Australasian Biometrics Conference 2023

Digging up the dirt on competition in multi-environment agricultural field trials

Alison Kelly Tolera Keno Ian Godwin Emma Mace David Jordan

November 30, 2023

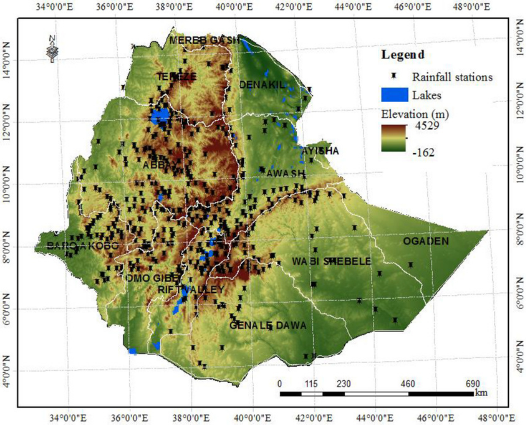
EIAR
Ethiopian Institute of Agricultural Research

The University of Queensland
Australia

QAAFI
Queensland Alliance for Agriculture and Food Innovation

Modernizing Ethiopian Research on Crop Improvement

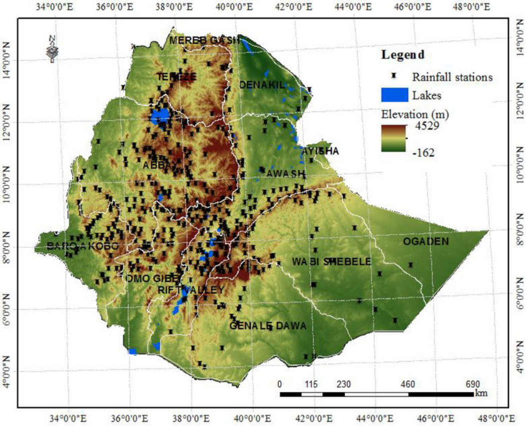
Project PhD study



Modernizing Ethiopian Research on Crop Improvement
The Environment

Environment

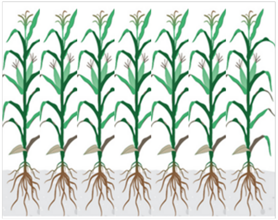
Mt Cook 3724m



Modernizing Ethiopian Research on Crop Improvement

Prediction into the target population of environments

Breeding trial



Farmers' fields

Genotype

- The aim of plant breeding is to select the best candidate varieties from small plot field trials and predict their performance in farmers' fields.

Competition in Multi-Environment Trials (METs)

Motivation

- The nature of agriculture in many developing countries is based on small-holder farms.

- The area of land available for experimental trials on agricultural research stations is often also limited in scale.

Competition in METs

Motivation

- Most farming and field research operations are conducted manually.
- In a hybrid plant breeding program there may also be limited seed available.
- Underlying all of these physical constraints is cost of operations.
- All of these constraints limit testing to small plots sizes, and maize is often tested in single row plots.

Competition in individual trials

Statistical considerations

- It is well established that spatial dependence exists between small plots in agricultural field trials
  (Wilkinson et al. 1983; Williams, 1986; Gleeson and Cullis, 1987; Gilmour et al. 1997).

- It is also well documented that there may be interference from adjacent plots in field trials (Pearce, 1957; Besag and Kempton, 1986; Goldringer, Brabant and Kempton 1994).

- Sometimes field operations are adjusted to minimise these edge effects, and sometimes they are ignored due to their expected small impact eg in larger plot sizes.

- When plots sizes become small, the interference effects can be important, as is likely the case in single-row plots.

Competition in individual trials

Literature

Genotypic interference model (Pearce 1957)

- The yield of each plot is influenced by the treatments (typically genotype) tested on that plot, as well as the treatments tested on the neighbouring plots.

Phenotypic interference model (Besag and Kempton 1986)

- Random treatment interference plus competition from the yield of neighbouring plots (ie residual errors).

Producer-Competitor model (Goldringer, Brabant and Kempton 1994)

- The treatment effects on plot i are

$$g_{(i)} + c_{(i-1)} + c_{(i+1)}$$

  where $g_{(i)}$ is the treatment effect on plot i and $c_{(i-1)}$ and $c_{(i+1)}$ are competitor effects from treatments in neighbouring plots $i-1$ and $i+1$ on either side.

Competition in individual trials

Recent developments

A model with genotypic and phenotypic interference (Stringer et al. 2011)

- A random treatment interference model (R-TIM)
- A residual correlation model capturing spatial variability and competition at the plot level
  - Spatial variability induces positive correlation between neighbouring plots
  - Competition typically induces negative correlation between neighbouring plots

Competition in METs

The model extension to METs

We developed a linear mixed model for multi-environment trial data, modelling genotype by environment interaction (GxE) for direct and neighbour effects (Keno 2023).

I will present

- the linear mixed model formulation for competition across multiple trials
- how we modelled spatial variability and competition at the plot level on a trial basis
- estimation of variance parameters in the model
- some thoughts on the impact of competition on GxE

The MET model for competition
The Linear Mixed Model

The general form of the linear mixed model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{\mathrm{g}}\mathbf{u}_{\mathrm{g}} + \mathbf{Z}_{\mathrm{o}}\mathbf{u}_{\mathrm{o}} + \mathbf{e}$$

where
$\mathbf{y}$ is the $(n \times 1)$ vector of yield response measured on n observations,
$\boldsymbol{\tau}$ is a $(t \times 1)$ vector of fixed effects with design matrix $\mathbf{X}^{(n \times t)}$
$\mathbf{u}_{\mathrm{g}}$ is a $(2mp \times 1)$ vector of direct and neighbour genetic effects for m genotypes in p trials, with design matrix $\mathbf{Z}_{\mathrm{g}}{}^{(n \times 2mp)}$
$\mathbf{u}_{\mathrm{o}}$ is a $(b \times 1)$ vector of random effects with design matrix $\mathbf{Z}_{\mathrm{o}}{}^{(n \times b)}$
$\mathbf{e}$ is the $(n \times 1)$ vector of residual errors.

The joint distribution of $\mathbf{u}_{\mathrm{g}}$, $\mathbf{u}_{\mathrm{o}}$ and $\mathbf{e}$ is

$$\begin{bmatrix} \mathbf{u}_{\mathrm{g}} \\ \mathbf{u}_{\mathrm{o}} \\ \mathbf{e} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G}_{\mathrm{g}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{\mathrm{o}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

The MET model for competition
Direct and neighbour effects

- Field trial j is comprised of n_j plots arranged in a rectangular array across c_j columns by r_j rows, and the data is assumed to be ordered as rows within columns within trials, $n = \sum_{j=1}^{p} n_j$

- The random genetic effects are partitioned into terms for direct and neighbour (competitor) effects

$$\mathbf{u}_{\mathrm{g}} = (\mathbf{u}_{\mathrm{d}}', \mathbf{u}_{\mathrm{n}}')'$$

$$\mathbf{Z}_{\mathrm{g}} = [\mathbf{Z}_{\mathrm{d}} \quad \mathbf{N}_{\mathrm{g}}\mathbf{Z}_{\mathrm{d}}]$$

where $\mathbf{Z}_{\mathrm{d}}$ is the usual $(n \times mp)$ design matrix for m genotypes in p trials, and $\mathbf{N}_{\mathrm{g}}$ is an $(n \times n)$ neighbour matrix,

$$\mathbf{N}_{\mathrm{g}} = \oplus_{j=1}^{p} \mathbf{N}_{\mathrm{g}_j}, \quad \text{and} \quad \mathbf{N}_{\mathrm{g}_j} = \mathbf{I}_{c_j} \otimes \mathbf{N}_{r_j}$$

where $\mathbf{N}_{r_j}$ is the first order neighbour incidence matrix indexed in the row direction for trial j.

The MET model for competition

Direct and neighbour effects

$$\mathbf{N}_{r_j}^{(r_j \times r_j)} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & & \\ 1 & 0 & 1 & 0 & 0 & & \\ 0 & 1 & 0 & 1 & 0 & & \\ 0 & 0 & 1 & 0 & 1 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & 1 & 0 & 1 \\ & & & & 0 & 1 & 0 \end{bmatrix}$$

- Then the genetic effects on plot i are from

$$u_{\mathrm{d}_{(i)}} + u_{\mathrm{n}_{(i-1)}} + u_{\mathrm{n}_{(i+1)}}$$

where i is indexed in the row direction, across the smallest plot dimension.

The MET model for competition

The G structure

The variance of the random genetic effects is

$$\mathrm{var}(\mathbf{u}_{\mathrm{g}}) = \mathbf{G}_{\mathrm{g}} \otimes \mathbf{I}_m$$

$$\mathbf{G}_{\mathrm{g}} = \begin{bmatrix} \mathbf{G}_{\mathrm{d}} & \\ \mathbf{G}_{\mathrm{dn}} & \mathbf{G}_{\mathrm{n}} \end{bmatrix}$$

$$\mathbf{G}_{\mathrm{d}} = \begin{bmatrix} \sigma^2_{\mathrm{d}_1} & & \\ \vdots & \ddots & \\ \sigma_{\mathrm{d}_1\mathrm{d}_p} & \cdots & \sigma^2_{\mathrm{d}_p} \end{bmatrix}$$

$$\mathbf{G}_{\mathrm{c}} = \begin{bmatrix} \sigma^2_{\mathrm{n}_1} & & \\ \vdots & \ddots & \\ \sigma_{\mathrm{n}_1\mathrm{n}_p} & \cdots & \sigma^2_{\mathrm{n}_p} \end{bmatrix}$$

$$\mathbf{G}_{\mathrm{dc}} = \begin{bmatrix} \sigma_{\mathrm{d}_1\mathrm{n}_1} & \cdots & \sigma_{\mathrm{d}_p\mathrm{n}_1} \\ \vdots & \ddots & \vdots \\ \sigma_{\mathrm{d}_1\mathrm{n}_p} & \cdots & \sigma_{\mathrm{d}_p\mathrm{n}_p} \end{bmatrix}$$

$\sigma^2_{\mathrm{d}_j}$ and $\sigma^2_{\mathrm{n}_j}$ are the variances of direct and neighbour genetic effects in trial j, respectively.

$\sigma_{\mathrm{d}_j\mathrm{d}_{j'}}$ and $\sigma_{\mathrm{n}_j\mathrm{n}_{j'}}$ are the covariances between direct effects and neighbour effects in trials j and j', respectively.

$\sigma_{\mathrm{d}_j\mathrm{n}_{j'}}$ is the covariance between direct and neighbour effects in trials j and j'.

The MET model for competition

The R structure

- The variance matrix for the residual errors is $\mathrm{var}(\mathbf{e}) = \oplus_{j=1}^{p} \mathbf{R}_j$

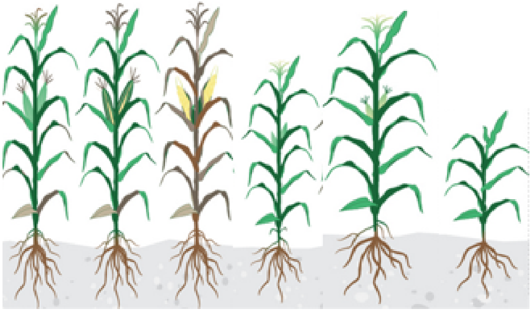
$$\mathbf{R}_j = \sigma^2_{\mathrm{e}_j}(\mathbf{\Sigma}_{c_j} \otimes \mathbf{\Sigma}_{r_j})$$

- In the competition model $\mathbf{\Sigma}_{r_j}$ can take the form of a constrained autoregressive process with two correlation parameters, α_{1_j} and α_{2_j} (Stringer et al., 2011)
- Note that when $\alpha_{2_j} = 0$ the residual correlation model reverts to the standard spatial autoregessive model (Gilmour et al. 1997).

Competition in METs

Motivation

- The motivation for this model extension comes from a MET series conducted in the Ethiopian mid-altitude maize breeding program.
- Due to limited land and seed availability, maize hybrids are grown in single row plots.
- This practical field arrangement is used for many maize breeding program trials conducted in sub-Saharan Africa.



Competition in METs

Field trial example

- A MET series was conducted in six trials over two years testing a total of 478 maize hybrids.

Trial	Columns	Rows	Replication	Genotypes	Mean yield (t/ha)
T1	8	44	1.45	242	2.88
T2	8	44	1.35	259	8.69
T3	8	56	1.87	240	8.98
T4	8	58	1.86	250	9.14
T5	8	56	1.87	240	8.12
T6	8	58	1.86	249	7.82

Competition in METs

Field trial example

- Genotype concurrence in this MET series is shown below.

Trial	T1	T2	T3	T4	T5	T6
T1	242					
T2	242	259				
T3	173	180	240			
T4	40	40	91	250		
T5	173	180	240	91	240	
T6	40	40	91	249	91	249

The MET model for competition

ASReml-R code

- The baseline linear mixed model for the competition MET can be fitted in ASReml-R.

```
asreml(fixed = yield ~ Site ,
random=~ str(Site:Genotype + Site:Gleft + and(Site:Gright) ,
                  ~diag(ns*2):id(Genotype)) +
                   at(Site):Replicate ,
residual =~ dsum(~ar1(Column):sar2(Row) | Site ) ,
equate.levels = c("Gleft","Gright"), data=maizeMET.df)
```

- The baseline model is extended to include more complex covariance structures for Site:(Direct+Neighbour) effects eg factor analytic
- The residual variance model is simplified for competition or spatial covariance alone, based on the AIC

Competition in METs

Field trial example

- A sequence of standard MET models were fitted, together with the competition MET model, in ASReml-R (Butler et al. 2019).

Model	Variance model Site:Genotype	Number of parameters	Change in AIC
1	diag(Site):Genotype	31	283
2	**fa(Site,2):Genotype**	40	85
3	diag((Site):(Direct+Neighbour))	37	257
4	**fa((Site,3):(Direct+Neighbour))**	64	**0**

Competition in METs

Field trial example

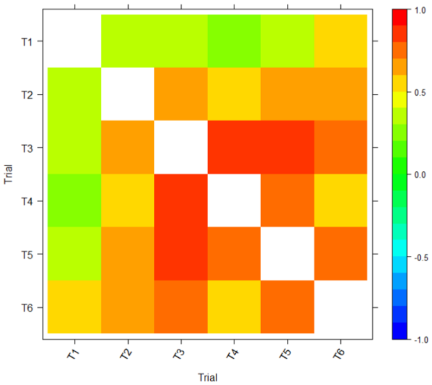
- Variance parameters were estimated using REML (Patterson and Thompson 1971).

Trial	Standard MET		Competition MET		
	σ^2_{g}	σ^2_{e}	σ^2_{d}	σ^2_{n}	σ^2_{e}
T1	0.06	1.13	0.03	0.14	0.90
T2	2.17	1.03	2.26	0.13	0.88
T3	1.96	1.54	2.25	0.10	1.20
T4	2.84	1.01	2.74	0.12	0.76
T5	1.17	1.38	1.07	0.10	1.25
T6	1.78	2.10	1.42	0.14	1.93

The MET model for competition

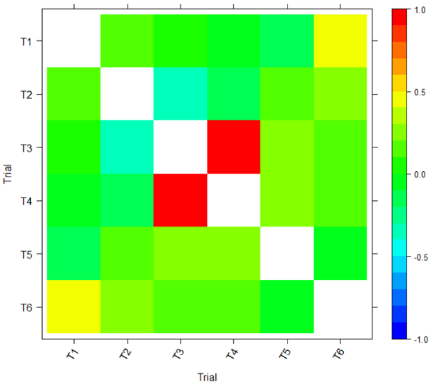
Direct and neighbour effects

Genetic correlations - Direct effects



Range from 0.24 to 0.86

Genetic correlations - Neighbour effects

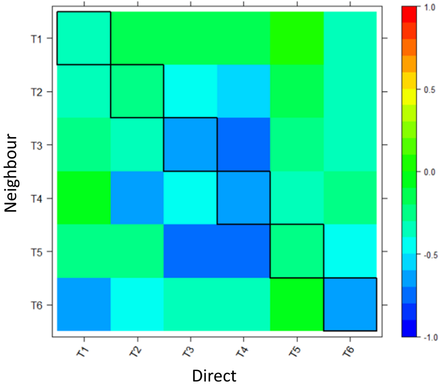


Range from -0.33 to 0.97

The MET model for competition
Direct and neighbour effects

In $\mathbf{G}_{\mathrm{dn}}$ the diagonal values show the negative correlation between direct and neighbour effects in the same Trial.

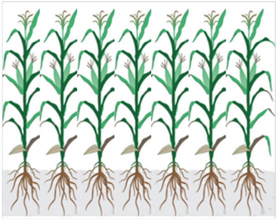
Range from -0.68 to -0.22



Competition in METs

Pure stand performance

Breeding trial



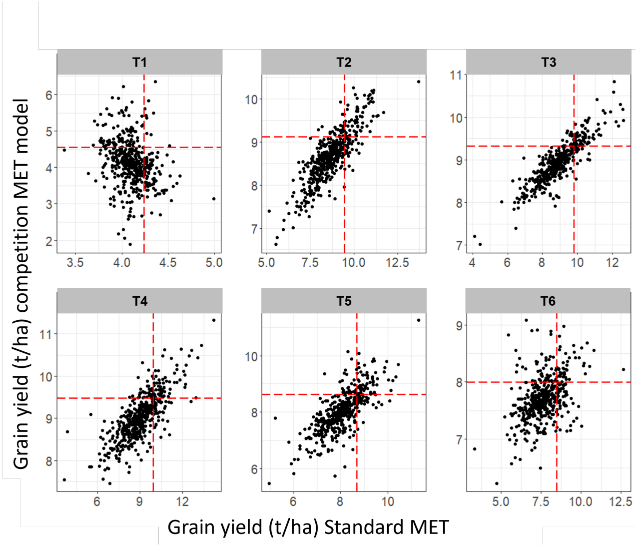
Farmers' fields

- The purpose of breeding trials to estimate performance of candidate varieties in farmers' fields (pure stand) (Goldringer et al. 1994)

$$\mathbf{u}_{\mathrm{d}} + 2\mathbf{u}_{\mathrm{n}}$$

The MET model for competition

Predictions from standard MET versus competition MET



- Competition changes ranking of genotypes in breeding trials
- The effect of the inter-plot competition on genotype selection is greater under stressed environments
- When there is a strong negative correlation between direct and neighbour effects, then the variance of the pure stand performance is reduced, relative to the standard MET

Competition in METs

Conclusion

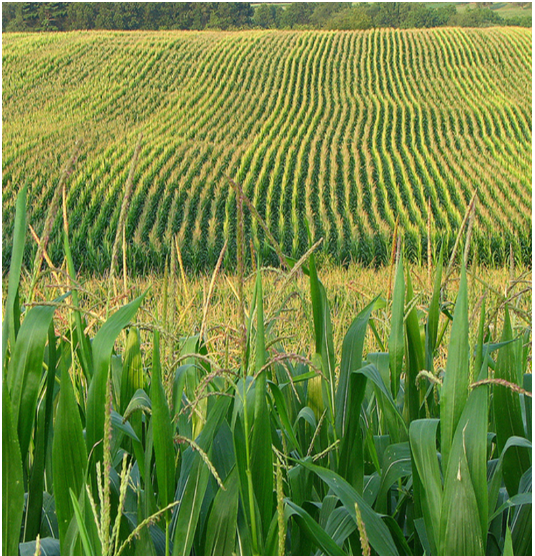
- Single-row plots are widely used in maize breeding programs in sub-Saharan Africa due to the limitations of land and resources for breeding trials

- In these trials inter-plot competition biases grain yield predictions resulting in changes in genotype rankings and reduced selection accuracy unless accounted for with appropriate statistical models

Competition in METs

Implications

- Use of the competition MET model in breeding programs that currently run multiple-row plots in early-generation yield trials could enable a change to single-row plots and increase the number of test genotypes that can be evaluated with the same resources, hence increasing genetic gain.



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

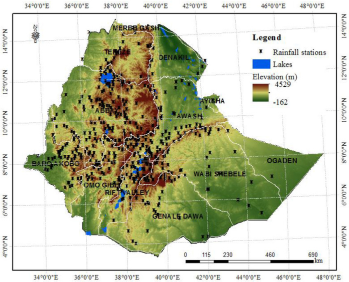
Modernizing Ethiopian Research on Crop Improvement

Acknowledgements

BILL & MELINDA GATES *foundation*

EIAR
[illegible]
Ethiopian Institute of Agricultural Research

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
QAAFI
Queensland Alliance for Agriculture and Food Innovation



Ameseginalehu