Fundamentals of Genomic Prediction and Data-Driven Crop Breeding (November 24-28, 2025)

Genomic Selection Workflow: Example and Implementation

**Module 4
November 27, 2025**

**Waseem Hussain, Mahender Anumalla, Margaret Catolos
Rice Breeding Innovations Platform
IRRI**

General Breeding Pipeline

Parental Pool

P1 x P2......P3 x P4.....P1 x P4

Segregating Generations

$F_2 - F_3 - F_4 - F_5$

Fixed Generation

F_6

Multilocation Evaluation

F_7

(884 Genotypes)

Multi-location Evaluation

Final Selections

Seed System (2-3)
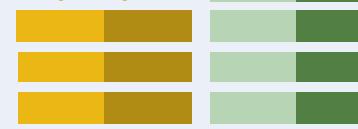
Phenotypic Selection

Option 1

Location 1 (884)   Location 2 (884)   Location 3 (884)

Field Evaluation across Locations

What we assume here

	G1	G2	G3	...	....	GN
G1	1	0	0	0	0	0
G2	0	1	0	0	0	0
G3	0	0	1	0	0	0
.	0	0	0	1	0	0
.	0	0	0	0	1	0
Gn	0	0	0	0	0	1

Perform Analysis
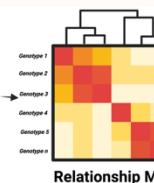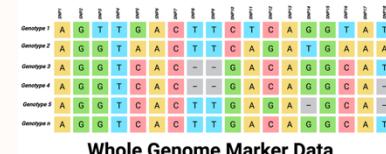
BLUEs and BLUPS

Genotypes are independent (No-covariance)

Final Selections

Genomic Selection

All 844 GENOTYPES

Whole Genome Marker Data   Relationship Matrix

Option 2

Training Set only

Location 1 (254)   Location 2 (254)   Location 3 (254)
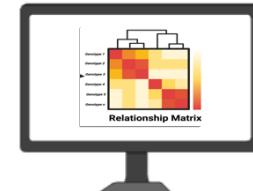
Limited Resources
Estimate Breeding Value
Perform Predictions
Dissect G x E better

Field Evaluation across Locations

Perform Analysis

Relationship Matrix

Leverage covariances (Relative Information)

Genotypes are Related
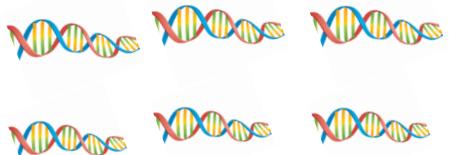
(Predict un-tested)

BLUPS

Final Selections

Steps of Implementing Genomic Selection
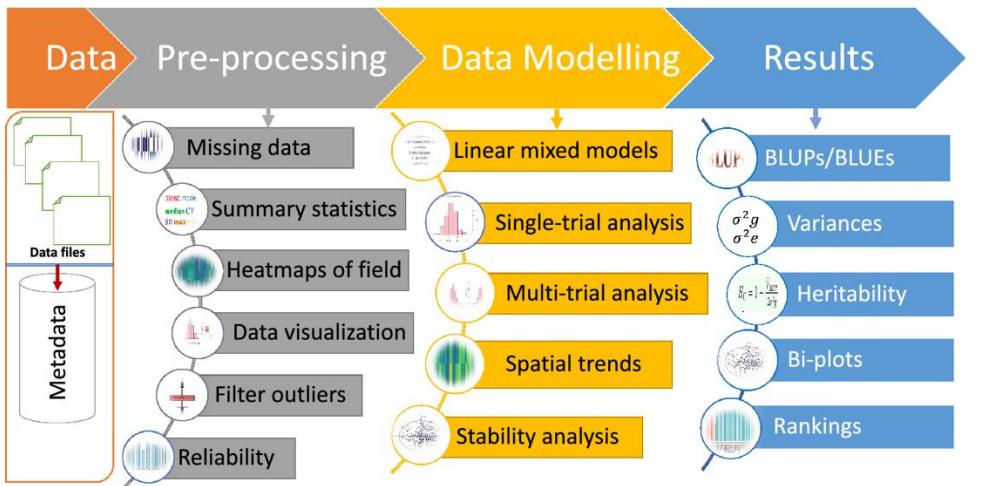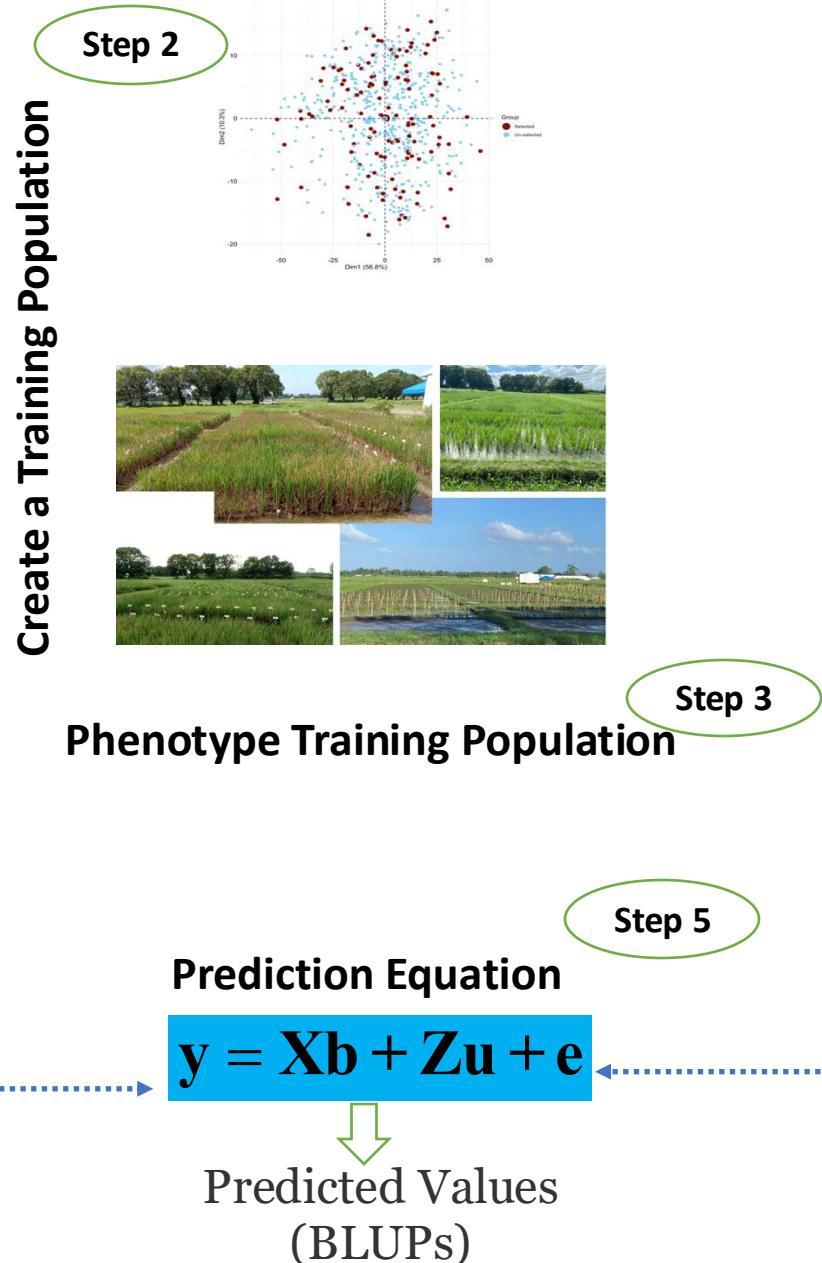
Whole F$_7$ Population

Genotype all Population

Step 1

Un-tested Population

Create a Training Population

Step 2

Step 3
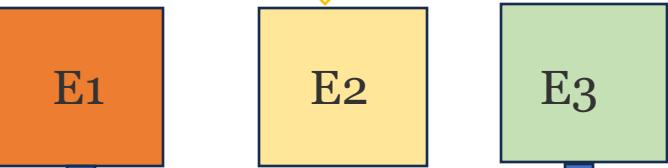
Phenotype Training Population

Pre-processing and Quality Check

✓ Calling rate (genotype and marker)
✓ Minor allele frequency
✓ Heterozygosity

Pre-processing and Data Analysis   Step 4

Step-Wise Analysis

E1   E2   E3

Extract BLUES Per Environment

Prediction Equation   Step 5

$$y = Xb + Zu + e$$

Predicted Values
(BLUPs)

Model G x E also

Data   Pre-processing   Data Modelling   Results

Missing data   Linear mixed models   BLUPs/BLUEs

Data files   Summary statistics   Single-trial analysis   Variances

Heatmaps of field   Multi-trial analysis   Heritability

Metadata   Data visualization   Spatial trends   Bi-plots

Filter outliers   Stability analysis   Rankings

Reliability

Genomic Selection Approach

One Step Approach

$$y = X_1\beta_1 + X_1\beta_1 \ldots + Z_1u_1 \ldots \ldots Z_2u_2 + \varepsilon..$$

Fit all in one Model
(Design, Environment, and Genotype factors)
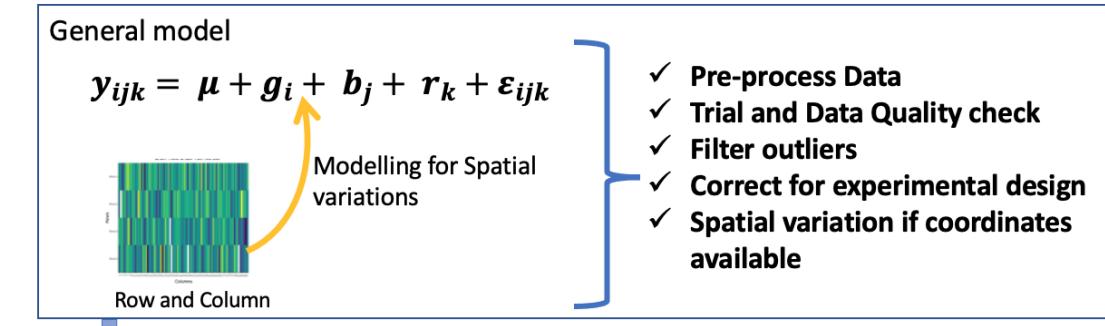
Can We Extract BLUP in Step 1 and Fit BLUPS again in Step 2?

https://doi.org/10.1093/g3journal/jkae250
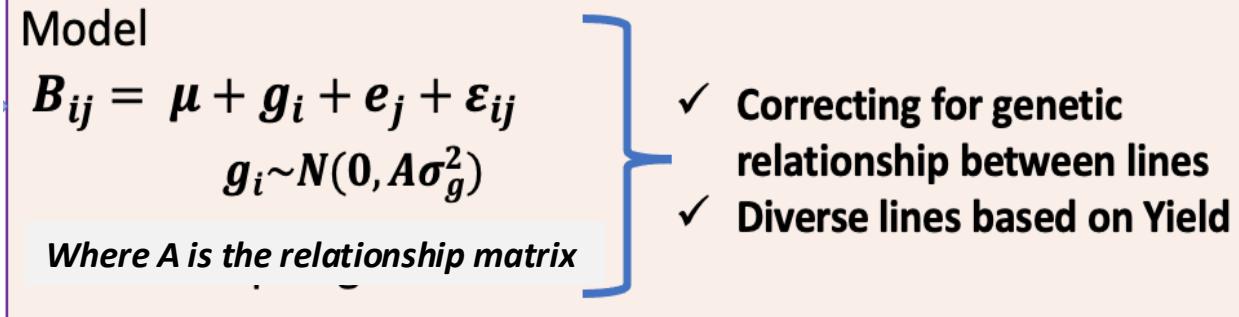
Double Shrinkage! Wrong Approach

Two Step

Stage 1: Extract BLUEs per Environment

General model

$$y_{ijk} = \mu + g_i + b_j + r_k + \varepsilon_{ijk}$$

Modelling for Spatial variations

Row and Column

- ✓ Pre-process Data
- ✓ Trial and Data Quality check
- ✓ Filter outliers
- ✓ Correct for experimental design
- ✓ Spatial variation if coordinates available

Stage 2: Extract the Breeding Values (BLUPs)

Model

$$B_{ij} = \mu + g_i + e_j + \varepsilon_{ij}$$

$$g_i \sim N(0, A\sigma_g^2)$$

Where A is the relationship matrix

- ✓ Correcting for genetic relationship between lines
- ✓ Diverse lines based on Yield

Example of IRRI's Breeding Program

Let us run it in R together!