



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

## Dissecting Mixed Models (BLUEs, BLUPs and Breeding Values)

**Module 2**

**November 25, 2025**

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

# Mixed Model

*A model  
containing both  
Fixed and Random  
component*

## Fixed Effect

1. Variation in fixed effects for subjects are same. Inferences based on parameters.
2. Effects are fixed when the study focuses on those individual effects.
3. When the entire population of interest is included in the study, the effect is fixed
4. Values of fixed effects are estimated by least squares
5. Inferences cannot be extended to other data sets

## Random Effect

1. Variation in random effects for subjects vary
2. Effects are random when the study focuses on characteristics of an underlying population
3. When a small fraction of the population is included, the effect is random.
4. Values of random effects are estimated with shrinkage (i.e., best linear unbiased predictors).
5. Inferences cannot be extended to other data sets

*Degrees of freedom less than 5, use fixed?*

# Foundation of mixed models

---

**Ronald Fisher (1918)** *Infinitesimal model ( $P = G + E$ )*

---

**Francis Galton (1888)**  
*Regression and  $h^2$*

---

**Sewall Wright (1922)**  
*Pedigree matrix ( $A$ )*

---

**Charles Henderson (1950)**  
BLUPs and BLUEs

# Why Mixed Models

(Predict BLUPs/BV's and  
Variance Components)

---

Model phenotypic observations on related individuals which can provide information about their underlying genotypic values

---

Mixed model allows efficient estimation of genetic parameters (such as variance components, breeding values and heritability)

---

More flexibility to handle the unbalanced and missing data.

---

Powerful to use performance information from all known relatives to Estimate Breeding values (co-variances)

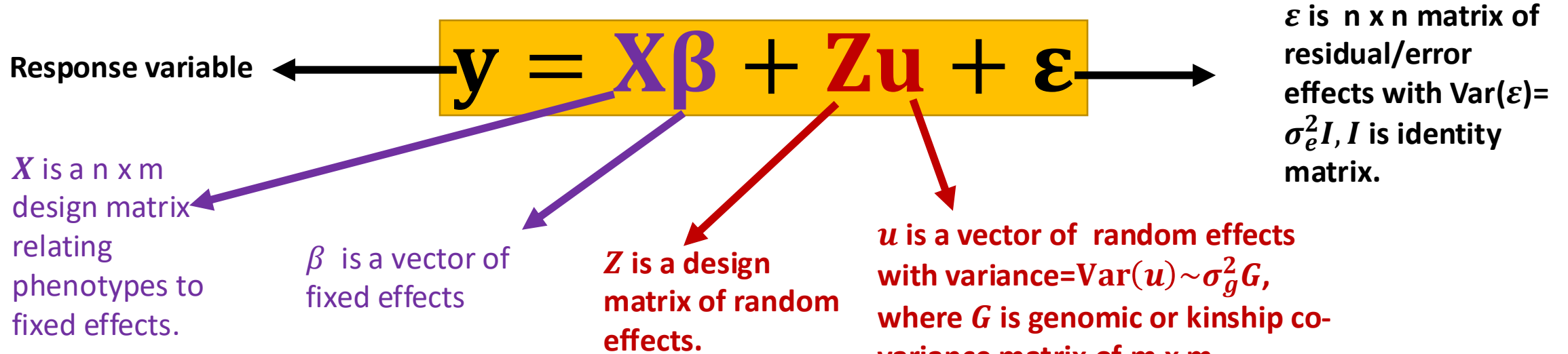
---

Fixed effects can be controlled and accounted to infer rightly about the predicted values

---

More control on nuisance effects blocks in case of Field Design

# Mixed Effect Model



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

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

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



# Mixed Effect Model

Example of RCBD design with 3 Replications and 4 genotypes

Replications as Fixed and Genotypes as Random

$$y = X\beta + Zu + \epsilon$$

$$\begin{bmatrix} 40 \\ 39 \\ 60 \\ 70 \\ 56 \\ 20 \\ 78 \\ 43 \\ 56 \\ 86 \\ 67 \\ 56 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} + [\epsilon]$$

Purpose is to get solution for  $\beta$  (BLUE) and  $u$  (BLUP) and estimate variance components

# Henderson (1950) Mixed model Equation

$$y = X\beta + Zu + \varepsilon$$

$$\begin{bmatrix} \acute{X}R^{-1}X & \acute{X}R^{-1}Z \\ \acute{Z}R^{-1}X & \acute{Z}R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} \acute{X}R^{-1}y \\ \acute{Z}R^{-1}y \end{bmatrix}$$

multiply  $R = \sigma^2 I$  on both sides

$$\begin{bmatrix} \acute{X}X & \acute{X}Z \\ \acute{Z}X & \acute{Z}Z + \lambda G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} \acute{X}y \\ \acute{Z}y \end{bmatrix}$$

Where,  $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_u^2}$  is shrinkage factor

BLUE  
BLUP

$\beta$  from mixed model is BLUE and equal to OLS and is given as  $\hat{\beta} = (\acute{X}V^{-1}X)^{-1} \acute{X}V^{-1}Y$

$u$  from mixed model is BLUP and is given as  $\hat{u} = G\acute{Z}V^{-1}(y - X\hat{\beta})$

where  $V = ZG\sigma_u^2\acute{Z} + R\sigma_\varepsilon^2$

# Solving Mixed model Equation

$$\underbrace{\begin{bmatrix} \acute{X}X & \acute{X}Z \\ \acute{Z}X & \acute{Z}Z + \lambda G^{-1} \end{bmatrix}}_{\text{LHS}} \begin{bmatrix} \beta \\ u \end{bmatrix} = \underbrace{\begin{bmatrix} \acute{X}y \\ \acute{Z}y \end{bmatrix}}_{\text{RHS}}$$

$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \acute{X}X & \acute{X}Z \\ \acute{Z}X & \acute{Z}Z + \lambda G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \acute{X}y \\ \acute{Z}y \end{bmatrix}$$

*Restricted maximum likelihood (REML) decomposes the matrices and get best solutions of parameters (BLUPs and BLUEs) and variance components through iterative methods (convergence)*



# Dissect the Mixed model Equation

N	Y	Rep	Gen
1	40	R1	Ge1
2	39	R1	Ge2
3	60	R1	Ge3
4	70	R1	Ge4
5	56	R2	Ge1
6	20	R2	Ge2
7	78	R2	Ge3
8	43	R2	Ge4
9	56	R3	Ge1
10	86	R3	Ge2
11	67	R3	Ge3
12	56	R3	Ge4

## What is then X

N	Rep1	Rep2	Rep3
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
5	0	1	0
6	0	1	0
7	0	1	0
8	0	1	0
9	0	0	1
10	0	0	1
11	0	0	1
12	0	0	1

**X is 12 x 3 matrix**

Dissect the  
Mixed model  
Equation

## What is then Z

N	Gen1	Gen2	Gen3	Gen4
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
5	1	0	0	0
6	0	1	0	0
7	0	0	1	0
8	0	0	0	1
9	1	0	0	0
10	0	1	0	0
11	0	0	1	0
12	0	0	0	1

**Z is 12 x 4 matrix**

# Dissect the Mixed model Equation

## What is then $\hat{X}X$

$\hat{X}$  is 3 x 12 matrix

$X$  is 12 x 3 matrix

$\hat{X}X$  is 3 x 3 matrix



	Rep1	Rep2	Rep3
Rep1	4	0	0
Rep2	0	4	0
Rep3	0	0	4

3 x 3 matrix, counting number of phenotypes in each replication

## What is then $\hat{X}Z$

$\hat{X}$  is 3 x 12 matrix

$Z$  is 12 x 4 matrix

$\hat{X}Z$  is 3 x 4 matrix



	Gen1	Gen2	Gen3	Gen4
Rep1	1	1	1	1
Rep2	1	1	1	1
Rep3	1	1	1	1

3 x 4 matrix, counting number of genotype in each replication

# Dissect the Mixed model Equation

## What is then $\hat{Z}X$

$\hat{Z}$  is 4 x 12 matrix

$X$  is 12 x 3 matrix

$\hat{Z}X$  is 4 x 3 matrix

	Rep1	Rep2	Rep3
Gen1	1	1	1

Gen2	1	1	1
------	---	---	---

Gen3	1	1	1
------	---	---	---

Gen4	1	1	1
------	---	---	---

4 x 3 matrix, counting number of genotype in each environment

## What is then $\hat{Z}Z$

$\hat{Z}$  is 4 x 12 matrix

$Z$  is 12 x 4 matrix

$\hat{Z}Z$  is 4 x 4 matrix

	Gen1	Gen2	Gen3	Gen4
Gen1	3	0	0	0

Gen2	0	3	0	0
------	---	---	---	---

Gen3	0	0	3	0
------	---	---	---	---

Gen4	0	0	0	3
------	---	---	---	---

4 x 4 diagonal matrix counting the number of phenotypes observed for each genotype

## Dissect the Mixed model Equation

What is then  $\hat{Z}Z + \lambda G^{-1}$

$\hat{Z}$  is 4 x 12 matrix

$Z$  is 12 x 4 matrix

$\hat{Z}Z$  is 4 x 4 matrix

Assume  $G=1$  (no relationship)

$$\lambda = \frac{\sigma_{\varepsilon}^2}{\sigma_u^2} = 1.89/10 = 0.18$$

4 x 4 diagonal matrix counting the number of phenotypes observed for each genotype +  $\lambda$  value in the diagonal elements

	Gen1	Gen2	Gen3	Gen4
Gen1	3.18	0	0	0
Gen2	0	3.18	0	0
Gen3	0	0	3.18	0
Gen4	0	0	0	3.18

Dissect the  
Mixed model  
Equation

What is then LHS

$$\begin{bmatrix} \hat{X}X & \hat{X}Z \\ \hat{Z}X & \hat{Z}Z + \lambda G^{-1} \end{bmatrix}$$

	Rep1	Rep2	Rep3	Gen1	Gen2	Gen3	Gen4
Rep1	4	0	0	1	1	1	1
Rep2	0	4	0	1	1	1	1
Rep3	0	0	4	1	1	1	1
Gen1	1	1	1	3.18	0	0	0
Gen2	1	1	1	0	3.18	0	0
Gen3	1	1	1	0	0	3.18	0
Gen4	1	1	1	0	0	0	3.18



# Dissect the Mixed model Equation

## What is then $\hat{X}y$

$\hat{X}$  is 3 x 12 matrix

$y$  is 12 x 1

$\hat{X}y$  is 3 x 1 matrix



Rep1	233
Rep2	323
Rep3	343

3 x1 matrix counting the sum of phenotypes in each environment

## What is then $\hat{Z}y$

$\hat{Z}$  is 4 x 12 matrix

$y$  is 12 x 1

$\hat{Z}y$  is 4 x 1 matrix



Gen21	110
Gen2	202
Gen3	303
Gen4	403

4 x1 matrix counting the sum of phenotypes for each genotype

## Dissect the Mixed model Equation

What is then RHS

$$\begin{bmatrix} \acute{X} & y \\ \acute{Z} & y \end{bmatrix}$$

Rep1	101
Rep2	202
Rep3	302
Gen1	332
Gen2	231
Gen3	432
Gen4	434

# Solutions

$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \acute{X}X & \acute{X}Z \\ \acute{Z}X & \acute{Z}Z + \lambda G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \acute{X}y \\ \acute{Z}y \end{bmatrix}$$

RHS
LHS

Rep1	-20	}	BLUES
Rep2	-43		
Rep3	12		
Gen1	-65	}	BLUPs
Gen2	-80		
Gen3	32		
Gen4	20		

$$\begin{bmatrix} \acute{X}X & \acute{X}Z \\ \acute{Z}X & \acute{Z}Z + \lambda G^{-1} \end{bmatrix}$$

Off-diagonal elements are replaced with values of G matrix

	Rep1	Rep2	Rep3	Gen1	Gen2	Gen3	Gen4
Rep1	4	0	0	1	1	1	1
Rep2	0	4	0	1	1	1	1
Rep3	0	0	4	1	1	1	1
Gen1	1	1	1	3.18	0.67	0.23	0.11
Gen2	1	1	1	0.67	3.18	0.99	0.89
Gen3	1	1	1	0.23	0.99	3.18	0.50
Gen4	1	1	1	0.11	0.89	0.50	3.18

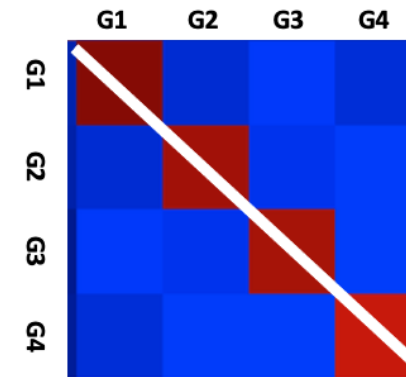
What  
when we  
know the  
G the

# Construction of G matrix

3 steps

1. Create a centered Z matrix
2. Create the cross product
3. Divide it by  $\sum_{i=1}^m 2p_j(1-p_j)$

$$G = \frac{ZZ^T}{\sum_{i=1}^m 2p_j(1-p_j)}$$



*First G matrix, VanRaden (2008)*

$\sigma^2 a$  is additive variance =  $\sum_{i=1}^m 2p_j(1-p_j)$

# Covariance Between Individuals

$$\text{Cov}(X, Y) = 2f_{xy}\sigma_a^2 (\text{assuming Dominance} = 0)$$

Where  $f_{xy}$  is the coefficient of co-ancestry between two individuals and  $\sigma_a^2$  is the additive genetic variance

## Example of Rice

	IRRI 154	IRRI 147	IRRI 219	IRRI 220
IRRI 154	1	2/3	2/16	2/18
IRRI 147		1	2/16	2/18
IRRI 219			1	2/18
IRRI 220				1

Note:

- If genotypes are independent off-diagonal elements=0
- Non-zero elements of off-diagonal reflect use of **relative information for BLUP estimation**

$$v(u) = A\sigma_a^2$$

Elements of **the additive relationship matrix (A)** has the elements given by the  $u(x, y) = 2f_{xy}$



# BLUEs, BLUPs, and Breeding Values

<b>Best</b>	Best solutions with minimum variance
<b>Linear</b>	Solutions are a linear combination of the observations
<b>Unbiased</b>	Low bias equal to their true values

BLUEs	BLUPs	Breeding Values
Associated with Fixed Effect	Associated with Random Effect	Associated with Random Effect, BLUPs of BV is called <b>Estimated Breeding Value</b>
$BLUE_i = y_i - \mu$	$BLUP_i = \frac{\sigma_a^2}{\sigma_a^2 + \frac{\sigma_\epsilon^2}{n}} y_i - \mu$	$BLUP \text{ of Breeding Value}_i = \frac{\sigma_a^2}{\sigma_a^2 + \frac{\sigma_\epsilon^2}{n}} y_i - \mu$
Values are simply average	Values are averaged but Shrunk, does not only depend upon the heritability but also <b>on information (n)</b> and effect error variance	Values are averaged but Shrunk, does not only depend upon the heritability but also on <b>information (n)</b> and effect error variance

BLUPs or BV's are always less than BLUEs, and depends upon the amount of Shrinkage

# BLUEs, BLUPs, and Breeding Values

---

$$\lambda = \frac{1 - h^2}{h^2}$$

- Less shrinkage with more observations
- Higher heritability, less shrinkage

BLUE is average sum of phenotypes in each environment =  $\frac{\text{sum}}{x}$

BLUE is simply computing averages

BLUP is the sum of phenotypes for each genotype divide by the number of phenotypes observed for each genotype =  $\frac{\text{sum}}{z + \lambda}$

Shrinkage towards 0, proportional to  $\lambda$

# Accuracy of Predictions (How Accurate we are!)

Accuracy is defined as the correlation between true and estimated breeding values

$$r_i = \rho(\hat{a}_i, a_i)$$

## Reliability and Prediction Error Variance (PEV)

Reliability is a squared correlation of accuracy and is **proportion of true genetic variance explained by the EBVs**

**Prediction error variance (PEV)**: fraction of additive variance not accounted for by the prediction

$$PEV = \text{var}(\hat{a}_i - a_i) = (1 - r_i^2)\sigma_a^2$$

➤ The closer the PEV to true values, the closer the reliability is to 1.

➤ PEV depends upon the n, the individuals with more information have small PEV

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + I\lambda \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

$$\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}^{-1} = \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & \mathbf{C^{22}} \end{pmatrix}$$

Now,  $PEV = \rho(\hat{a}_i, a_i) = C^{22}\sigma_e^2$

For each level of random effect or individual breeding value  $PEV_i = (d_i\sigma_e^2)$ ,

Where,  $d_i$  is the diagonal element of  $C^{22}$

# Accuracy of Predictions (How Accurate we are!)

=====

	Rep1	Rep2	Rep3	Gen1	Gen2	Gen3	Gen4
Rep1	4	0	0	1	1	1	1
Rep2	0	4	0	1	1	1	1
Rep3	0	0	4	1	1	1	1
Gen1	1	1	1	3.18	0	0	0
Gen2	1	1	1	0	3.20	0	0
Gen3	1	1	1	0	0	3.17	0
Gen4	1	1	1	0	0	0	3.15

$C^{22}$

Assuming  $\sigma_e^2 = 0.5$ ,  
Then,  $PEV =$

$\begin{bmatrix} 1.59 \\ 1.60 \\ 1.58 \\ 1.57 \end{bmatrix}$

and **Reliabilities** of predictions based on  $PEV$  is given by  $r_i^2 = 1 - \frac{PEV}{\sigma_u^2}$



# Literature to Read on Matrix

- <https://courses.lumenlearning.com/wmopen-collegealgebra/chapter/introduction-matrices-and-matrix-operations/>
- <https://courses.lumenlearning.com/boundless-algebra/chapter/introduction-to-matrices/>
- [https://math.libretexts.org/Bookshelves/Applied Mathematics/Book%3A Applied Finite Mathematics \(Sekhon and Bloom\)/02%3A Matrices/2.01%3A Introduction to Matrices](https://math.libretexts.org/Bookshelves/Applied_Mathematics/Book%3A_Applied_Finite_Mathematics_(Sekhon_and_Bloom)/02%3A_Matrices/2.01%3A_Introduction_to_Matrices)
- <https://byjus.com/maths/singular-matrix/>
- <https://www.onlinemathlearning.com/singular-matrix.html>