_b \mathbf{Z}_b^T + \mathbf{D}_w) +$$

$$\mathbf{G}_i \otimes \mathbf{I}_m$$

Partitioned
model

$$\text{var}(\mathbf{g}) = \mathbf{G}_e \otimes \mathbf{I}_m$$

Standard
model

WAKE UP



Divide and Conquer

PARTITION PARTITION PARTITION

$$\mathbf{y} = \mathbf{Z}_\gamma \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

PEG

$$\boldsymbol{\gamma} = \mathbf{X}_\tau \boldsymbol{\tau} + \mathbf{g}$$

FIXED+RANDOM

$$\mathbf{g} = \mathbf{a} + \mathbf{d} + \mathbf{i}$$

GAnDI

$$\mathbf{d} = \mathbf{Z}_b \mathbf{d}_b + \mathbf{d}_w$$

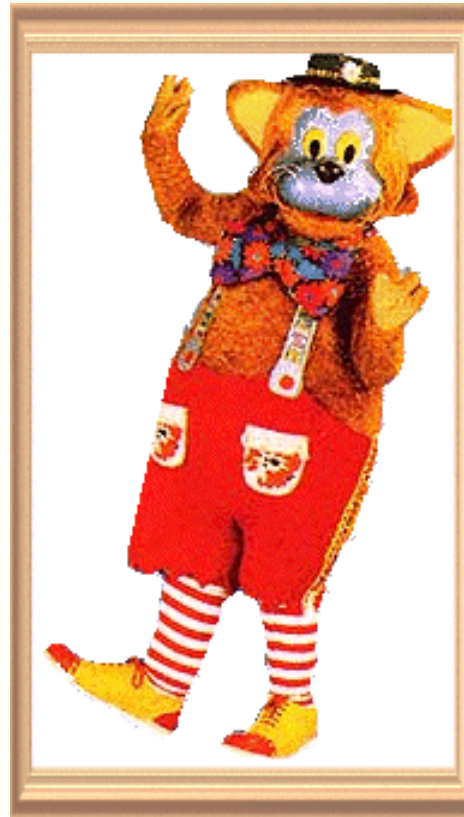
PARTITION

Example



- Sugarcane data from the joint sugar breeding program of BSES Ltd and CSIRO
- Response is Commercial Sugar Cane
- SIX environment
- TWO stage 2 (2002) CATs: Clonal Assessment Trials
 - 2267 varieties tested
- FOUR Stage 3 (2003) FATs: Final Assessment Trials
 - 105 varieties tested in (selected from the CATs)
- Pedigree information (mother, father) on all varieties and going back several generations

CATs FATs NOT FAT CAT



Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g \mathbf{a} + \mathbf{Z}_g \mathbf{d} + \mathbf{Z}_g \mathbf{i} + \boldsymbol{\varepsilon}$$

- Unable to fit \mathbf{d} for this example as \mathbf{D} as contained over 2 million pieces of information

- Partition \mathbf{d} and \mathbf{D}

$$\mathbf{d} = \mathbf{Z}_b \mathbf{d}_b + \mathbf{d}_w \quad \mathbf{D} = \mathbf{Z}_b^T \mathbf{D}_b \mathbf{Z}_b + \mathbf{D}_w$$

- In our example
 - m (number of individuals) = 2267
 - v (number of families) = 187
 - So reduced number of bits of information to max. 19845 and able to fit partitioned \mathbf{d}

Summary of the (some) of the models fitted

	Structure of site genetic variance matrix			q	AIC	Log-likeli
	G_a	G_d	G_i			
1			XFA2*	59	3179	-2469
2	DIAG	DIAG(1, 2, 3, 4)	DIAG(2, 5, 6)	54	3124	-2438
3	XFA1			53	3057	-2406
4	XFA2	XFA1(1, 2, 3, 4)	XFA1(2, 5, 6)	72	0	-858

AIC=Akaike Information Criteria = $-2\loglik+2q$ (lower better)

AIC=relative to model 4

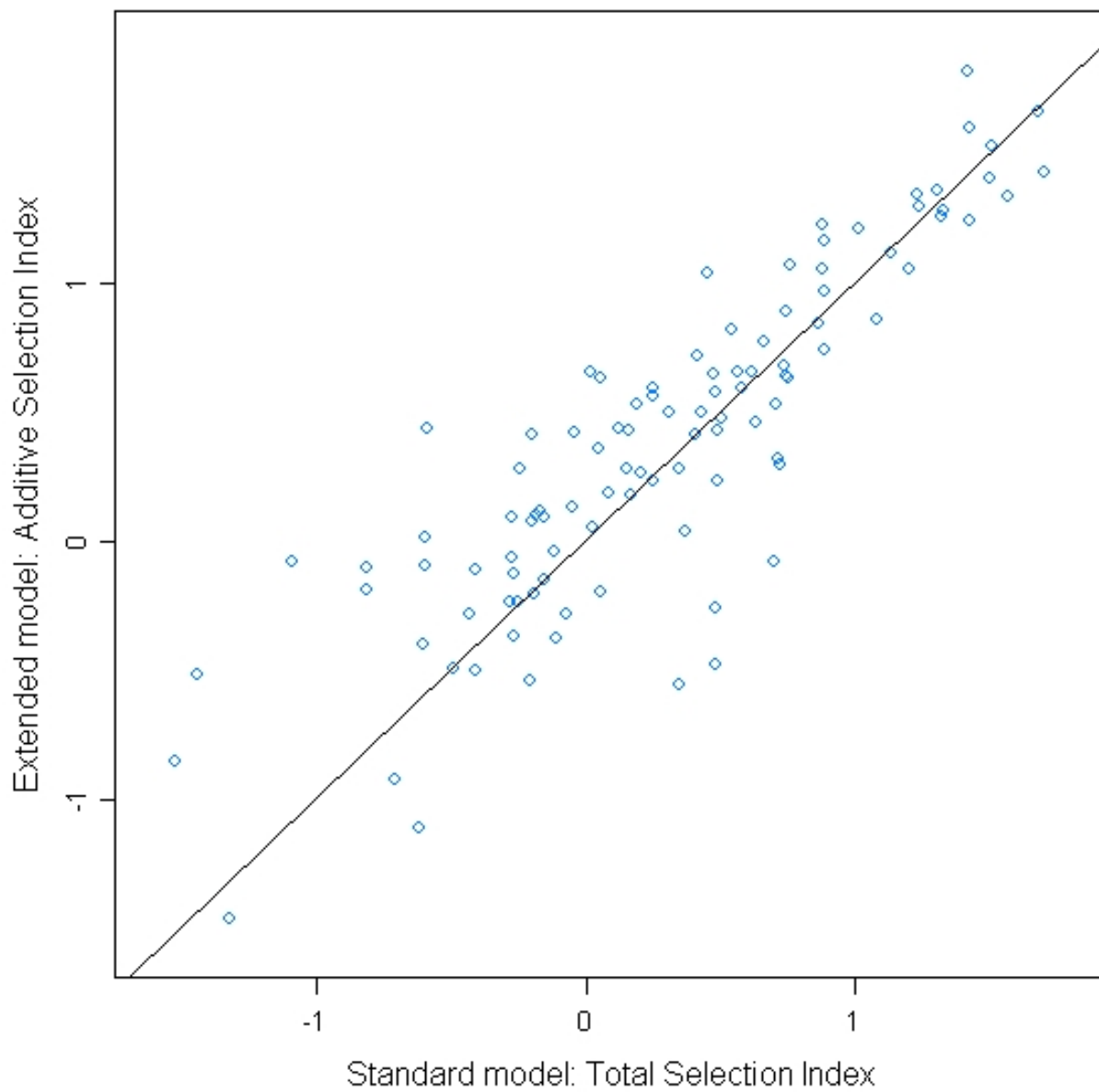
q =number of parameters

XFA=factor analytic structure (Smith et al, 2001)

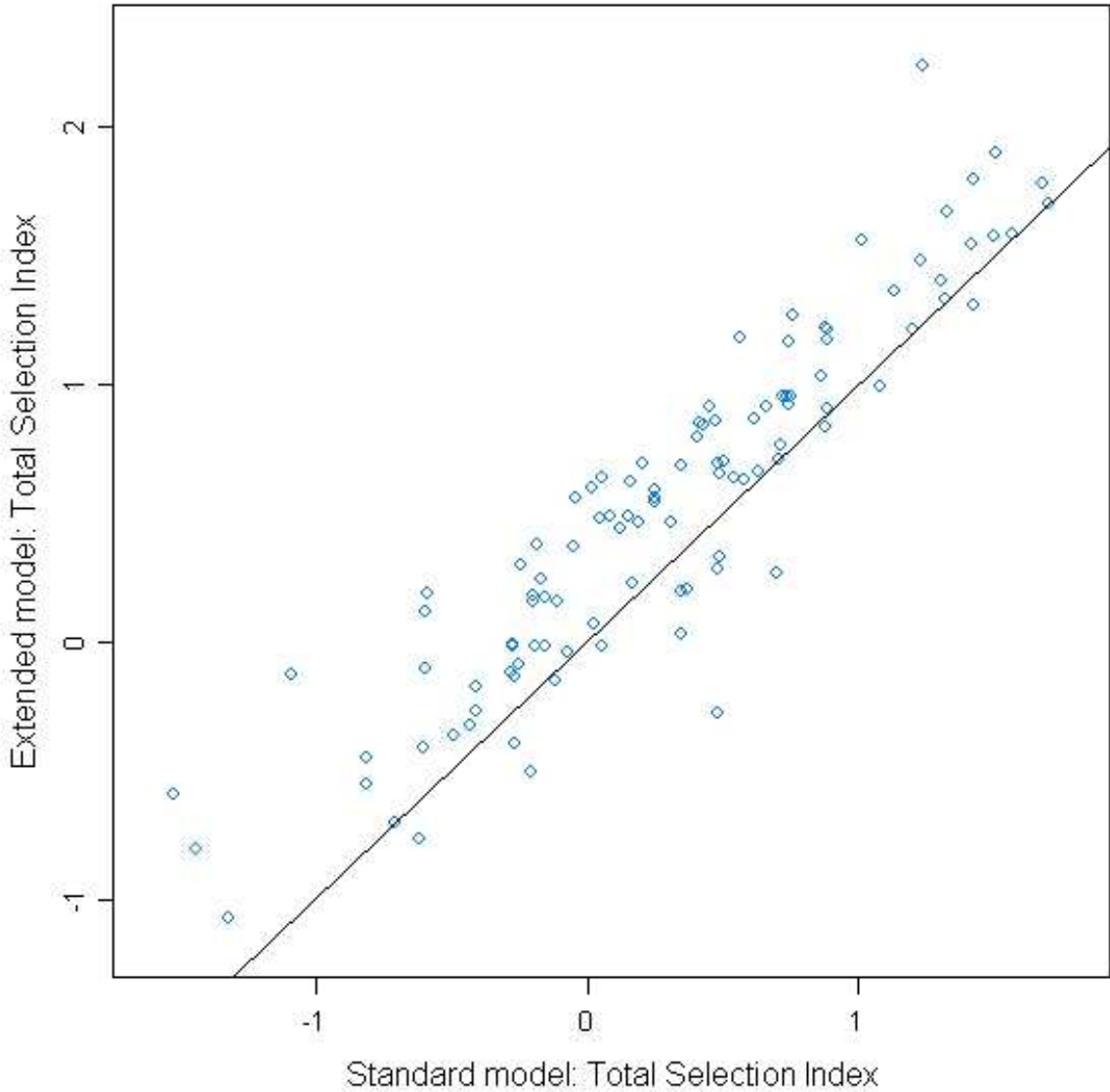
Loglikelihood=MAX (higher better)

Fitted in ASReml

CURRENT APPROACH **g** versus PARTIONED APPROACH **a**



CURRENT APPROACH g verses PARTIONED APPROACH g



Conclusions

Models underlying genetic structure

Better fitting models

Selection of

- best commercial lines (total genetic effect)
- best parents (additive effect)
- best combination of parents (dominance effects)

From single analysis of standard MET without resorting to specialised expt. such as Diallel crosses

Future work

- **A** matrix
 - Based on “average” relationships between individuals
- Develop a more accurate measure may be determined if have marker data
- Currently move towards using DARTs in plant breeding setting

Thanks

- GRDC providing scholarship
- Supervisors
 - Ari Verbyla
 - Brian Cullis
 - Wayne Pitchford



References

- Oakey, H., Verbyla, A., Cullis, B., Wei, X. and Pitchford, W. (2006). *Joint modelling of additive and non-additive (genetic line) effects in multi-environment trials*. **Theoretical and Applied Genetics**. Accepted.
- Oakey, H., Verbyla, A., Pitchford, W., Cullis, B. and Kuchel, H. (2006). *Joint modelling of additive and non-additive genetic line effects in single field trials*. **Theoretical and Applied Genetics**. **113**, 809-819.
- Verbyla A. and Oakey, H (2006) *The variance-covariance matrix for relatives undergoing mendelian sampling and inbreeding*. Research Report 4, BiometricsSA, University of Adelaide, Sept. 7th, 11 pages.
- Falconer, D.S. and Mackay, T.F.C. (1996). *Introduction to quantitative genetics* 4th edn. Longman Group Ltd.
- Gilmour, A. R., Cullis, B. R., Gogel, B., Welham, S. J., and Thompson, R. (2005). *ASReml, user Guide. release 2.0*. Hemel Hempstead, UK: VSN International Ltd.
- Gilmour, A. R., Cullis, B., and Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics* **2**, 269–293.
- Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. *Biometrics* **32**, 69–83.
- Hoescele, I. and VanRaden P.M. (1991) Rapid inversion of dominance relationship matrices for noninbred populations by including sire by dam subclass effects. *Journal of Dairy Science* **74**:557-569.
- Smith, A.B., Cullis, B.R. and Thompson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *Journal of Agricultural Science* **143**, 1–14.