



Fundamentals of Genomic Prediction and Data-Driven Crop Breeding (August 4-8, 2025)

G x E Interactions in Plant Breeding

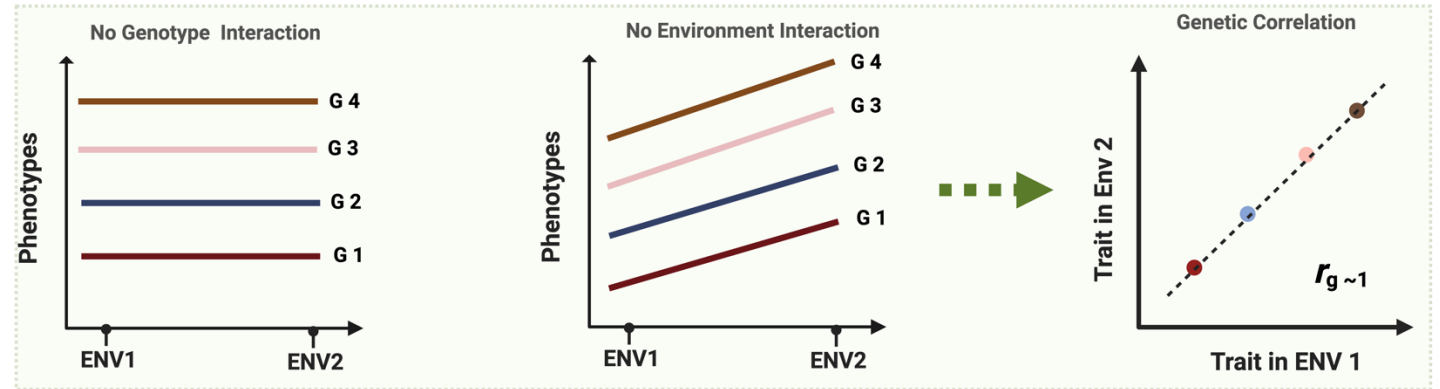
Module 3
August 6, 2025

Waseem Hussain and Mahender Anumalla
Rice Breeding Innovations Platform
IRRI

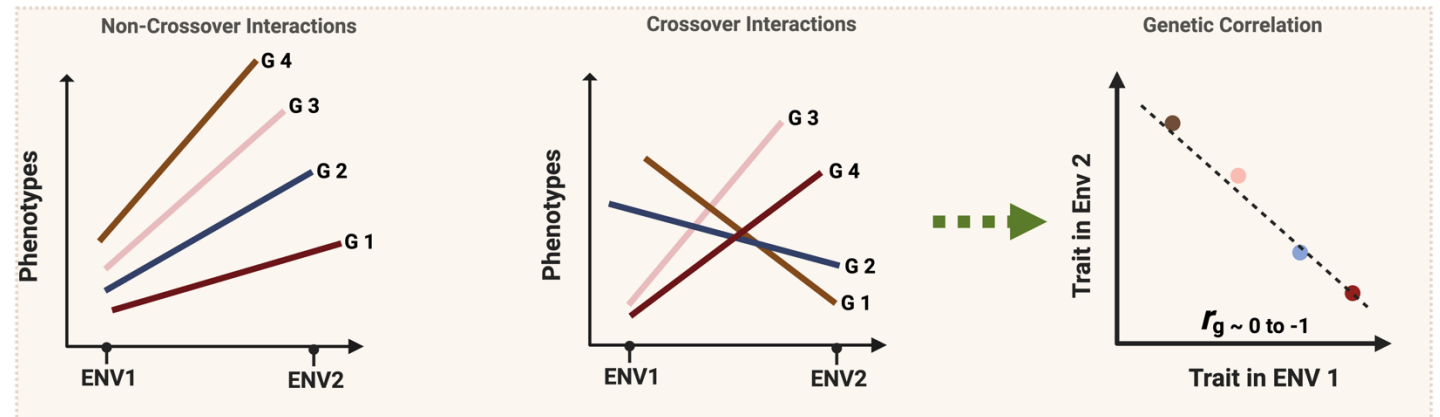
What is G x E Environment Interaction

- **Phenotypic Plasticity:** a changing the phenotypic expression due to change in environment
- Phenotypic plasticity: The range of phenotypes a single genotype can express as a function of its environment.
- When phenotypic plasticity differs between genotypes, this is described as **G x E**
- $G \times E$ is commonly defined as the joint effect of genetics and the environment that deviates from their individual additive effects

G x E Interaction Absent

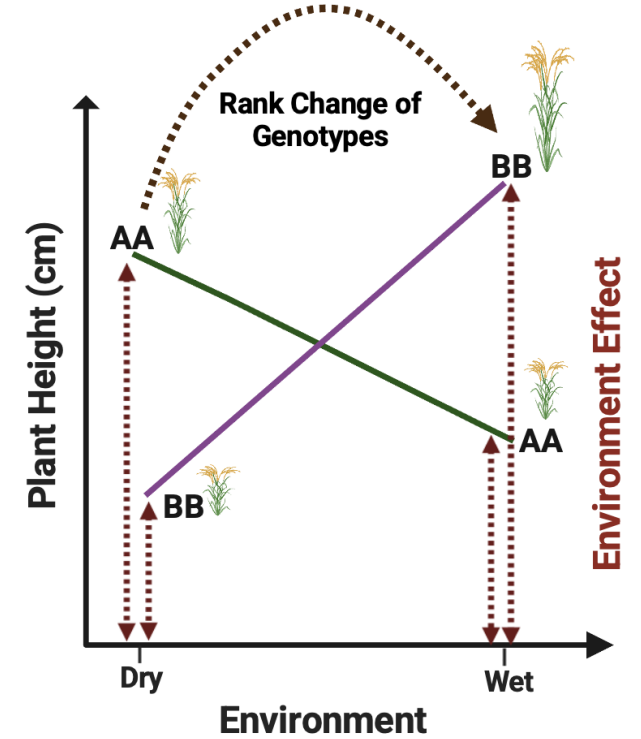
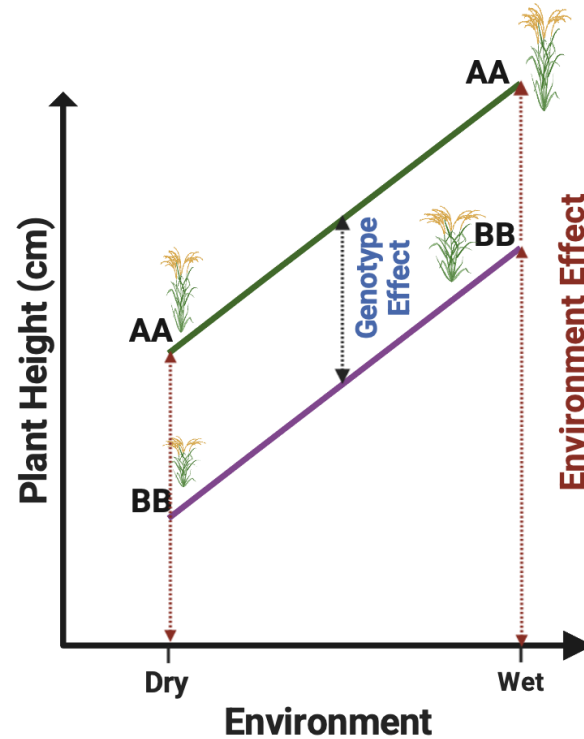


G x E Interaction Present



Why Multi-Environment Trials

- Assess the performance of genotypes across a range of environments
- Choose based on a Specific adaptation or broader adaptation (Stable genotype)
- We choose and assess based on Statistical Models



Basic Model of G x E

$$P = Genotype + Environment$$

$$Y_i = \mu + G_i + E_j + \epsilon_{ij} \text{ (Model1)}$$

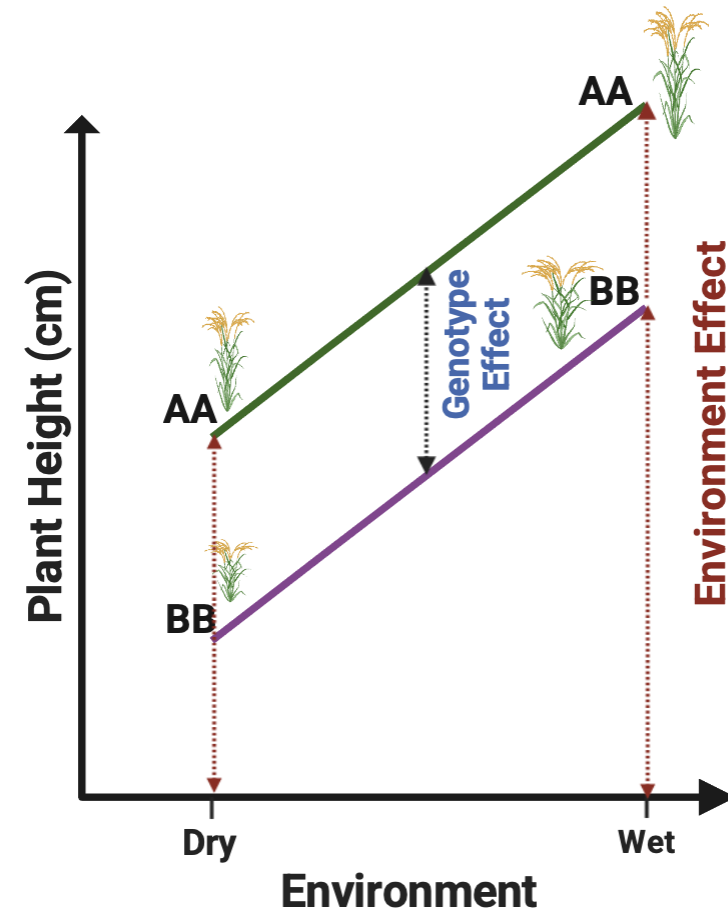
Where Y_i is the mean of genotype i in environment j

μ is the mean intercept (fixed)

G_i is fixed genotype effect of genotypes

E_j is the fixed environmental effect

And ϵ_{ij} is the error (random) with a mean 0 and constant variance $\epsilon_{ij} \sim N(0, \sigma^2)$



- The norms of reaction of genotypes will be parallel
- Genotype means change because of the Environment; the effect depends upon the environments
- Not suitable for Unbalanced data sets; mean will change

Basic Model of G x E

$$Y_i = \mu + G_i + E_j + GE_{ij} + \epsilon_{ij} \text{ (Model 2)}$$

Where Y_i is the mean of genotype i in environment j

μ is the mean intercept (fixed)

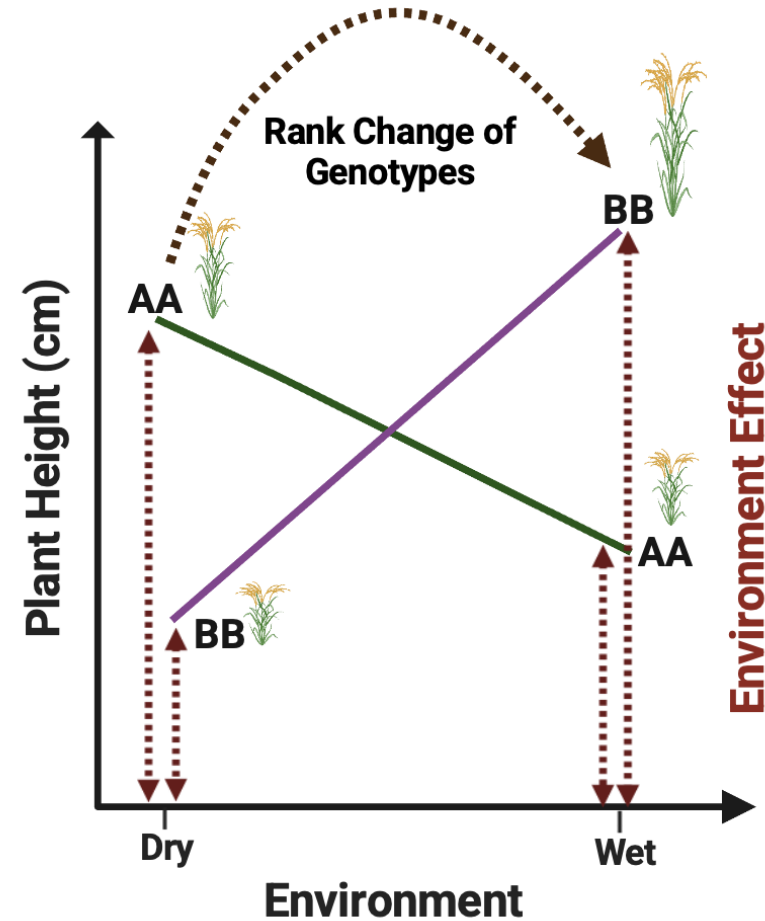
G_i is fixed genotype effect of genotypes

E_j is the fixed environmental effect

And ϵ_{ij} is the error (random) with a mean

0 and constant variance $\epsilon_{ij} \sim N(0, \sigma^2)$

- Specific for genotype-by-environment combinations, G x E
- Simply tell amount of genotype effects in relation to G x E



- As many parameters as combinations of genotypes x Environment
- Not possible to assess the performance of genotypes not included in trials

Alternative: Regression

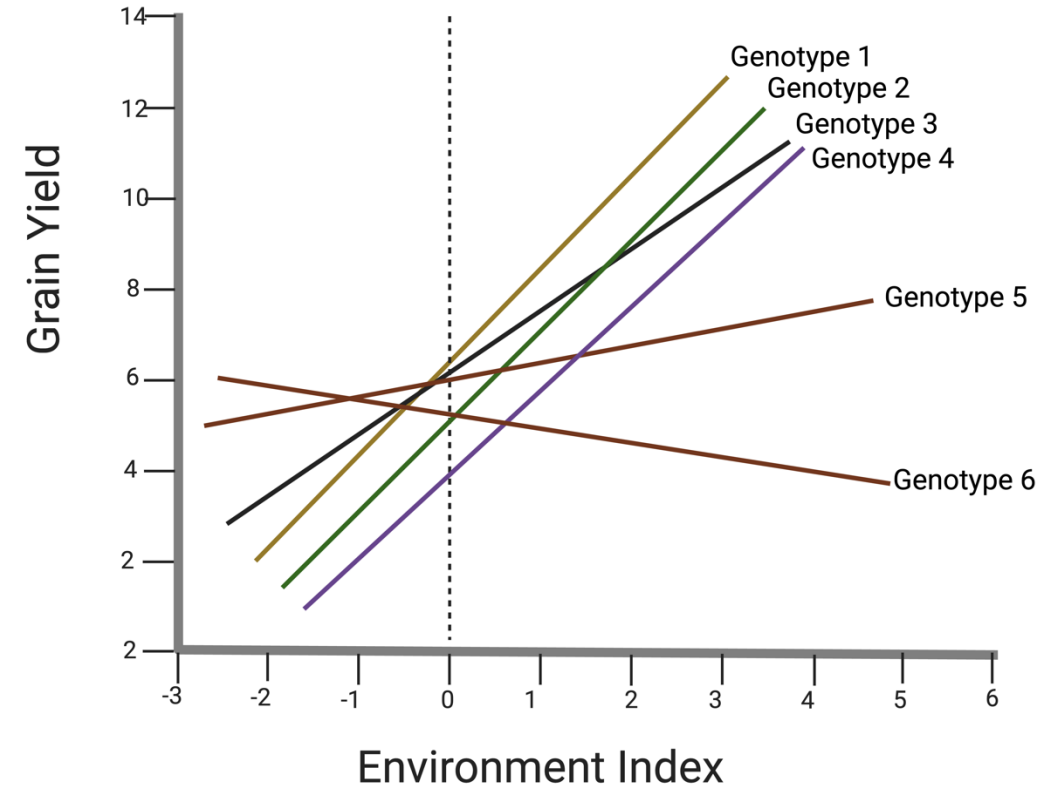
Model of G x E

$$Y_i = \mu + G_i + E_j + b_i E_j + \epsilon_{ij} \text{ (Model3)}$$

- Explains G x E as a set of parallel regression lines
- A popular one is Finlay and Wilkson (1963), which describes $G \times E$ as a regression line on Environmental Index or Quality
- The G x E is modeled as differential genotypic sensitivity, represented by the parameters β_j , to the environmental characterization E_j , **with the average sensitivity being zero.**

How Does it Work

- Good environments have high average genotype performance
- Bad environments would have low average genotype performance
- The phenotypic response of individual genotypes is regressed on the average performance, and G x E expresses itself by the difference in the slopes.



Alternative: Regression

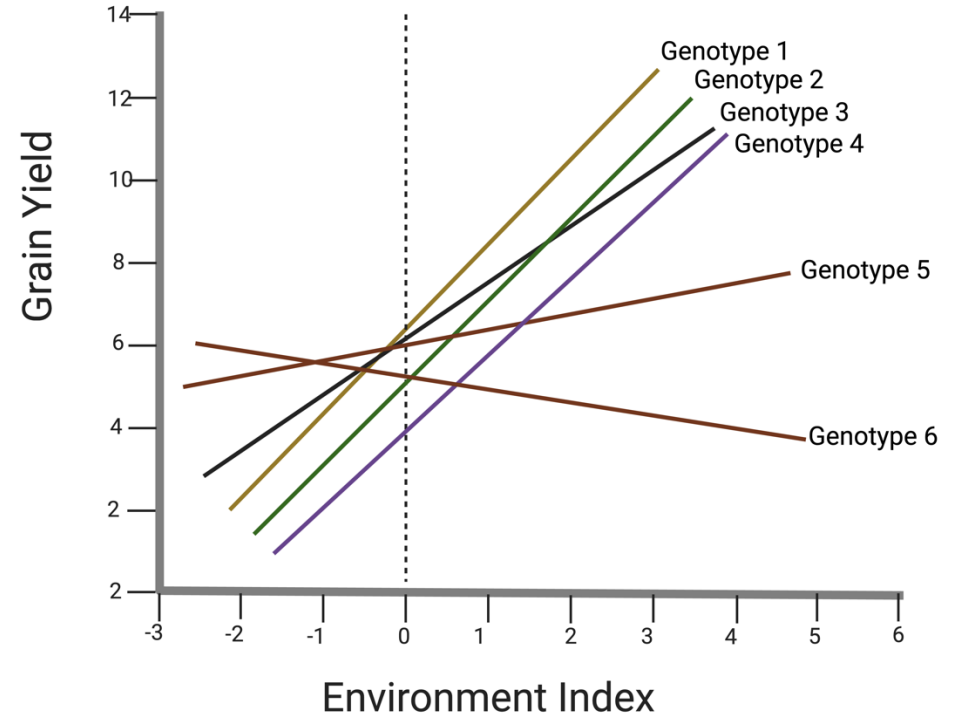
Model of G x E

$$Y_i = (\mu + G_i) + (1 + b_i)E_j + \epsilon_{ij} = G_i + \beta_j E_j \text{ (Model 4)}$$

- Now, average β_j is not zero in Model 4
- In general, when β_j is zero or 1, the regression model reduces to Additive Model

Limitations of Regression Models

- Environmental characterization is based on a single dimension.
- Environmental quality can be hard to summarize within a single explanatory variable
- A substantial amount of GEI can remain unexplained.



Alternative: Multidimensional G x E

(Flexibility to add more than one environment variable and popular)

Additive Main Effects and Multiplicative interaction (AMMI)

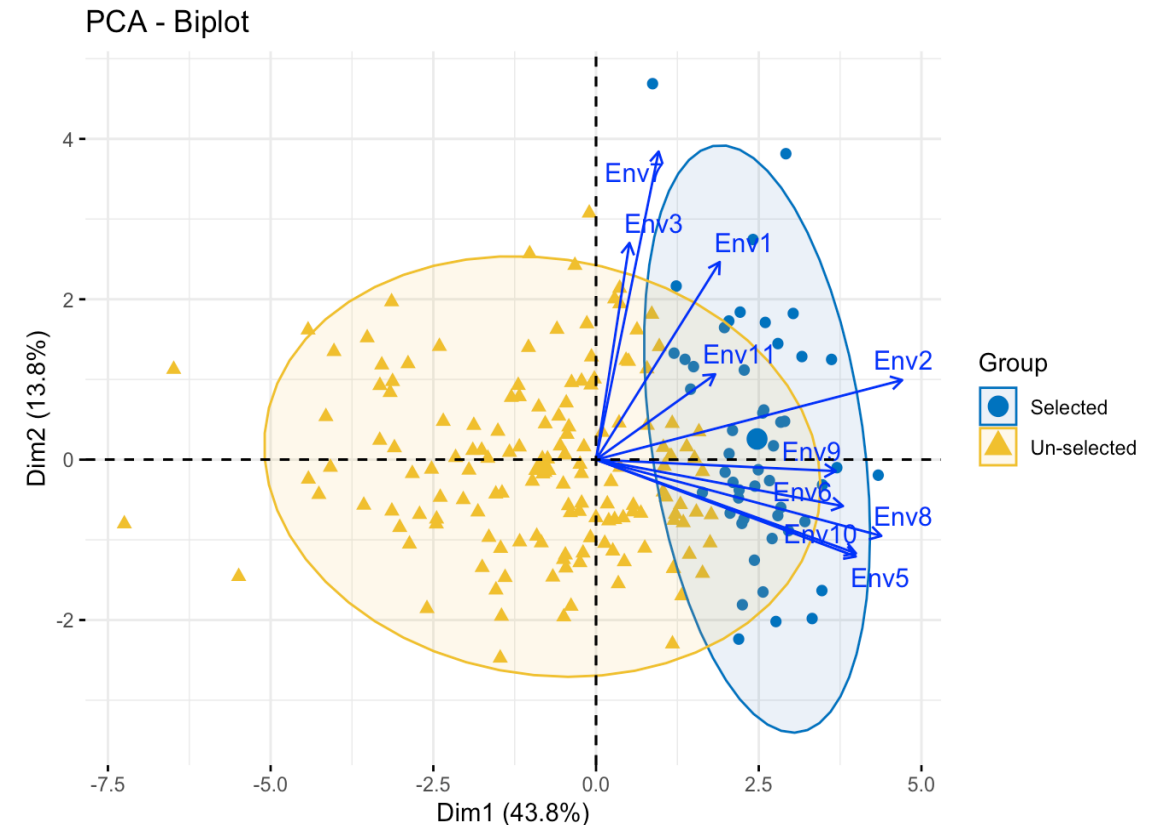
$$Y_i = \mu + G_i + E_j + \sum_{k=1}^K a_{ki} b_{kj} + \epsilon_{ij} \text{ (Model 5)}$$

- where the G x E is now explained by K multiplicative terms ($k=1 \dots K$),
- Each K is a product of a genotypic sensitivity a_{ik} (genotypic score) and a hypothetical environmental characterization b_{ik} (environmental score).

a_{ik} and b_{ik} come from the Principal Component Analysis (PCA)

PCA maximizes the variation explained by -
product of a_{ik} and b_{ik} ; with PCA1 >
PCA2...>PCA k

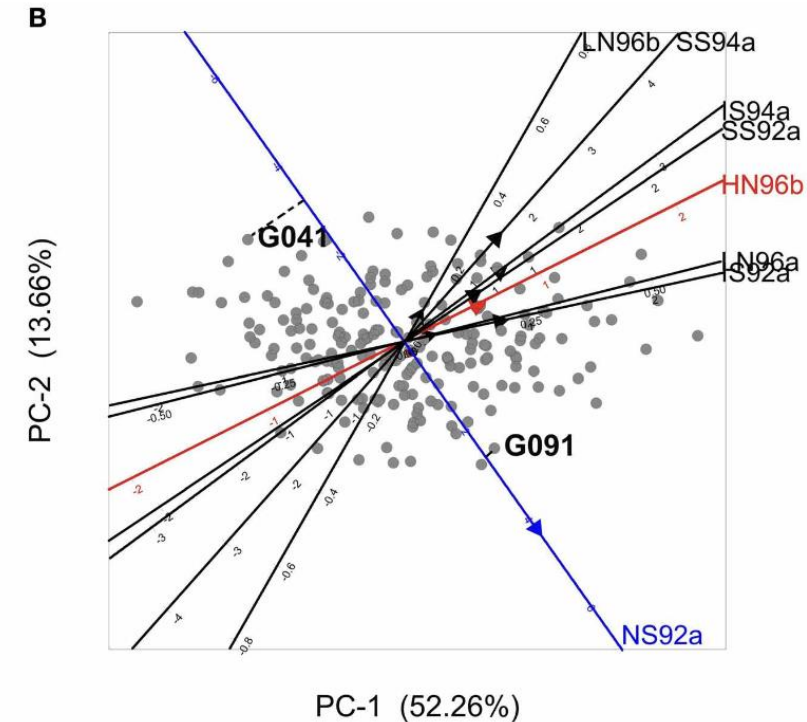
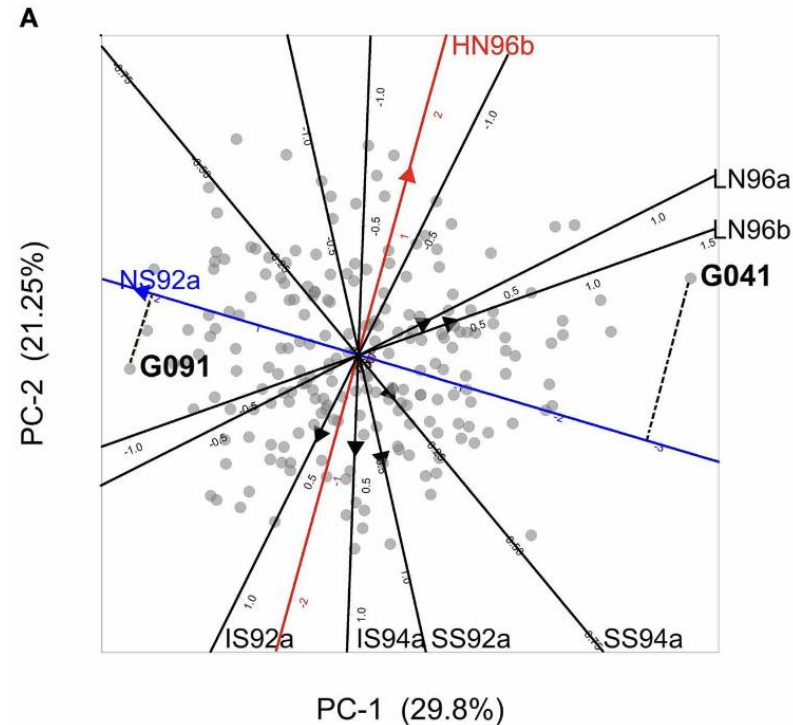
A desirable feature of the AMMI model is that the genotypic and environmental scores can be used to construct powerful graphical representations called **biplots** (Gabriel, 1978)



Alternative: GEI or GGE Models

Joint modelling of Genotypic and G x E main effects

- Breeders are interested in both genotypic and G x E interactive Effects, which is why it is called GGE or GGI Models.
- Biplots of GEI or GGE are called GGE **biplots** (Gabriel, 1978)



- GGE biplots approximate overall performance (G + GEI).
- Whereas AMMI biplots approximate only the GEI part of the phenotype

Alternative: Factorial Regression Models

What if we have Environmental Covariables, Rainfall, Temperature etc

$$Y_i = \mu + G_i + E_j + \sum_{k=1}^K ac_{ki}bc_{kj} + \epsilon_{ij} \text{ (Model 5)}$$

- if we have explicit information about the environment, the data can be used directly in the model by including it as explanatory variables.
- GEI is then described as differential genotypic sensitivity to explicit environmental factors such as temperature, precipitation, water availability etc.
- Such models are known as Factorial Regression Models

$ac_{ki}bc_{kj}$ represents an explicit environmental covariable and not a hypothetical environmental covariable represented by AMMI or GGI Models

FRM help to identify genotypes that are differentially sensitive to changes in identified environmental quality components, for example, in a particular nutrient or in water availability.

Mixed Effect Model

Modeling variances and Covariance instead of Just mean

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

Model focuses on the modeling of GEI in terms of heterogeneity of variances and covariances

Y is the Response variable

X is a $n \times m$ design matrix relating phenotypes to fixed effects

β is a vector of fixed effects

Z is a design matrix of random effects.

u is a vector of random effects with $\text{Var}(\mathbf{u}) \sim \sigma_g^2 \mathbf{G}$, where **G** is Kinship matrix

ε is $n \times n$ matrix of residual/error effects with $\text{Var}(\boldsymbol{\varepsilon}) = \sigma_e^2 \mathbf{I}$, **I** is identity matrix.

$$\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 \mathbf{K} & 0 \\ 0 & \sigma_\varepsilon^2 \mathbf{R} \end{pmatrix} \right)$$

Note: $\text{cov}(\mathbf{u}, \boldsymbol{\varepsilon}) = 0$

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, E(\mathbf{u}) = \mathbf{0}, E(\boldsymbol{\varepsilon}) = \mathbf{0}$$

- More flexibility to handle the unbalanced and missing data
- Powerful to use performance information from all known relatives to Estimate Breeding values (co-variances)
- Fixed effects can be controlled and accounted to infer rightly about the predicted values
- Overcomes limitations of ANOVA

Mixed Effect Model

Modeling variances and Covariance instead of Just mean

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

Y is the Response variable

X is a $n \times m$ design matrix relating phenotypes to fixed effects

β is a vector of fixed effects

Z is a design matrix of random effects.

u is a vector of random effects with $\text{variance} = \text{Var}(\mathbf{u}) \sim \sigma_g^2 \mathbf{G}$, where **G** is Kinship matrix

ε is $n \times n$ matrix of residual/error effects with $\text{Var}(\boldsymbol{\varepsilon}) = \sigma_e^2 \mathbf{I}$, **I** is identity matrix.

$$\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 \mathbf{K} & 0 \\ 0 & \sigma_\varepsilon^2 \mathbf{R} \end{pmatrix} \right)$$

Note: $\text{cov}(\mathbf{u}, \boldsymbol{\varepsilon}) = 0$

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, E(\mathbf{u}) = \mathbf{0}, E(\boldsymbol{\varepsilon}) = \mathbf{0}$$

- **Model 1 and model 2:** same variance (homogeneous)
- **Model 3:** Different variances across environments (heterogeneous)
- **Model 4:** Same as Model 3 with Spatial variations.
- **Model 5:** Spatial variations specific to each environment.
- **Model 6:** Heterogeneous in variance but no correlation between environments
- **Model 7:** Uniform correlation between environments
- **Model 8 and 9:** Factor analytical models with k multiplicative components
- **Model 10:** Factor analytical model with Spatial variations specific to each environment.

Mixed Effect Model



```
graph TD; A[Mixed Effect Model] --> B[Single Stage Analysis]; A --> C[Two Stage Analysis]
```

Single Stage Analysis

Each trial or environment is analyzed separately

Stage-wise analysis is more appropriate.

- Trials with unbalanced data sets,
- Different experimental design factors across trials,
- Avoid the computational challenges of analyzing a huge number of trials.

Two Stage Analysis

All trials combined and analyzed together

Commentary | [Open access](#) | Published: 05 February 2022

Open-source analytical pipeline for robust data analysis, visualizations and sharing in crop breeding

[Waseem Hussain](#) , [Mahender Anumalla](#), [Margaret Catolos](#), [Apurva Khanna](#), [Ma. Teresa Sta. Cruz](#), [Joie Ramos](#) & [Sankalp Bhosale](#)

[Plant Methods](#) **18**, Article number: 14 (2022) | [Cite this article](#)

8126 Accesses | **2** Citations | **27** Altmetric | [Metrics](#)

Abstract
