



THE UNIVERSITY  
of EDINBURGH



Biotechnology and  
Biological Sciences  
Research Council



THE ROYAL  
SOCIETY

# Day 4

## Spatial variation and genotype by environment interaction

*Multi-environment trials*

Daniel Tolhurst, Chris Gaynor, Gregor Gorjanc, Jon Bancic



# Lecture overview

## 1. Multi-environment plant breeding field trials ←

- Genotype by environment (GxE) interaction
- Linear mixed models for MET data

## 2. A framework for simulating GxE interaction

- MET-TPE concepts
- Applications to comparing statistical approaches and different breeding strategies over time.



**Multi-environment plant  
breeding field trials**

# Outline

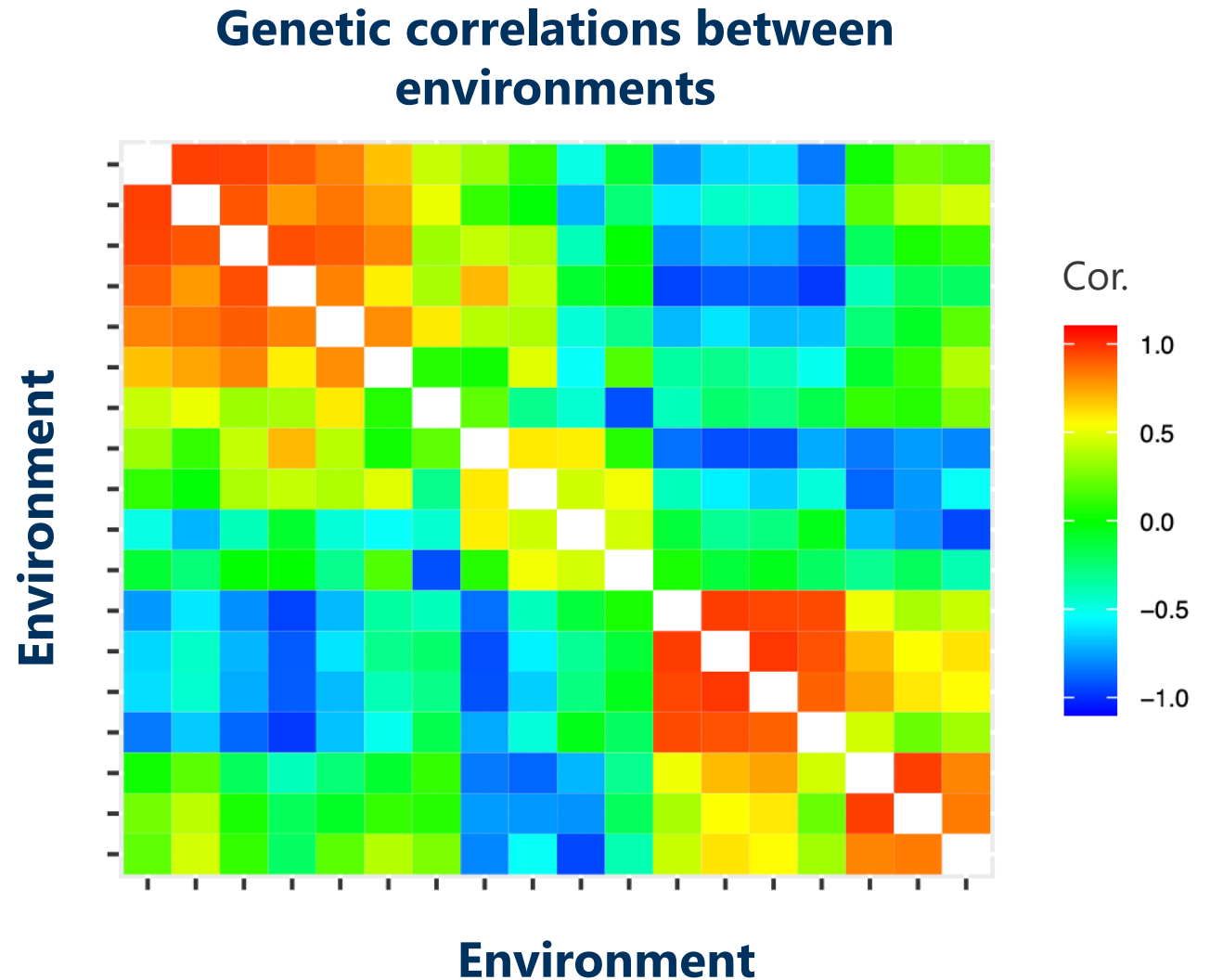
- **Background**
- **Genotype by environment (GxE) interaction**
- **Linear mixed models for GxE**
  - Main effects only
  - Diagonal model
  - Compound symmetry model
  - Unstructured model
  - Factor analytic model

# Background

- **Genotype by environment (GxE) interaction complicates plant and animal breeder's selection decisions**
  - The relative response of individuals (genotypes) changes when placed in different environments
  - Some genotypes may tolerate harsh environments, while others do not
  - Can reflect substantial re-rankings between environments

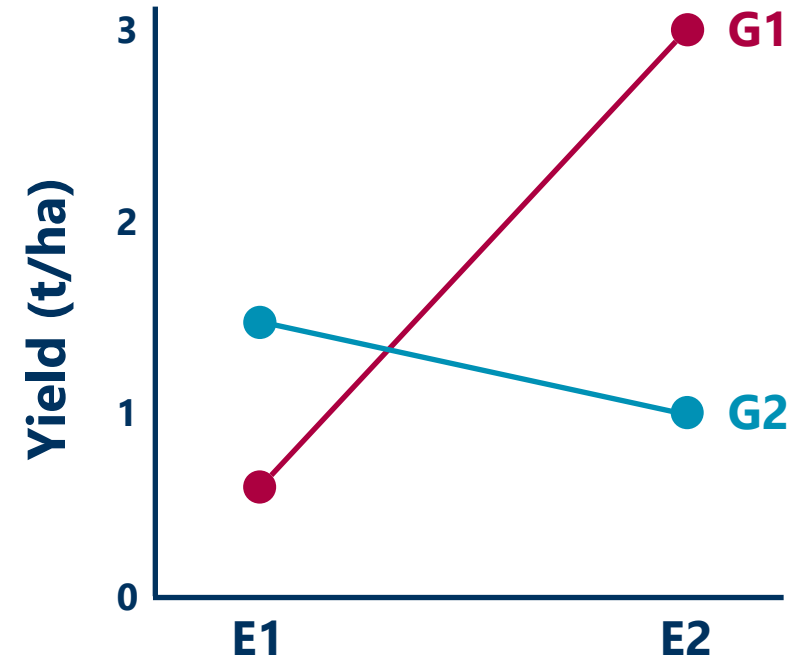
# GxE interaction

- Often viewed using the genetic correlations between environments
  - Cor. of 1 = **perfect agreement** in rankings between environments
  - Cor. of 0 = **dissimilarity** in rankings between environments
  - Cor. of -1 = **complete reversal** in rankings between environments



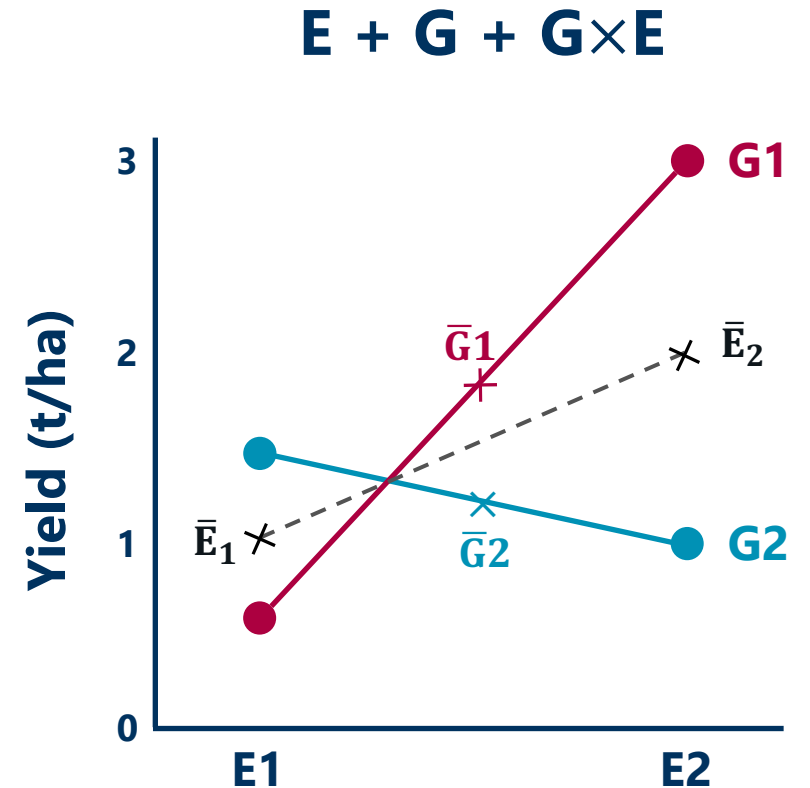
# Toy example

- Consider two wheat varieties, G1 & G2, grown in two environments, E1 & E2
- The crops are harvested and measured for grain yield (tonnes/hectare)
- The response of the two varieties changes between the two environments, but is this just GxE?



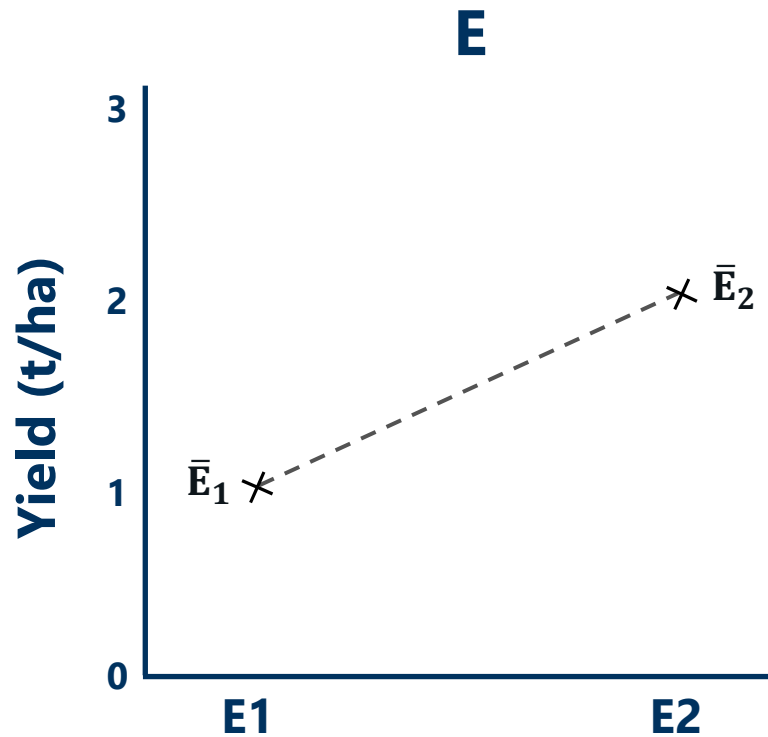
# Toy example

- **Environmental main effects**
  - $\bar{E}_1 \neq \bar{E}_2$ , E2 yields more on average
- **Genotype main effects**
  - $\bar{G}_1 \neq \bar{G}_2$ , G1 yields more on average
- **GxE interaction effects**
  - Relative response of G1 and G2 changes (lines are not parallel)
  - G2 yields more in E1 but G1 yields more in E2

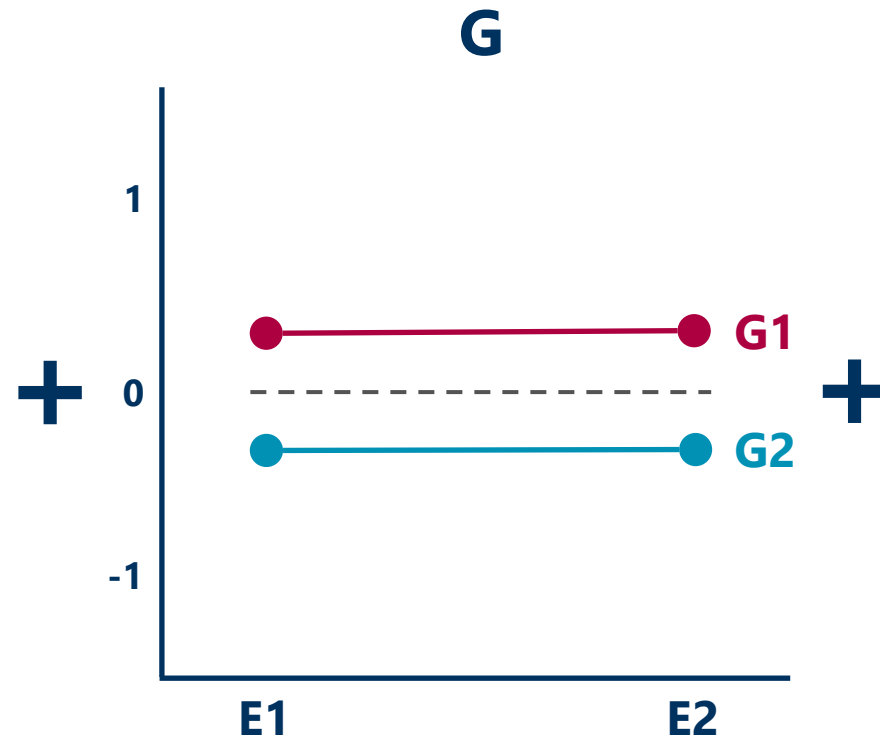




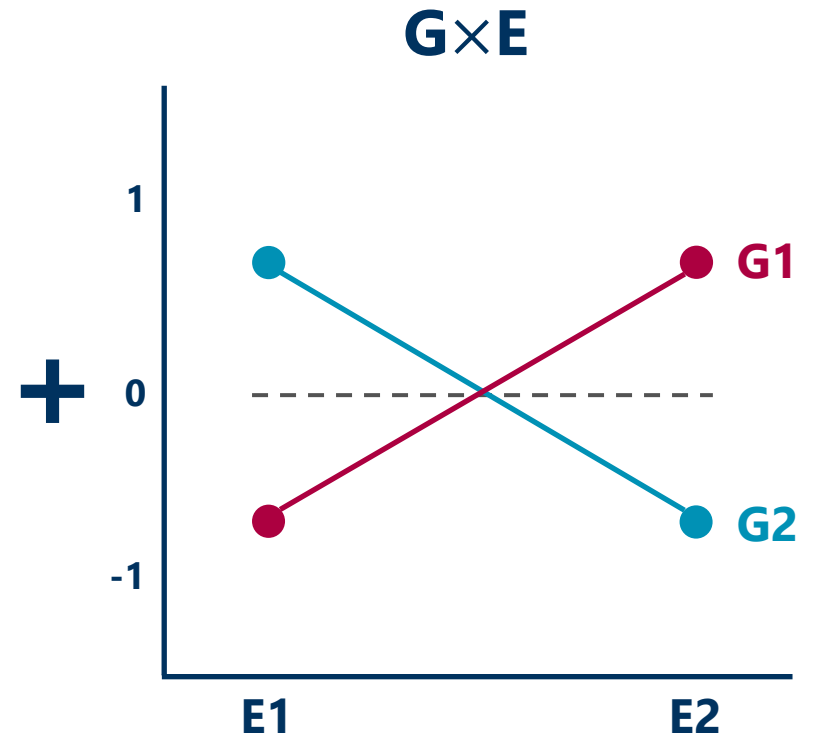
# Toy example



**Environmental main effects** as  $\bar{E}_1 \neq \bar{E}_2$   
– E2 yields more on average

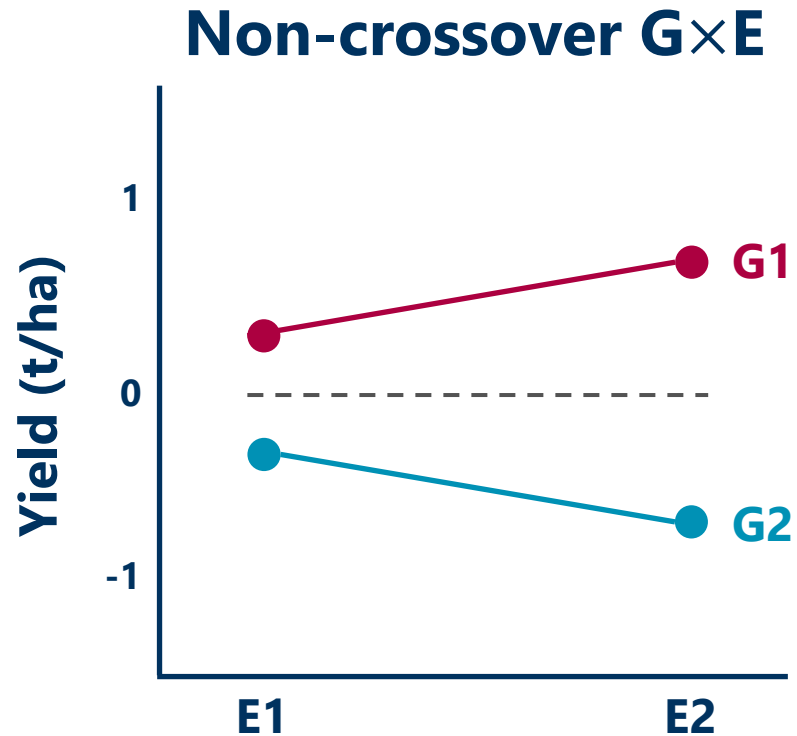


**Genotype main effects** as  $\bar{G}_1 \neq \bar{G}_2$   
– G1 yields more on average



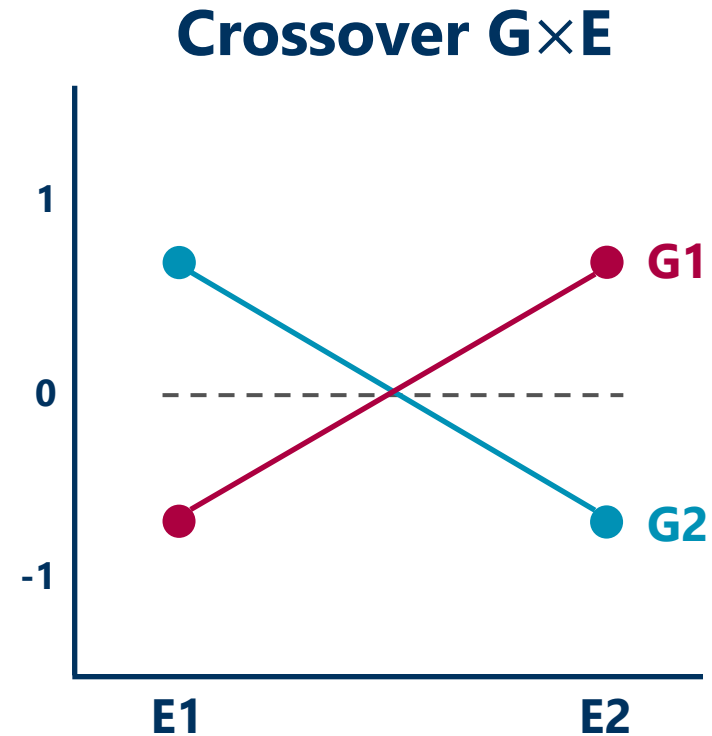
**G×E interaction effects** as the relative response of G1 and G2 changes (lines are not parallel)  
– G2 yields more in E1 but G1 yields more in E2

# Non-crossover and crossover GxE



**Non-crossover GxE interaction effects** as the relative response of G1 and G2 changes but the lines don't intersect

- G1 yields more in E1 and E2



**Crossover GxE interaction effects** as the relative response of G1 and G2 changes (lines are not parallel)

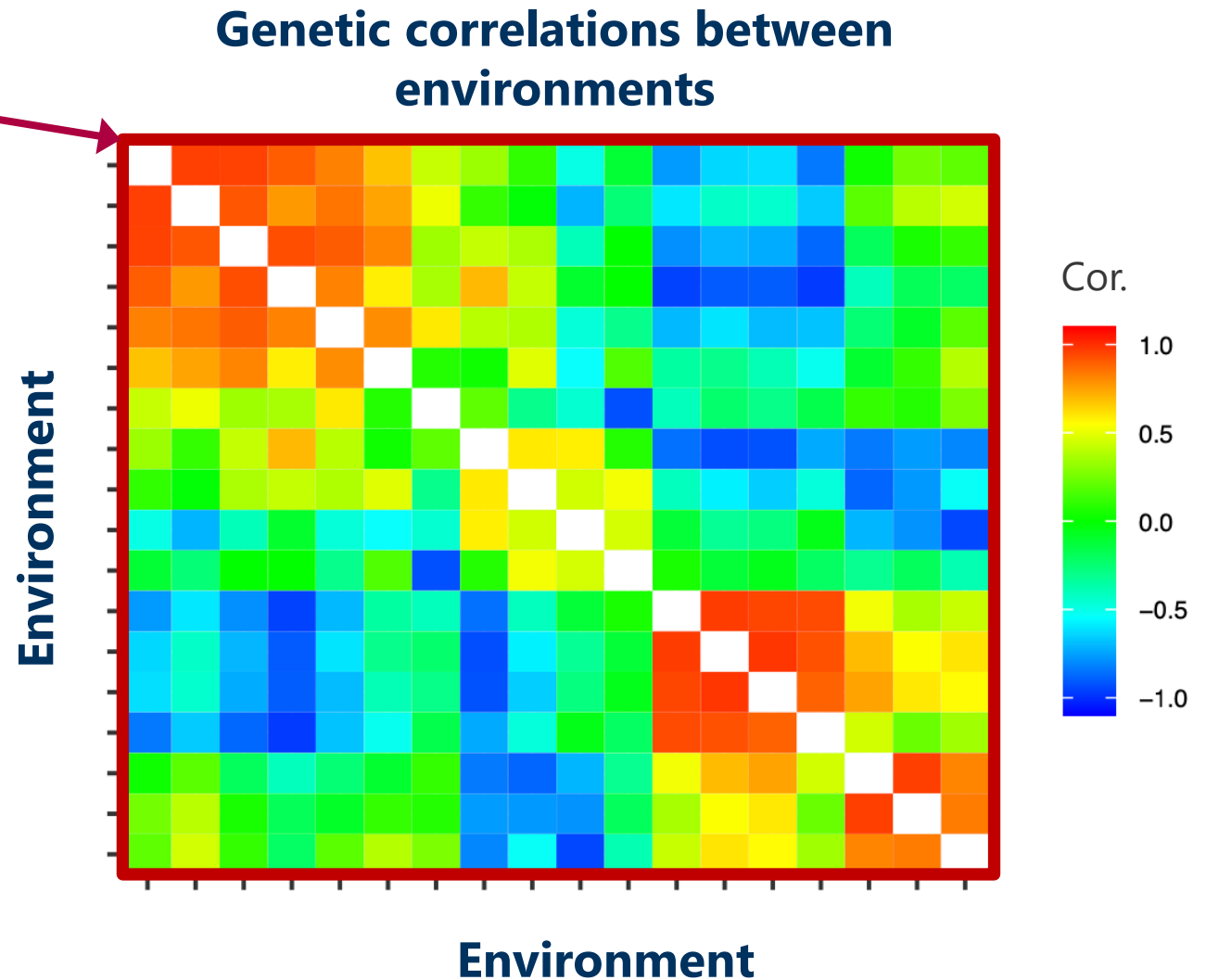
- G2 yields more in E1 but G1 yields more in E2

# So how can we handle GxE?

- **GxE has been historically handled in one of three ways:**
  1. **Ignore GxE** by selecting the most favourable genotypes on average
  2. **Reduce GxE** by grouping similar environments together and selecting within each group
  3. **Leverage GxE** by selecting the most favourable individuals in terms of average performance and stability (adaptability) ←
- May also consider some combination of 2 and 3

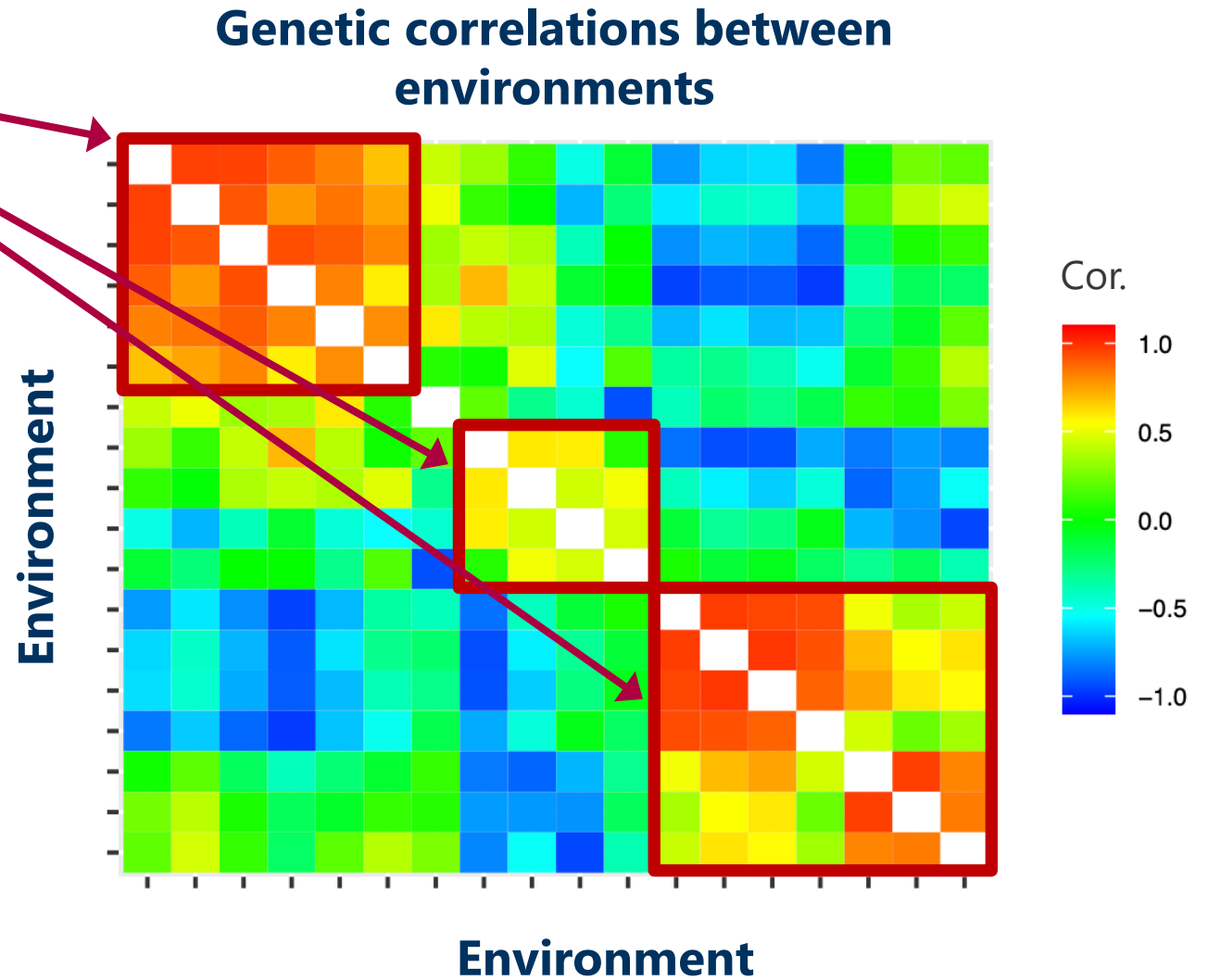
# Ignore $G \times E$ interaction

- Select the most favourable genotypes on average across all environments
- Selection for increasing overall genetic gain, but...
- Ignores important crossover GxE (re-rankings)
- **Potential to release poorly adapted genotypes to growers!**



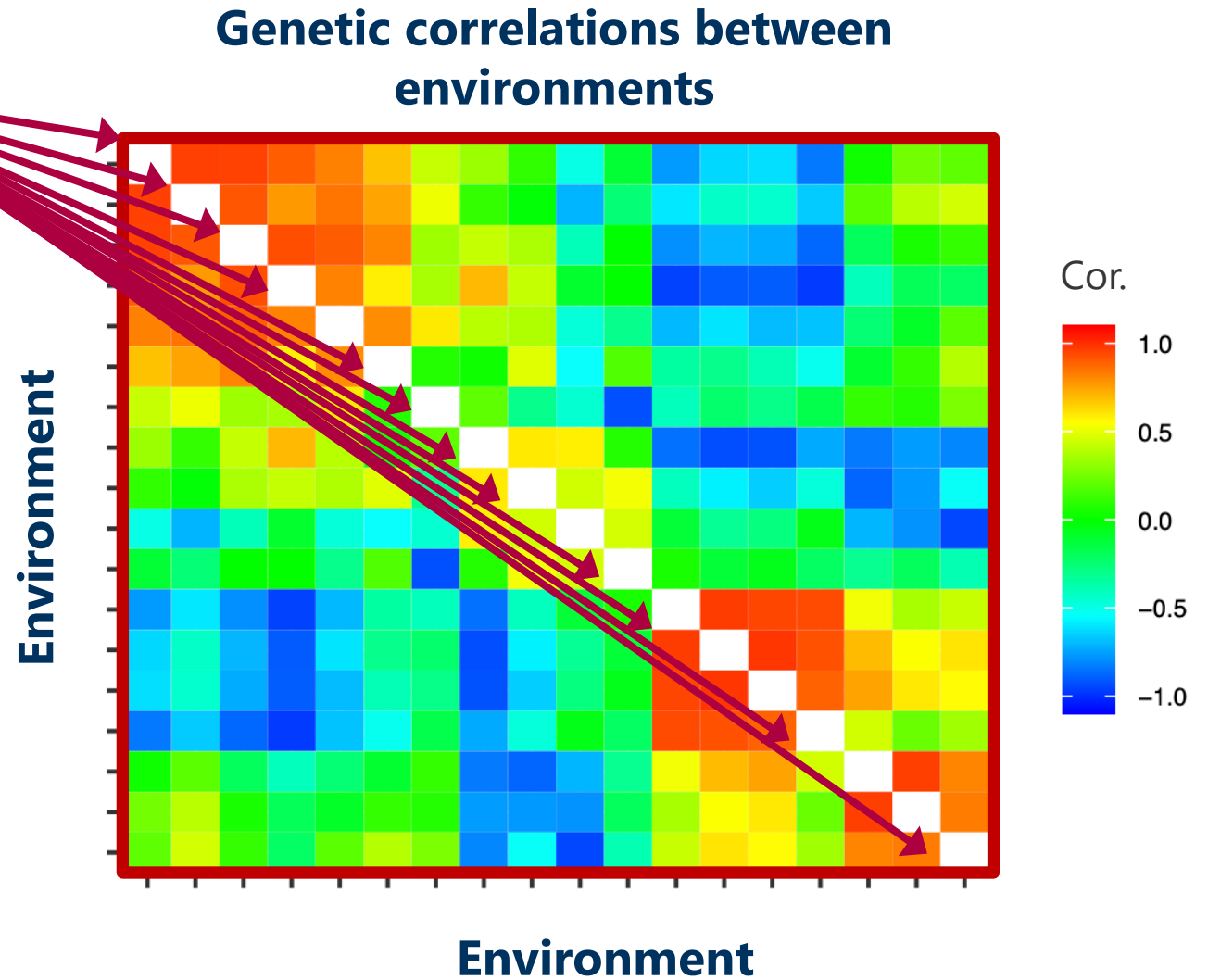
# Reduce $G \times E$ interaction

- Group similar environments together
- Select the most favourable genotypes on average within each group
- Selection for increasing genetic gain within groups, but...
- **Assumes you can explain the different groupings, and they are repeatable**



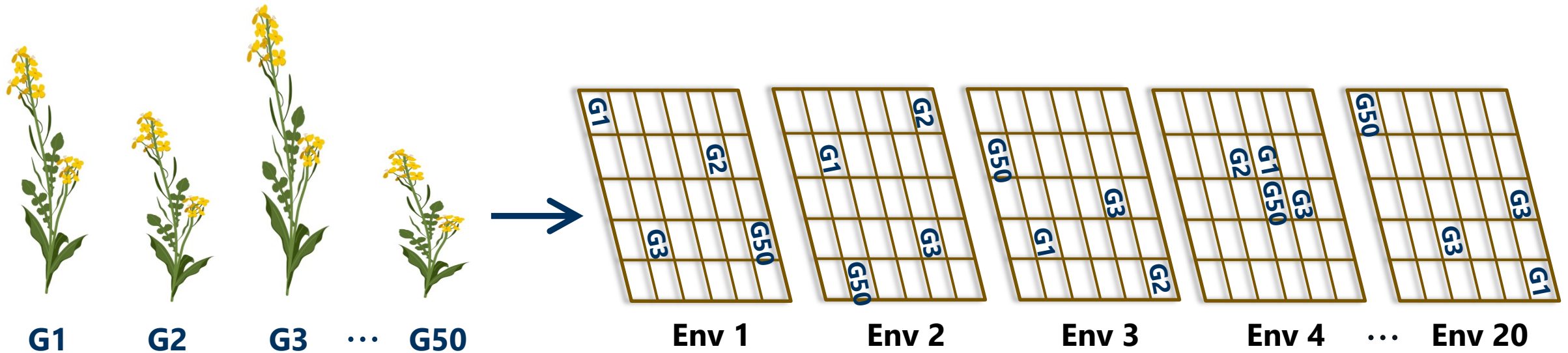
# Leverage $G \times E$ interaction

- Select the most favourable genotypes for average performance and stability
- **Selection for increasing overall genetic gain, and releasing well adapted genotypes to growers!**



# Multi-environment trial (MET) dataset

- Gauge GxE by accumulating and analysing multi-environment trial (MET) data



G1 G2 G3 ... G50

Env 1

Env 2

Env 3

Env 4

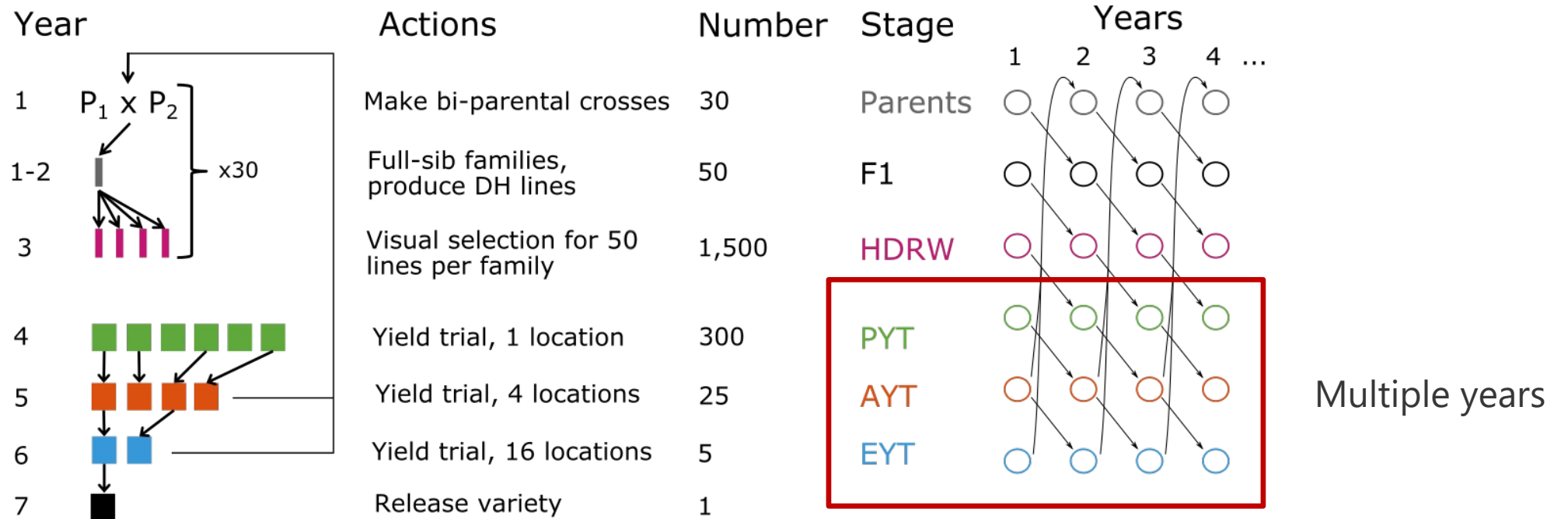
... Env 20

Plant genotypes

May be incomplete and highly unbalanced across environments

# Multi-environment trial (MET) dataset

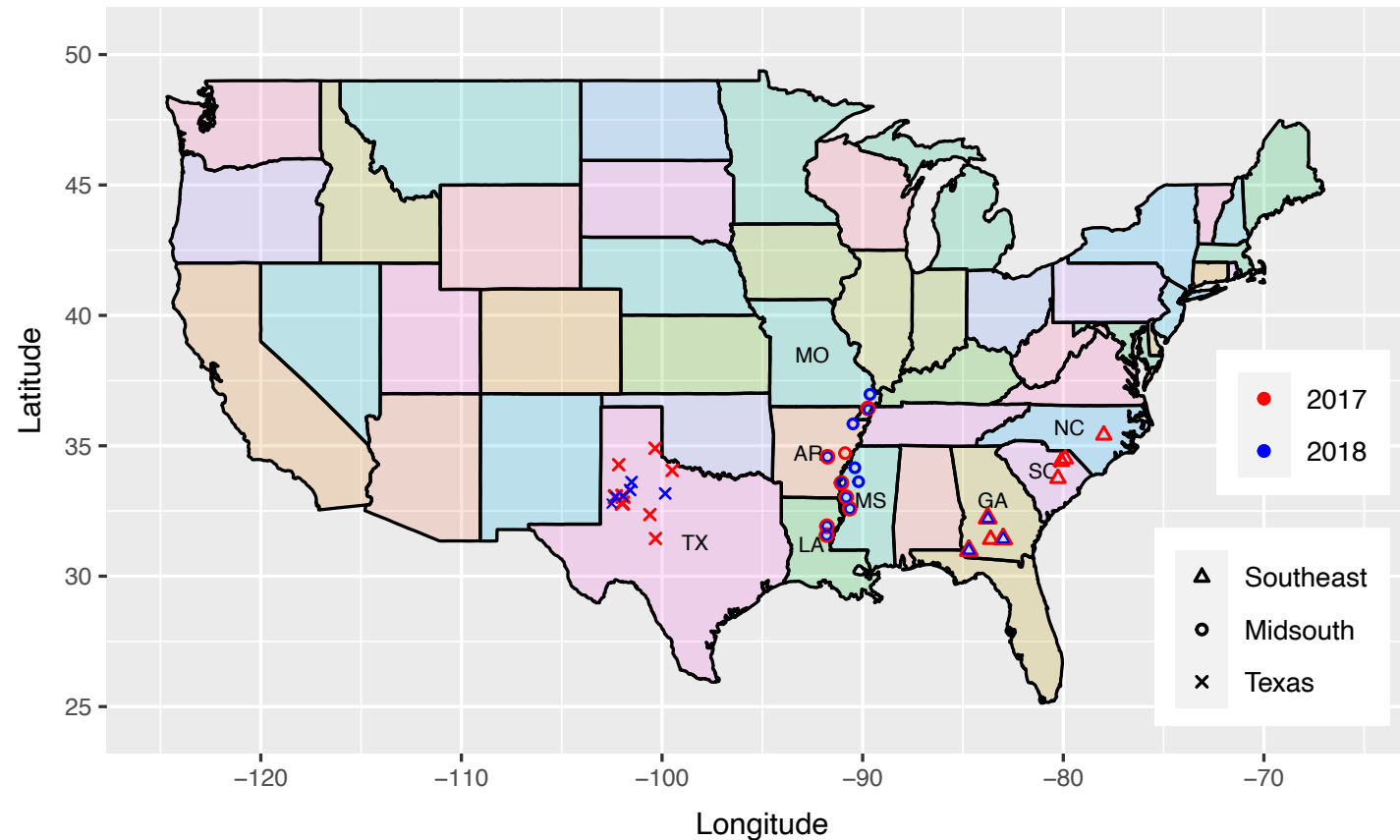
- Gauge GxE by accumulating and analysing multi-environment trial (MET) data





# Multi-environment trial (MET) dataset

- Gauge GxE by accumulating and analysing multi-environment trial (MET) data



Multiple locations

# Multi-environment trial (MET) dataset

- Gauge GxE by accumulating and analysing multi-environment trial (MET) data

	env <fctr>	block <fctr>	col <fctr>	row <fctr>	id <fctr>	phe.Trait1 <dbl>
1	1	1	1	1	481	1.64066056
2	1	1	1	2	543	2.51841946
3	1	1	1	3	793	4.15104719
4	1	1	1	4	768	2.21278404
5	1	1	1	5	814	4.16124862
6	1	1	1	6	645	4.95910345
7	1	1	1	7	500	2.07052692
8	1	1	1	8	437	4.31971130
9	1	1	1	9	444	1.75114321
10	1	1	1	10	781	2.88818150

 multiple year-location combinations



**Linear mixed models for  
multi-environment field  
trials**

# Randomised complete block (RCB) design

$$y_{ij} = \mu + g_i + b_j + e_{ij}$$

phenotype   mean   genotype   block   residual

- Simple to construct
- Balanced, complete and resolvable
- Genotypes and blocks are orthogonal
- **But**, assumes blocks are homogeneous

Block 1		Block 2	
G1	G4	G5	G12
G14	G11	G10	G16
G6	G8	G7	G2
G10	G16	G3	G15
G2	G5	G9	G8
G3	G13	G1	G6
G12	G15	G11	G4
G9	G7	G14	G13



# Vector notation

$$\mathbf{y} = \mathbf{1}_n \mu + \mathbf{X} \boldsymbol{\tau} + \mathbf{Z} \mathbf{g} \mathbf{e} + \mathbf{W} \mathbf{b} + \mathbf{e}$$

phenotype    mean    env.    genotype  
in env.    block    residual  
in env.

- $\mathbf{y}$  is the  $n$ -vector of phenotypes (ordered as plots in blocks in envs)
- $\mu$  is the overall mean,  $\mathbf{1}_n$  is a  $n$ -vector of ones
- $\boldsymbol{\tau}$  is the  $n_e$ -vector of environmental main effects with  $n \times n_e$  design matrix  $\mathbf{X}$  which links plots to environments
- $\mathbf{g} \mathbf{e}$  is the  $n_g \times n_e$ -vector of **genotype effects in environments**, with  $n \times (n_g \times n_e)$  design matrix  $\mathbf{Z}$  which links plots to genotypes in envs.
- $\mathbf{b}$  is the  $n_b \times n_e$ -vector of block effects, with  $n \times (n_b \times n_e)$  design matrix  $\mathbf{W}$  which links plots to blocks in environments
- $\mathbf{e}$  is the  $n$ -vector of residuals

# Model assumptions

## Complex genetic and residual variance structures:

$$E(\mathbf{y}) = \mathbf{1}_n \mu + \mathbf{X}\boldsymbol{\tau} \quad \text{and} \quad \text{Var}(\mathbf{y}) = \mathbf{Z}(\mathbf{G}_e \otimes \mathbf{G})\mathbf{Z}' + \mathbf{W}\mathbf{B}\mathbf{W}' + \mathbf{R}$$

$$\begin{bmatrix} \mathbf{g}_e \\ \mathbf{b} \\ \mathbf{e} \end{bmatrix} \sim \mathbf{N} \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G}_e \otimes \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

- $\mathbf{G}_e \otimes \mathbf{G}$  is a  $(n_g \times n_e) \times (n_g \times n_e)$  variance matrix with  $n_e \times n_e$  between-environment genetic variance matrix,  $\mathbf{G}_e$ , and  $n_g \times n_g$  genotype relationship matrix,  $\mathbf{G}$
- $\mathbf{B}$  is a diagonal block variance matrix
- $\mathbf{R}$  is a  $n \times n$  residual variance matrix

# Models for the block effects

$$\text{Var}(\mathbf{b}) = \mathbf{B} = \bigoplus_{m=1}^{n_e} \mathbf{B}_m = \begin{bmatrix} \mathbf{B}_1 & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{B}_{n_e} \end{bmatrix}$$

$$\mathbf{B}_m = \sigma_{b_m}^2 \mathbf{I}_{n_b}$$

- $\sigma_{b_m}^2$  is the block variance for environment  $m$
- $\mathbf{I}_{n_b}$  is an identity matrix of order  $n_b$



# Models for the residuals

$$\text{Var}(\mathbf{e}) = \mathbf{R} = \bigoplus_{m=1}^{n_e} \mathbf{R}_m = \begin{bmatrix} \mathbf{R}_1 & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{R}_{n_e} \end{bmatrix}$$

$$\mathbf{R}_m = \sigma_{s_m}^2 \boldsymbol{\Sigma}_{\mathbf{c}_m}(\rho_{c_m}) \otimes \boldsymbol{\Sigma}_{\mathbf{r}_m}(\rho_{r_m}) + \sigma_{r_m}^2 \mathbf{I}_n$$

- $\sigma_{s_m}^2$  is the autoregressive scaling component for environment  $m$
- $\boldsymbol{\Sigma}_{\mathbf{c}_m}$  is a  $n_c \times n_c$  matrix with column autocorrelation  $\rho_{c_m}$  for env.  $m$
- $\boldsymbol{\Sigma}_{\mathbf{r}_m}$  is a  $n_r \times n_r$  matrix with row autocorrelation  $\rho_{r_m}$  for env.  $m$
- $\sigma_{r_m}^2$  is the random error variance component for env.  $m$

# Models for the genotype by environment effects

$$\text{Var}(\mathbf{ge}) = \mathbf{G}_e \otimes \mathbf{G} = \begin{bmatrix} \mathbf{G}_{e_1;1} \mathbf{G} & \dots & \mathbf{G}_{e_1;n_e} \mathbf{G} \\ \vdots & \ddots & \vdots \\ \mathbf{G}_{e_{n_e};1} \mathbf{G} & \dots & \mathbf{G}_{e_{n_e};n_e} \mathbf{G} \end{bmatrix}$$

- $\mathbf{G}_e$  is the  $n_e \times n_e$  between-environment genetic variance matrix
- $\mathbf{G}$  is the  $n_g \times n_g$  genotype relationship matrix

# 0. Main effects only model

$$\mathbf{G}_e = \begin{bmatrix} \sigma_g^2 & \dots & \sigma_g^2 \\ \vdots & \ddots & \vdots \\ \sigma_g^2 & \dots & \sigma_g^2 \end{bmatrix} \quad \text{and} \quad \mathbf{G}_e \otimes \mathbf{G} = \begin{bmatrix} \sigma_g^2 \mathbf{G} & \dots & \sigma_g^2 \mathbf{G} \\ \vdots & \ddots & \vdots \\ \sigma_g^2 \mathbf{G} & \dots & \sigma_g^2 \mathbf{G} \end{bmatrix}$$

$$ge_{im} = g_i \quad \text{i.e.} \quad \mathbf{g}_e = (\mathbf{1}_{n_e} \otimes \mathbf{g})$$

- $\text{Var}(\mathbf{g}) = \sigma_g^2 \mathbf{I}_{n_g}$ , where  $\sigma_g^2$  is the genotype main effect variance component
- Assumes genotype effects are the same across environments!
  - **Does not model GxE interaction (ignores GxE)**

# 1. Diagonal model

$$\mathbf{G}_e = \begin{bmatrix} \sigma_{ge_1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{ge_{n_e}}^2 \end{bmatrix} \quad \text{and} \quad \mathbf{G}_e \otimes \mathbf{G} = \begin{bmatrix} \sigma_{ge_1}^2 \mathbf{G} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \sigma_{ge_{n_e}}^2 \mathbf{G} \end{bmatrix}$$

$$ge_{im} \quad \text{i.e.} \quad \mathbf{ge} = (\mathbf{ge}'_1, \mathbf{ge}'_2, \dots, \mathbf{ge}'_{n_e})'$$

- $\text{Var}(\mathbf{ge}_m) = \sigma_{ge_m}^2 \mathbf{I}_{n_g}$ , where  $\sigma_{ge_m}^2$  is the genetic variance for env.  $m$
- Assumes genotype effects across environments are independent
  - **Does not leverage GxE interaction**

## 2. Compound symmetry model

$$\mathbf{G}_e = \begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & \dots & \sigma_g^2 \\ \vdots & \ddots & \vdots \\ \sigma_g^2 & \dots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix} \text{ and } \mathbf{G}_e \otimes \mathbf{G} = \begin{bmatrix} (\sigma_g^2 + \sigma_{ge}^2)\mathbf{G} & \dots & \sigma_g^2\mathbf{G} \\ \vdots & \ddots & \vdots \\ \sigma_g^2\mathbf{G} & \dots & (\sigma_g^2 + \sigma_{ge}^2)\mathbf{G} \end{bmatrix}$$

$$ge_{im} = g_i + g \times e_{im} \quad \text{i.e.} \quad \mathbf{ge} = (\mathbf{1}_{n_e} \otimes \mathbf{g}) + \mathbf{g} \times \mathbf{e}$$

- $\text{Var}(\mathbf{g}) = \sigma_g^2 \mathbf{I}_{n_g}$ , where  $\sigma_g^2$  is the genotype main effect variance
- $\text{Var}(\mathbf{ge}) = \sigma_{ge}^2 \mathbf{I}_{n_g \times n_e}$ , where  $\sigma_{ge}^2$  is the GxE interaction variance

## 2. Compound symmetry model

$$\mathbf{G}_e = \begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & \dots & \sigma_g^2 \\ \vdots & \ddots & \vdots \\ \sigma_g^2 & \dots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix} \text{ and } \mathbf{G}_e \otimes \mathbf{G} = \begin{bmatrix} (\sigma_g^2 + \sigma_{ge}^2)\mathbf{G} & \dots & \sigma_g^2\mathbf{G} \\ \vdots & \ddots & \vdots \\ \sigma_g^2\mathbf{G} & \dots & (\sigma_g^2 + \sigma_{ge}^2)\mathbf{G} \end{bmatrix}$$

$$ge_{im} = g_i + g \times e_{im} \quad \text{i.e.} \quad \mathbf{g}\mathbf{e} = (\mathbf{1}_{n_e} \otimes \mathbf{g}) + \mathbf{g} \times \mathbf{e}$$

- Assumes genotype effects across environments are correlated
  - **But**, not sensible, assumes same variance within environments and same covariance between pairs of environments

### 3. Unstructured model

$$\mathbf{G}_e = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1n_e} \\ \vdots & \ddots & \vdots \\ \sigma_{1n_e} & \dots & \sigma_{n_e}^2 \end{bmatrix} \quad \text{and} \quad \mathbf{G}_e \otimes \mathbf{G} = \begin{bmatrix} \sigma_1^2 \mathbf{G} & \dots & \sigma_{1n_e} \mathbf{G} \\ \vdots & \ddots & \vdots \\ \sigma_{1n_e} \mathbf{G} & \dots & \sigma_{n_e}^2 \mathbf{G} \end{bmatrix}$$

$$\mathbf{ge} = (\mathbf{ge}'_1, \mathbf{ge}'_2, \dots, \mathbf{ge}'_{n_e})'$$

- $\sigma_m^2$  is the genetic variance for environment  $m$
- $\sigma_{jm}$  is the genetic covariance between environments  $j$  and  $m$

### 3. Unstructured model

$$\mathbf{G}_e = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1n_e} \\ \vdots & \ddots & \vdots \\ \sigma_{1n_e} & \dots & \sigma_{n_e}^2 \end{bmatrix} \quad \text{and} \quad \mathbf{G}_e \otimes \mathbf{G} = \begin{bmatrix} \sigma_1^2 \mathbf{G} & \dots & \sigma_{1n_e} \mathbf{G} \\ \vdots & \ddots & \vdots \\ \sigma_{1n_e} \mathbf{G} & \dots & \sigma_{n_e}^2 \mathbf{G} \end{bmatrix}$$

$$\mathbf{ge} = (\mathbf{ge}'_1, \mathbf{ge}'_2, \dots, \mathbf{ge}'_{n_e})'$$

- Fully parameterised model with  $n_e(n_e - 1)/2$  parameters
  - Becomes computationally prohibitive for large number of envs.
  - Captures noise and cannot be directly used to identify repeatable GxE Interaction



## 4. Factor analytic model

$$\mathbf{G}_e = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi} = \begin{bmatrix} \lambda_{11}^2 + \dots + \lambda_{r1}^2 + \psi_1 & \dots & \lambda_{11}\lambda_{1n_e} + \dots + \lambda_{r1}\lambda_{rn_e} \\ \vdots & \ddots & \vdots \\ \lambda_{11}\lambda_{1n_e} + \dots + \lambda_{r1}\lambda_{rn_e} & \dots & \lambda_{1n_e}^2 + \dots + \lambda_{rn_e}^2 + \psi_{n_e} \end{bmatrix}$$

$$g_{e_{im}} = \lambda_{1m}f_{1i} + \dots + \lambda_{km}f_{ki} + \delta_{im} \quad \text{i.e.} \quad \mathbf{g}_e = (\mathbf{\Lambda} \otimes \mathbf{I}_k)\mathbf{f} + \boldsymbol{\delta}$$

factor 1

factor  $k$

lack-of-fit

- $\lambda_{rm}$  is the latent covariate (loading) for environment  $m$  and factor  $r$  ( $r = 1, \dots, k$ ), typically  $k$  is small
- $f_{ri}$  is the slope (score) for genotype  $i$  and factor  $r$ ,  $\text{var}(f_{ri}) = 1g_{ii}$
- $\delta_{im}$  is the lack-of-fit of genotype  $i$  in env.  $m$ ,  $\text{var}(\delta_{im}) = \psi_1g_{ii}$

## 4. Factor analytic model

$$\mathbf{G}_e = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi} = \begin{bmatrix} \lambda_{11}^2 + \dots + \lambda_{r1}^2 + \psi_1 & \dots & \lambda_{11}\lambda_{1n_e} + \dots + \lambda_{r1}\lambda_{rn_e} \\ \vdots & \ddots & \vdots \\ \lambda_{11}\lambda_{1n_e} + \dots + \lambda_{r1}\lambda_{rn_e} & \dots & \lambda_{1n_e}^2 + \dots + \lambda_{rn_e}^2 + \psi_{n_e} \end{bmatrix}$$

$$ge_{im} = \lambda_{1m}f_{1i} + \dots + \lambda_{km}f_{ki} + \delta_{im} \quad \text{i.e.} \quad \mathbf{g}\mathbf{e} = (\mathbf{\Lambda} \otimes \mathbf{I}_k)\mathbf{f} + \boldsymbol{\delta}$$

factor 1                      factor  $k$       lack-of-fit

- **Reduced rank model with  $n_e(k + 1) - k(k - 1)/2$  parameters, which is much smaller than  $n_e(n_e - 1)/2$** 
  - Captures GxE with a small number of factors
  - Can be used to capture repeatable GxE interaction

# Making selections

- Overall performance for genotype  $i$ :

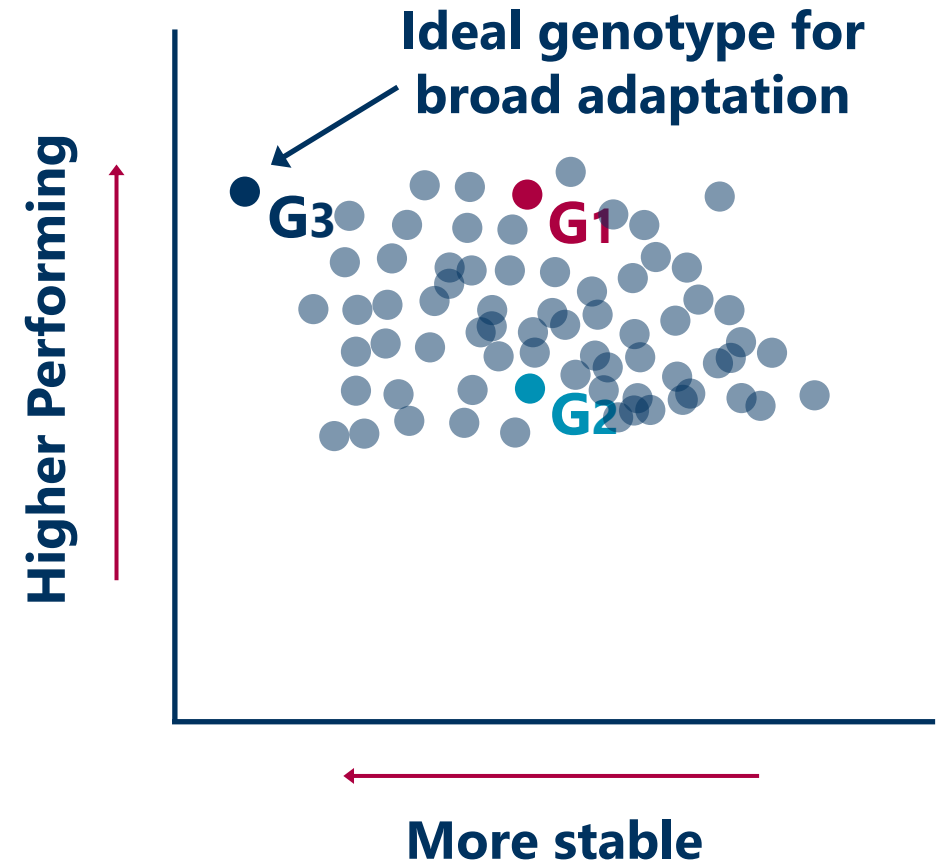
$$OP_i = \bar{\lambda}_1 f_{1i}$$

$\bar{\lambda}_1$  is the mean loading for factor 1

- Stability for genotype  $i$ :

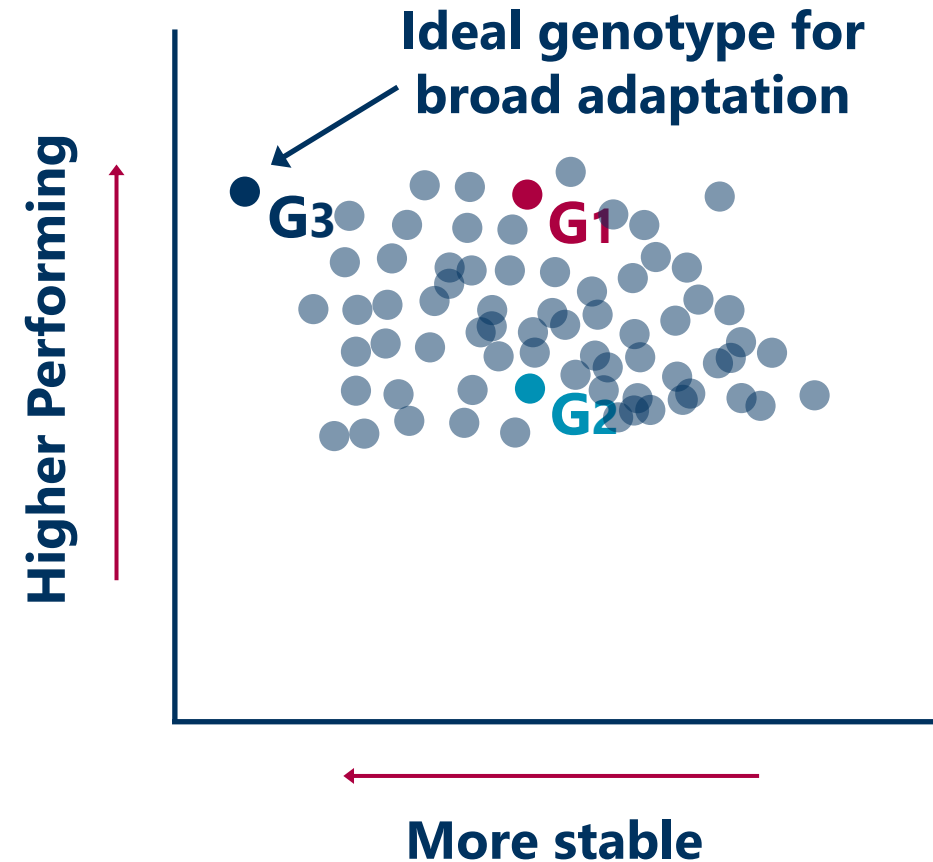
$$RMSD_i = \sqrt{\left(\sum_{r=2}^k \lambda_{rm} f_{ri}\right)^2 / n_e}$$

Root mean square of the deviations around the regression for factor 1

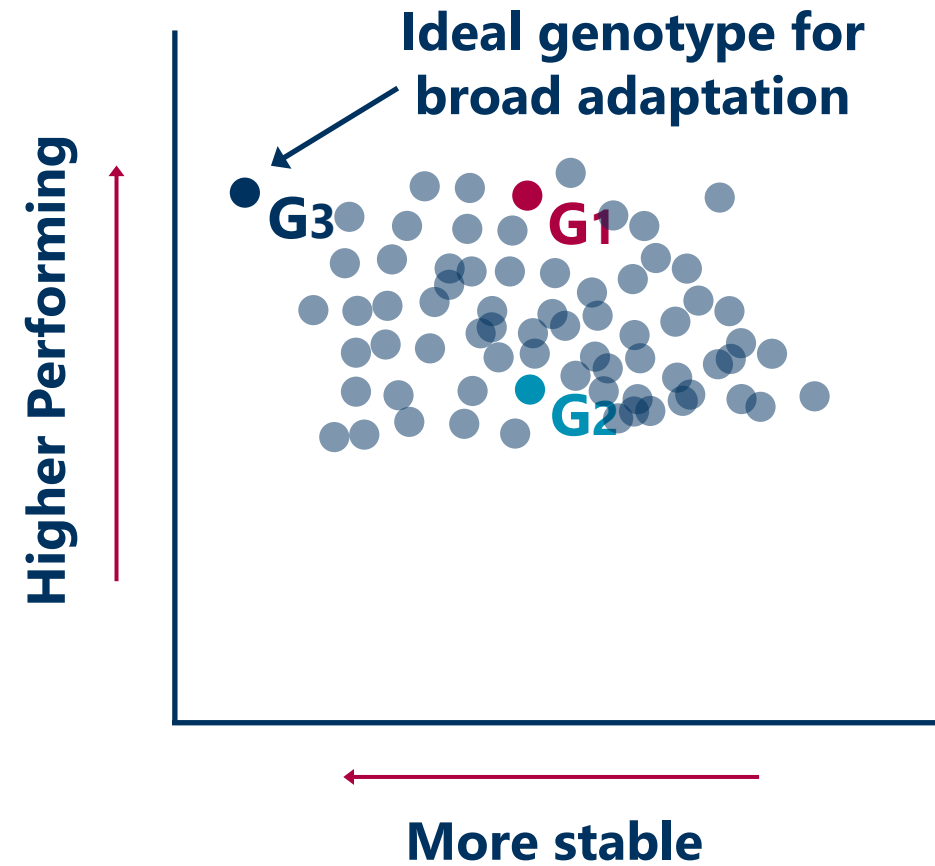
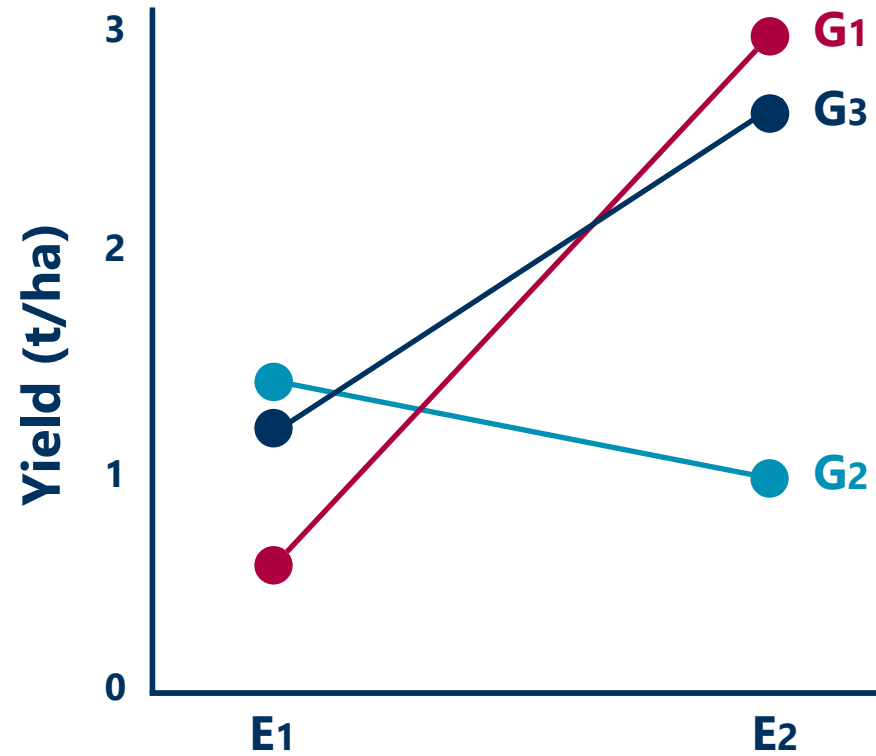


# Making selections

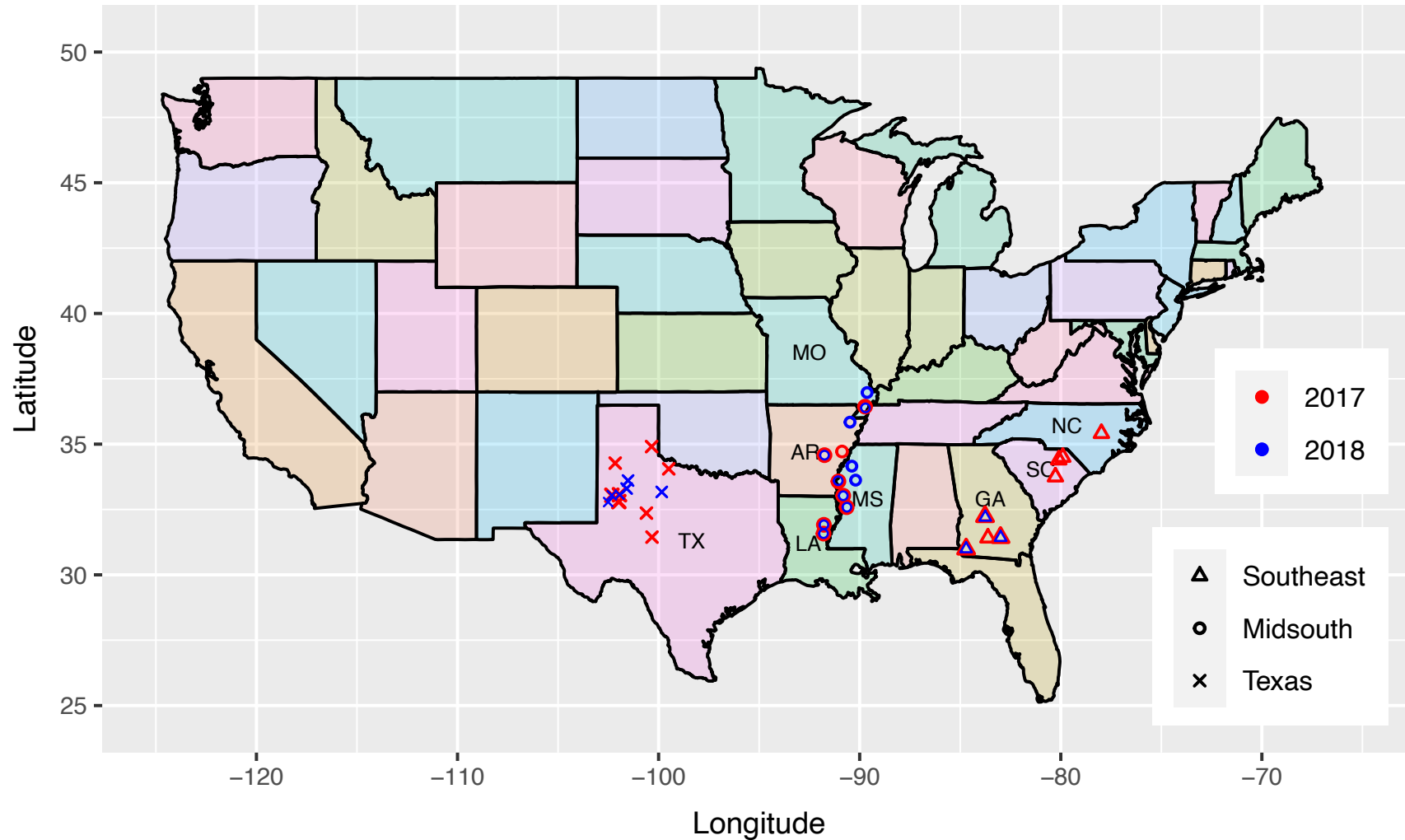
- **G3 is broadly adapted** because it is high performing on average and stable
- **G1 is likely to be specifically adapted** because it is high performing on average but unstable
- Summaries for a very large number of genotypes and environments
- Can be included within a selection index



# Making selections

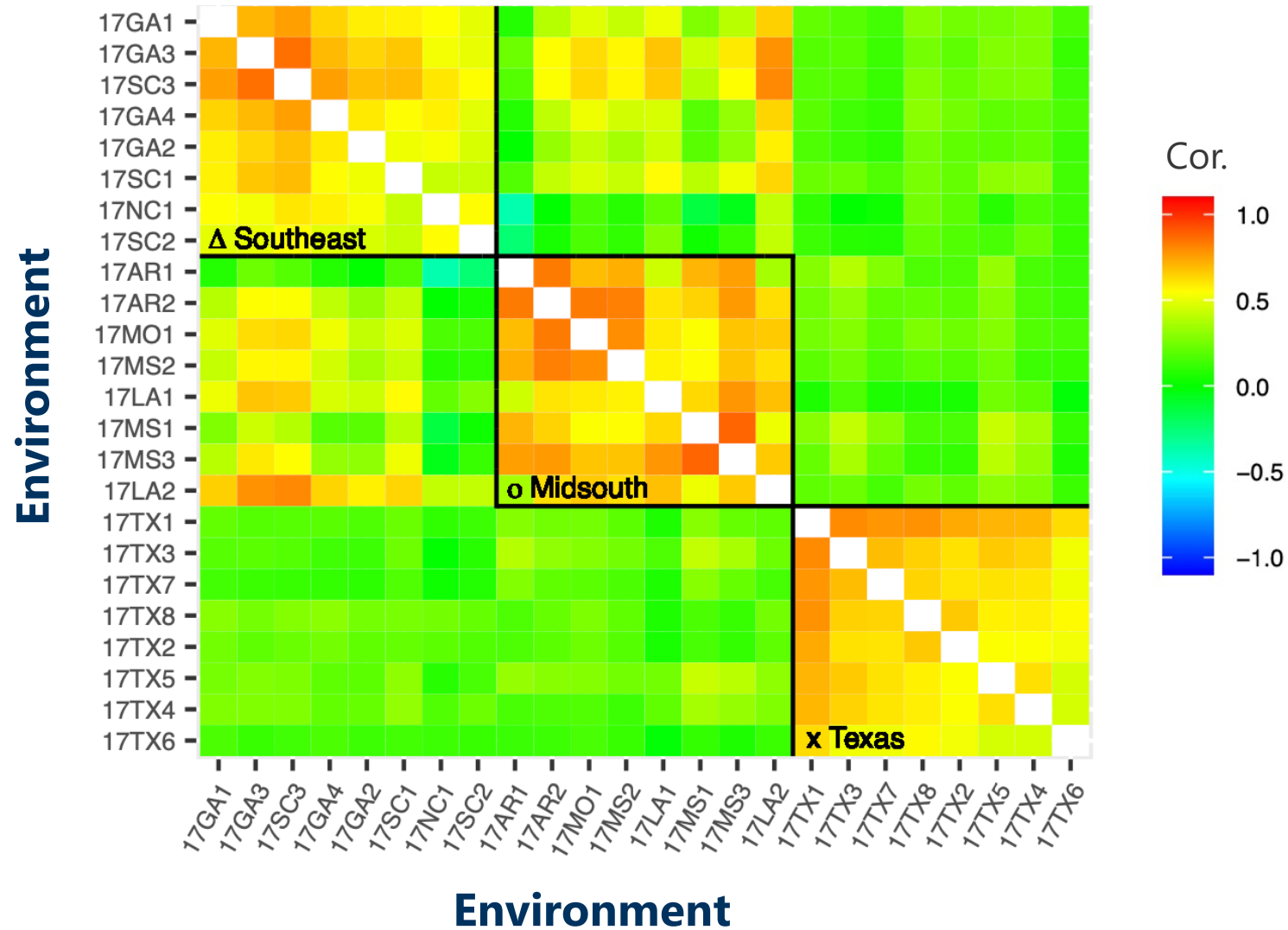


# Small example – environments

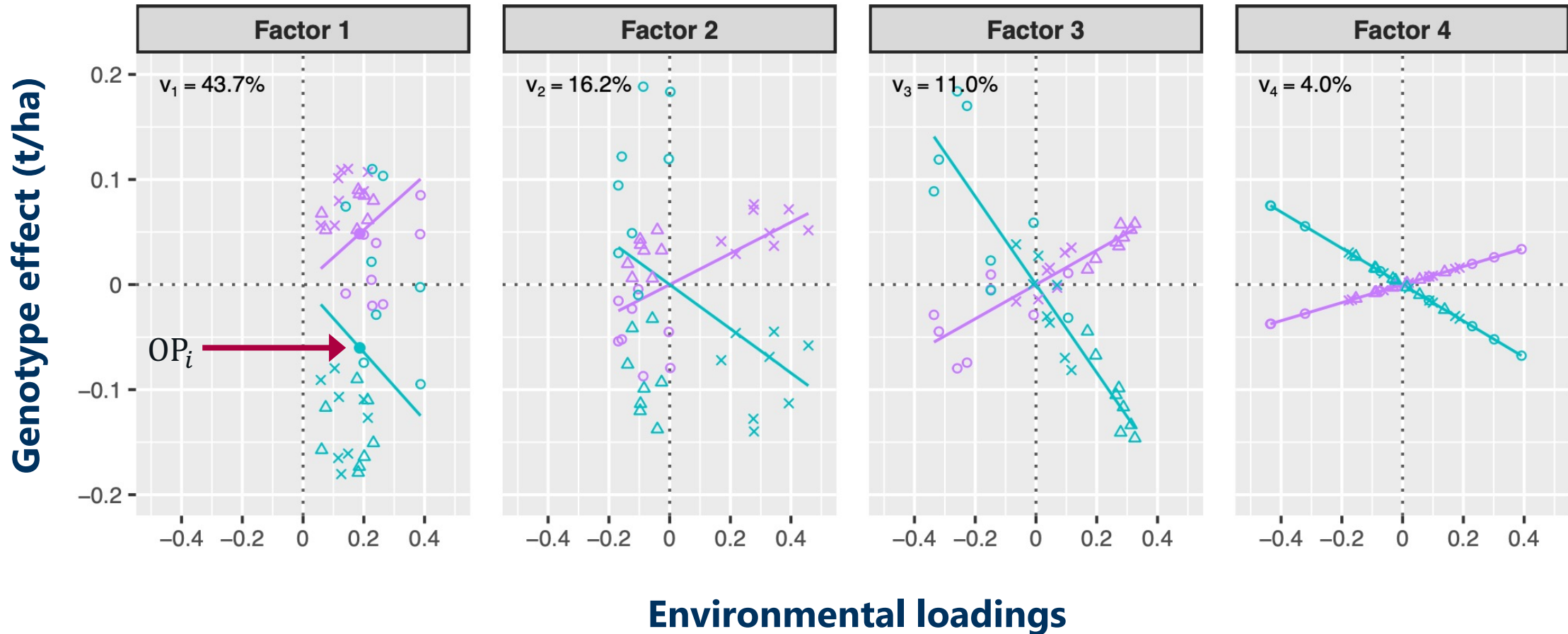


# Small example – observed GxE interaction

## Genetic correlations between environments

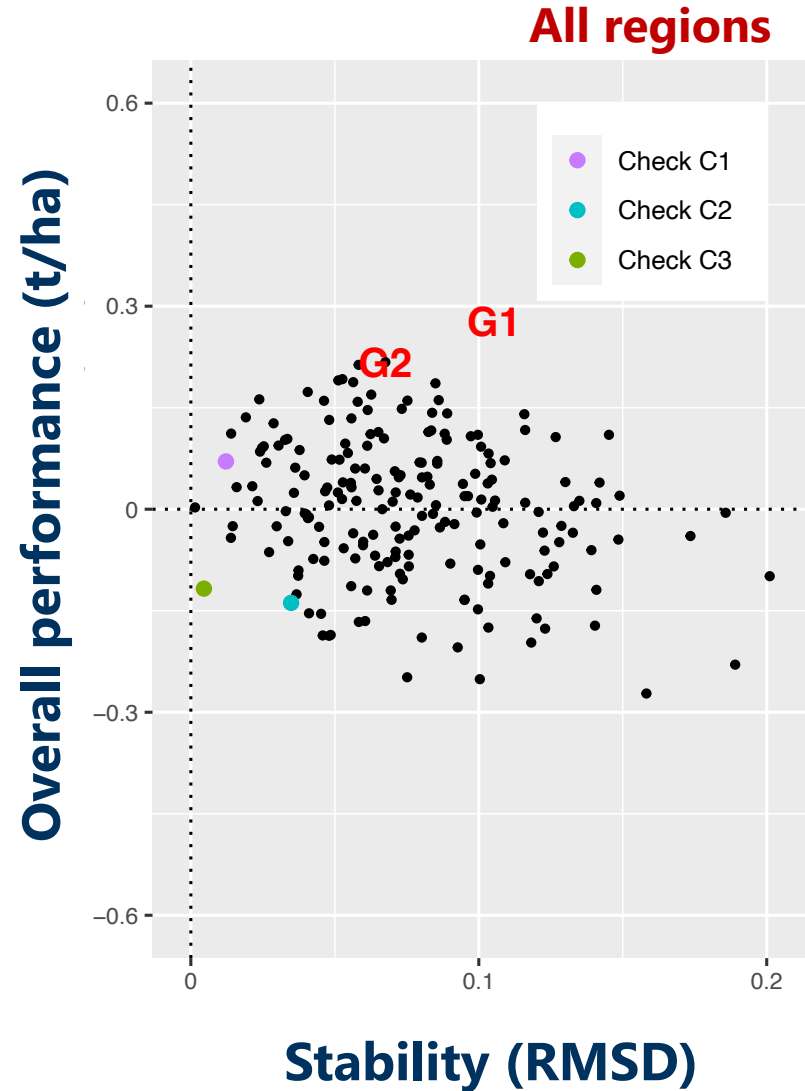
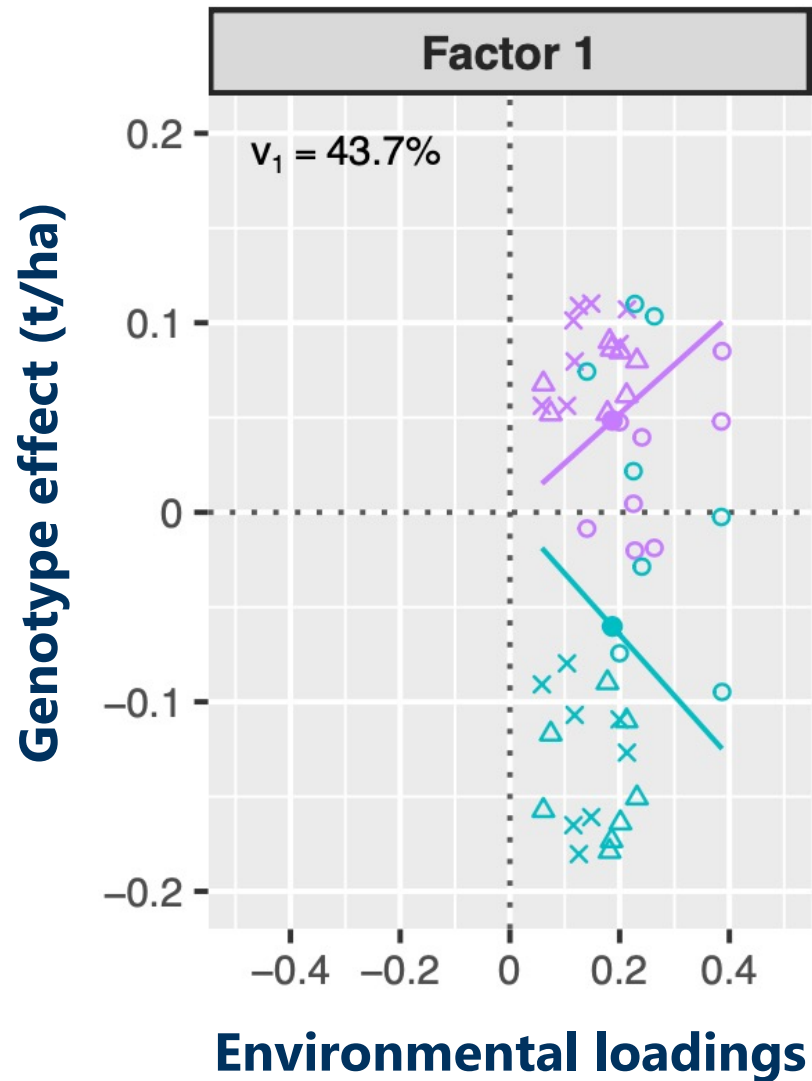


# Small example – factor analytic regression plot



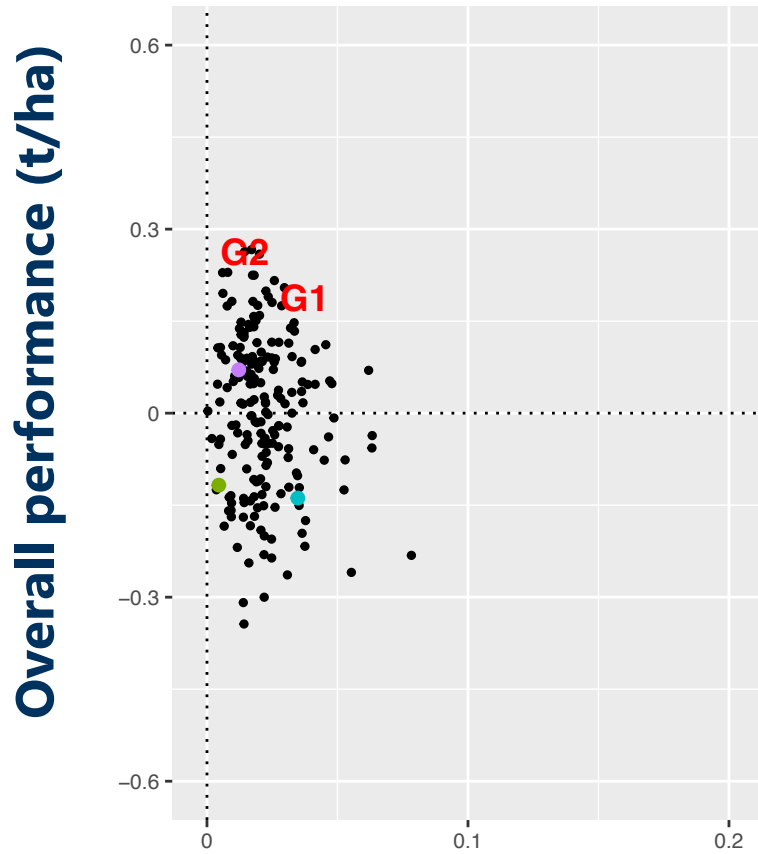


# Small example – making selections

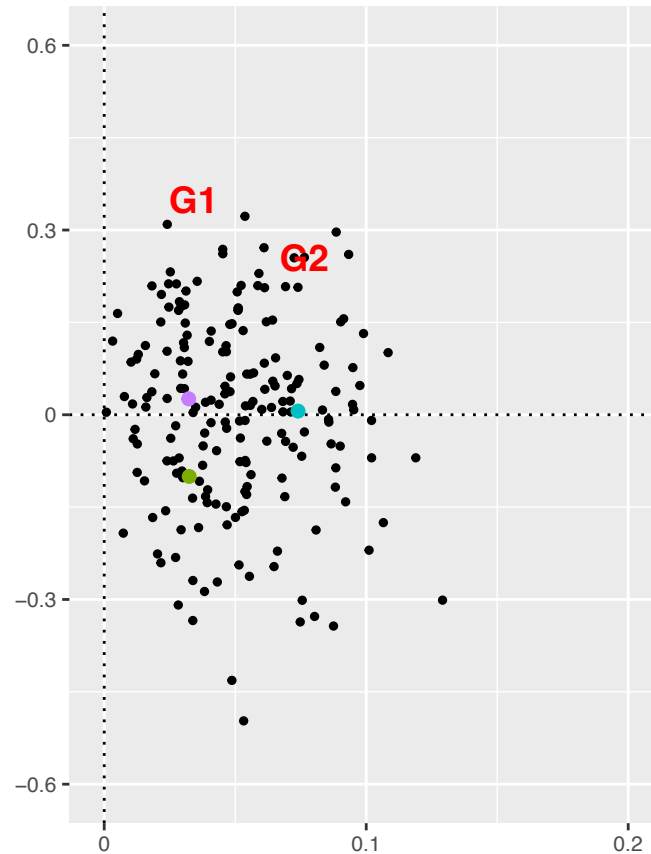


# Small example – making selections

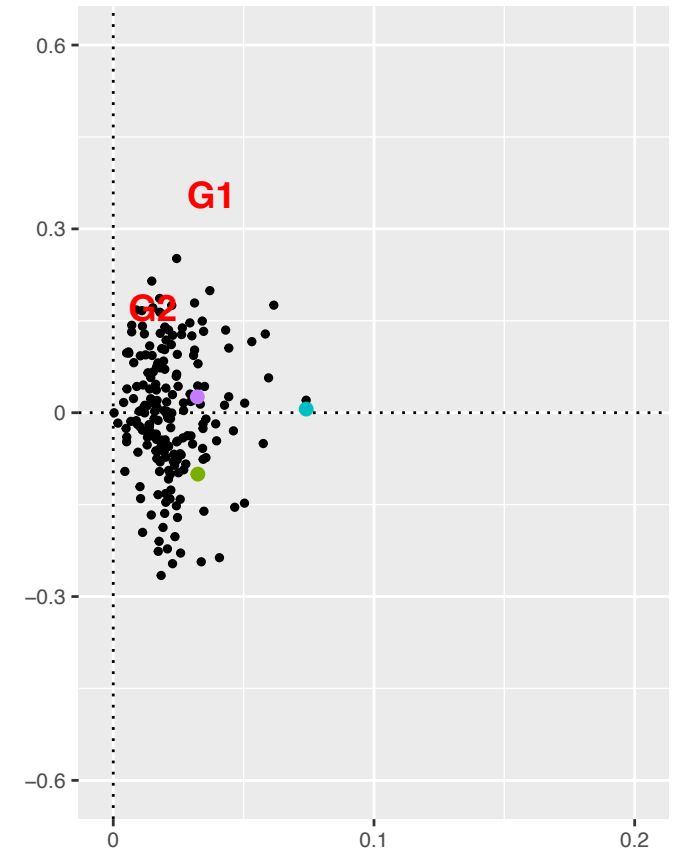
**Southeast**



**Midsouth**



**Texas**



**Stability (RMSE)**

# Lecture overview

## 1. Multi-environment plant breeding field trials ←

- Genotype by environment (GxE) interaction ✓
- Linear mixed models for MET data ✓

## 2. A framework for simulating GxE interaction

- MET-TPE concepts
- Applications to comparing statistical approaches and different breeding strategies over time