Module 3: G x E Interaction in Breeding and Predictive Genetics

Fundamentals of Genomic Prediction and Data-Drive Crop Breeding

(August 4-8, 2025)

Waseem Hussain
Senior Scientist-I
International Rice Research Institute
Rice Breeding Innovations Platfrom
waseem.hussain@irri.org
whussain2.github.io

Mahender Anumalla
Scientist-I
International Rice Research Institute
South-Asian Hub, Hyderabad
m.anumalla@irri.org

November 20, 2025

Contents

Load the Libraries

```
> # Load the Required Libraries
>   rm(list=ls()) # Remove previous work
>   library(lme4)
>   library(sommer)
>   library(AGHmatrix)
>   library(ggplot2)
>   library(DT)
>   library(statgenSTA)
>   library(statgenGxE)
>   library(metan)
>   library(readxl)
>   library(arm)
```

Data Set for Analysis

The data set is from IRRI Rainfed Rice Breeding Trials evaluated under 11 rainfed environments of Africa (ESA). The experiment was evaluated using Alpha Lattice Design with 4 Blocks, 2 Replications, 192 genotypes and 8 checks. Data has information on three traits of grain yield, plant height (HT) and days to flowering (DTF).

```
> # Read the data
>   demo.data<- read_excel("./Data/demo.data.xlsx", sheet ="Filtered")
> # Convert variables into appropriate data types
>   demo.data$Genotype<-as.factor(demo.data$Genotype) # Genotypes as factor
>   demo.data$Block<-as.factor(demo.data$Block) # Block as factor
>   demo.data$Row<-as.factor(demo.data$Row) # Row as factor
>   demo.data$Rep<-as.factor(demo.data$Rep) # Replication as factor
>   demo.data$Column<-as.factor(demo.data$Column) # Column as factor
>   demo.data$Environment<-as.factor(demo.data$Environment)  # Env. as factor
>   demo.data$Yield<-as.numeric(demo.data$Yield)  # Env. as factor
> # Order the factor levels
>   demo.data$Environment<- factor(demo.data$Environment,
+                       levels = c("Env1","Env2", "Env3",  "Env4", "Env5","Env6",
+                                  "Env7",  "Env8",  "Env9",  "Env10", "Env11"))
> # Visualize as table
>   kable(head(demo.data))
```

Environment	Genotype	Rep	Block	Row	Column	Line.type	DTF	HT	Yield
Env1	Genotype44	1	1	1	1	Entry	96	115.6	4956.395
Env1	Genotype131	1	1	1	2	Entry	89	116	5059.207
Env1	Genotype17	1	1	1	3	Entry	101	99	4948.038
Env1	Genotype146	1	1	1	4	Entry	92	102.8	6012.658
Env1	Genotype123	1	1	1	5	Entry	94	112.2	4456.759
Env1	Genotype116	1	1	1	6	Entry	98	108	4473.946

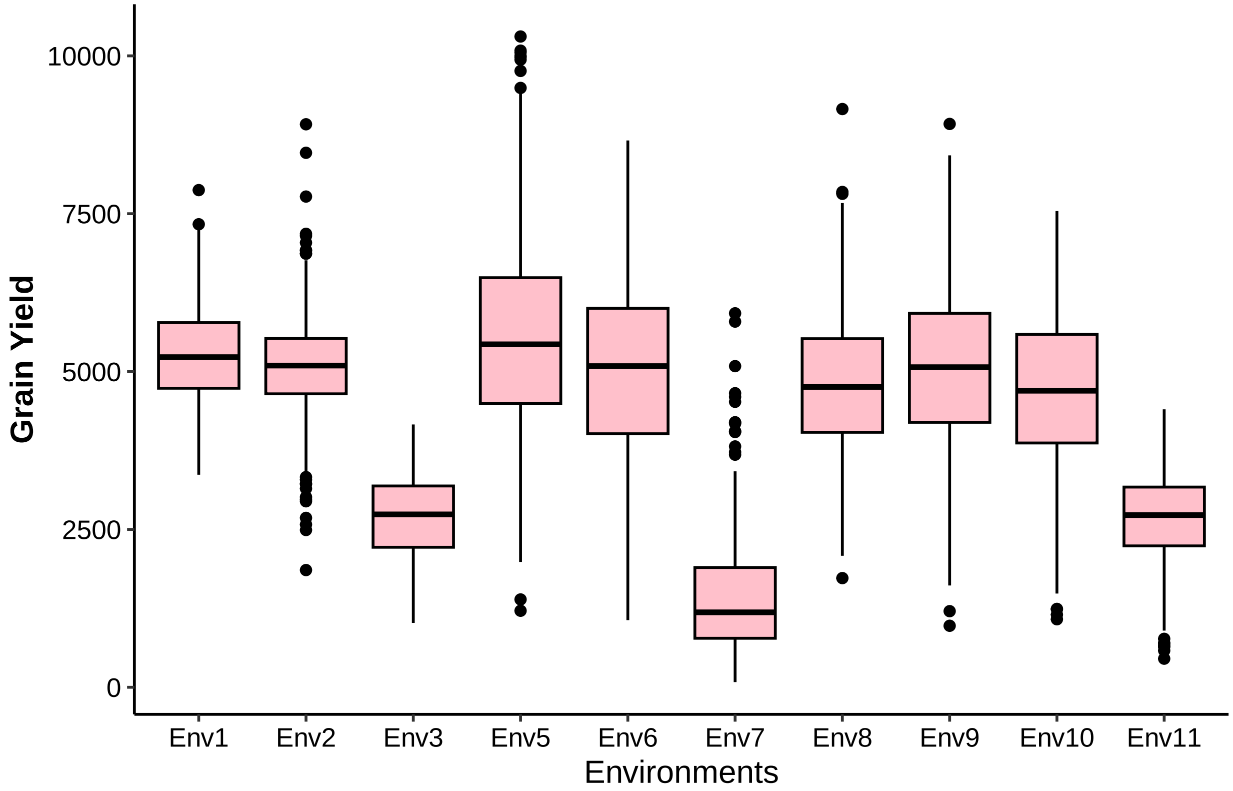
```
>   str(demo.data)
```

```
tibble [4,000 x 10] (S3: tbl_df/tbl/data.frame)
 $ Environment: Factor w/ 11 levels "Env1","Env2",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Genotype   : Factor w/ 200 levels "Genotype1","Genotype10",..: 140 37 79 53 28 20 196 83 108 18 ...
 $ Rep        : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
 $ Block      : Factor w/ 8 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Row        : Factor w/ 10 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Column     : Factor w/ 50 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Line.type  : chr [1:4000] "Entry" "Entry" "Entry" "Entry" ...
 $ DTF        : chr [1:4000] "96" "89" "101" "92" ...
 $ HT         : chr [1:4000] "115.6" "116" "99" "102.8" ...
 $ Yield      : num [1:4000] 4956 5059 4948 6013 4457 ...
```

Data Visualization

Here we will visualize the data as boxplot.

```
>   ggplot(demo.data, aes(x=Environment, y=Yield))+
+   geom_boxplot(color="black", fill="pink")+
+   # fill by timepoint to give different color
+   #scale_fill_manual(values = c("", ""))+
+   #scale_color_manual(values = c("", ""))+
+   theme_classic()+ #choose the theme for background
+   labs(title="",x="Environments", y = "Grain Yield")+#Add the labels to the plots
+   theme (plot.title = element_text(color="black", size=14,
+         face="bold",hjust=0.5), # add and modify the title to plot
+   axis.title.x = element_text(color="black", size=12),
+   axis.title.y = element_text(color="black", size=12,
+       face="bold")) + # add and modify title to y axis
+   theme(axis.text= element_text(color = "black", size = 10))+ # modify the axis text
+   theme(legend.position="none") # remove the theme from the plot
```



Basic ANOVA Model

Here we will use ***lm()*** function of R to get the ANNOVA for the terms

```
> # Run linear model using lm function of R
> model1<-lm(formula =
+ Yield~Genotype+Environment+Genotype*Environment+Environment:Rep+ Environment:Rep:Block,
+          data=demo.data)
> # Get ANNOVA using anova function
> anova(model1)
```

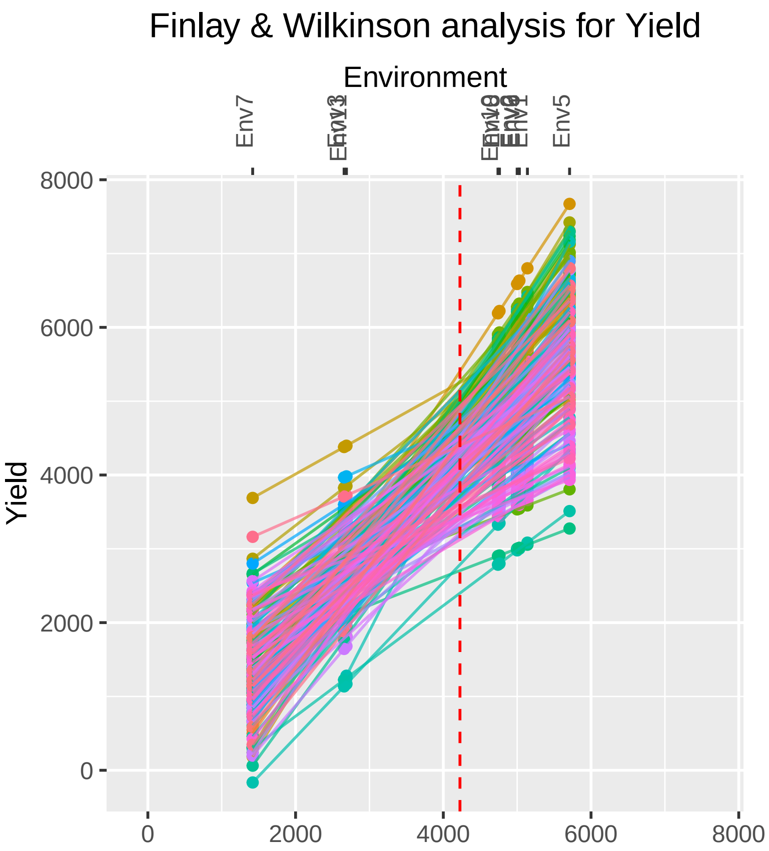
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Genotype	199	1185402954	5956798.8	14.585107	0.00e+00
Environment	9	7105461322	789495702.4	1933.065053	0.00e+00
Genotype:Environment	1779	3073244084	1727512.1	4.229780	0.00e+00
Environment:Rep	10	57746059	5774605.9	14.139012	0.00e+00
Environment:Rep:Block	76	57387089	755093.3	1.848831	1.72e-05
Residuals	1890	771907223	408416.5	NA	NA

```
> #test<-data.frame(model1$coefficients)
```

Regression Model

Here we will run popular Finlay and Wilkson (1963) regression model

```
> ## Create a TD object from dropsPheno.
>   dropsss<- statgenSTA::createTD(data =demo.data, genotype = "Genotype",
+                                   trial = "Environment")
> ## Perform a Finlay-Wilkinson analysis for all trials.
>   dropsFW <- gxeFw(TD = dropsss, trait = "Yield")
> ## Create line plot for Finlay Wilkinson analysis.
>   plot(dropsFW, plotType = "line")
```



AMMI Model

```
> # Developm AMMI Model
> Model_ammi<-performs_ammi(demo.data,
+               env = Environment,
+               gen = Genotype,
+               rep = Rep,
+               block=Block,
+               resp = Yield)
```

```
----------------------------------------------
Convergence information
----------------------------------------------
Number of iterations: 19
Final RMSE: 8.775333e-11
Number of axis: 1
Convergence: TRUE
----------------------------------------------
```

```
variable Yield
---------------------------------------------------------------------------
AMMI analysis table
---------------------------------------------------------------------------
         Source   Df   Sum Sq  Mean Sq F value    Pr(>F) Proportion Accumulated
            ENV    9 7.11e+09 7.89e+08 1933.07  0.00e+00         NA          NA
       REP(ENV)   10 5.77e+07 5.77e+06   14.14  2.06e-24         NA          NA
 BLOCK(REP*ENV)   76 5.74e+07 7.55e+05    1.85  1.72e-05         NA          NA
            GEN  199 1.19e+09 5.96e+06   14.59 9.87e-264         NA          NA
        GEN:ENV 1779 3.07e+09 1.73e+06    4.23 3.07e-196         NA          NA
            PC1  208 8.80e+08 4.23e+06   10.36  0.00e+00       28.3        28.3
            PC2  206 5.76e+08 2.79e+06    6.84  0.00e+00       18.5        46.8
            PC3  204 4.02e+08 1.97e+06    4.82  0.00e+00       12.9        59.7
            PC4  202 3.78e+08 1.87e+06    4.58  0.00e+00       12.2        71.9
            PC5  200 2.67e+08 1.34e+06    3.27  0.00e+00        8.6        80.5
            PC6  198 2.10e+08 1.06e+06    2.60  0.00e+00        6.7        87.2
            PC7  196 1.59e+08 8.11e+05    1.99  0.00e+00        5.1        92.3
 [ reached 'max' / getOption("max.print") -- omitted 5 rows ]
---------------------------------------------------------------------------

All variables with significant (p < 0.05) genotype-vs-environment interaction
Done!
```

```
> # ANNOVA
>   Model_ammi$Yield$ANOVA
```

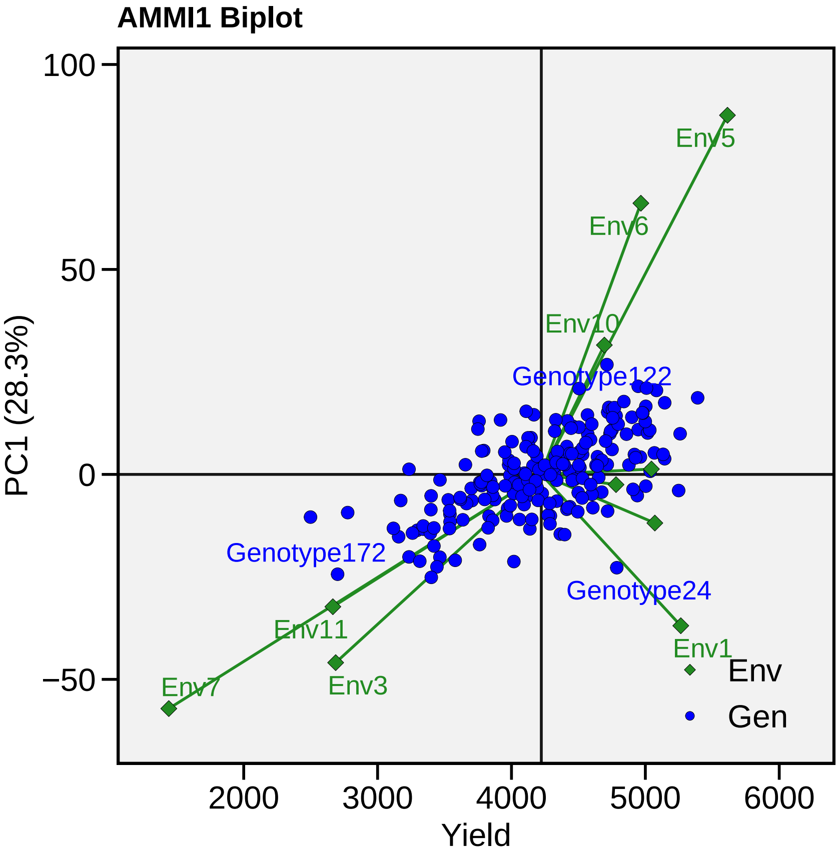
Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Proportion	Accumulated
ENV	9	7105461322	789495702.4	1933.065053	0.00e+00	NA	NA
REP(ENV)	10	57746059	5774605.9	14.139012	0.00e+00	NA	NA
BLOCK(REP*ENV)	76	57387089	755093.3	1.848831	1.72e-05	NA	NA
GEN	199	1185402954	5956798.8	14.585107	0.00e+00	NA	NA
GEN:ENV	1779	3073244084	1727512.1	4.229780	0.00e+00	NA	NA
PC1	208	879908129	4230327.5	10.360000	0.00e+00	28.3	28.3
PC2	206	575551396	2793938.8	6.840000	0.00e+00	18.5	46.8
PC3	204	401741898	1969323.0	4.820000	0.00e+00	12.9	59.7
PC4	202	378242194	1872486.1	4.580000	0.00e+00	12.2	71.9
PC5	200	267021758	1335108.8	3.270000	0.00e+00	8.6	80.5
PC6	198	209876756	1059983.6	2.600000	0.00e+00	6.7	87.2
PC7	196	158991524	811181.2	1.990000	0.00e+00	5.1	92.3
PC8	194	126188632	650456.9	1.590000	0.00e+00	4.1	96.4
PC9	192	111924983	582942.6	1.430000	2.00e-04	3.6	100.0
PC10	190	0	0.0	0.000000	1.00e+00	0.0	100.0
Residuals	1890	771907223	408416.5	NA	NA	NA	NA
Total	5953	15360596001	2580311.8	NA	NA	NA	NA

```
> # PCA Components
>   Model_ammi$Yield$PCA
```

PC	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Proportion	Accumulated
PC1	208	879908129	4230327.5	10.36	0e+00	28.3	28.3
PC2	206	575551396	2793938.8	6.84	0e+00	18.5	46.8
PC3	204	401741898	1969323.0	4.82	0e+00	12.9	59.7

PC	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Proportion	Accumulated
PC4	202	378242194	1872486.1	4.58	0e+00	12.2	71.9
PC5	200	267021758	1335108.8	3.27	0e+00	8.6	80.5
PC6	198	209876756	1059983.6	2.60	0e+00	6.7	87.2
PC7	196	158991524	811181.2	1.99	0e+00	5.1	92.3
PC8	194	126188632	650456.9	1.59	0e+00	4.1	96.4
PC9	192	111924983	582942.6	1.43	2e-04	3.6	100.0
PC10	190	0	0.0	0.00	1e+00	0.0	100.0

```
> # AMMI  Biplot
>   plot_scores(Model_ammi, type = 1)
```

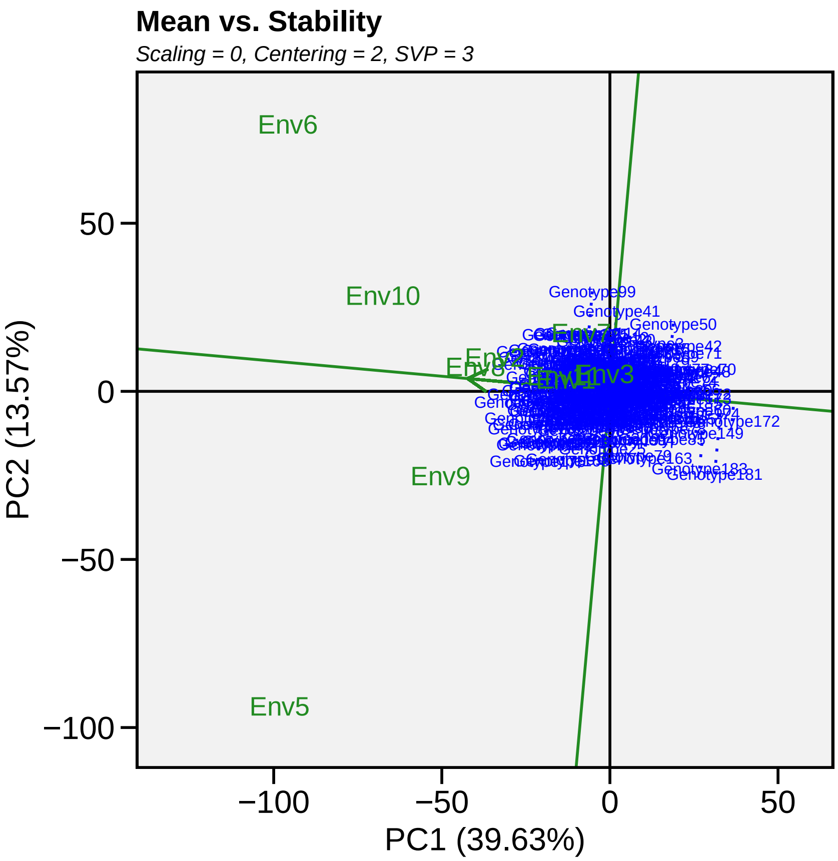


```
> # Stability indices
>   stab_indexes <- AMMI_indexes(Model_ammi)
>   head (stab_indexes$Yield)[1:5,1:7]
```

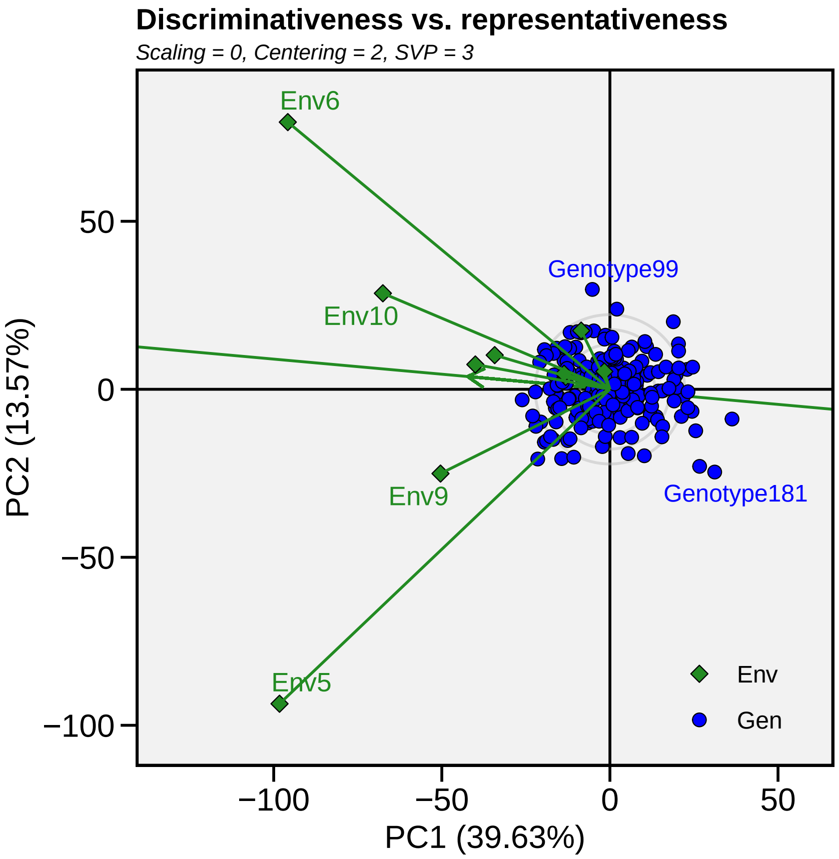
GEN	Y	Y_R	ASTAB	ASTAB_R	ssiASTAB	ASI
Genotype1	3971.519	141	321.1343	45	186	2.658307
Genotype10	3527.785	178	491.1672	90	268	2.297568
Genotype100	4718.930	40	830.1967	168	208	5.078426
Genotype101	4946.232	20	552.1578	108	128	3.219760
Genotype102	5003.387	14	1015.8038	186	200	5.386090

GGE Model

```
> gge_model <- gge(demo.data,
+                  env = Environment,
+                  gen = Genotype,
+                  resp = Yield,
+                  centering = 2,
+                  svp = "symmetrical",
+                  scaling = 0)
>
> # Mean performance vs. stability
>   plot(gge_model, type = 2, size.text.gen = 2)
```



```
> # Discriminativeness vs. representativeness
>   plot(gge_model, type = 4, size.text.gen = 3)
```



Mixed Model

Here we will run simple mixed model analysis using **lme4** R package. We will run the simple model accounting for Genotype, environment and G x E Interaction effects. For detailed analysis I suggested you check this **Manuscript** and my **[GitHub Page] (https://github.com/whussain2/Analysis-pipeline)**.

Mixed models are powerful tools to handle assumptions of linear model Read this one

```
> demo.data$Environment<-as.factor(demo.data$Environment)
> Model3.lme4<-lmer(Yield~Rep+(1|Genotype)+(1|Environment:Genotype)+
+           +(1|Environment:Rep:Block), data=demo.data)
```

Summary of MET results

In summary we will get following summarized results: 1) Description of model we used, 2) Random effects and variances, 3) Fixed effects, 4) Correlation of fixed effects

```
>   summary(Model3.lme4)
```

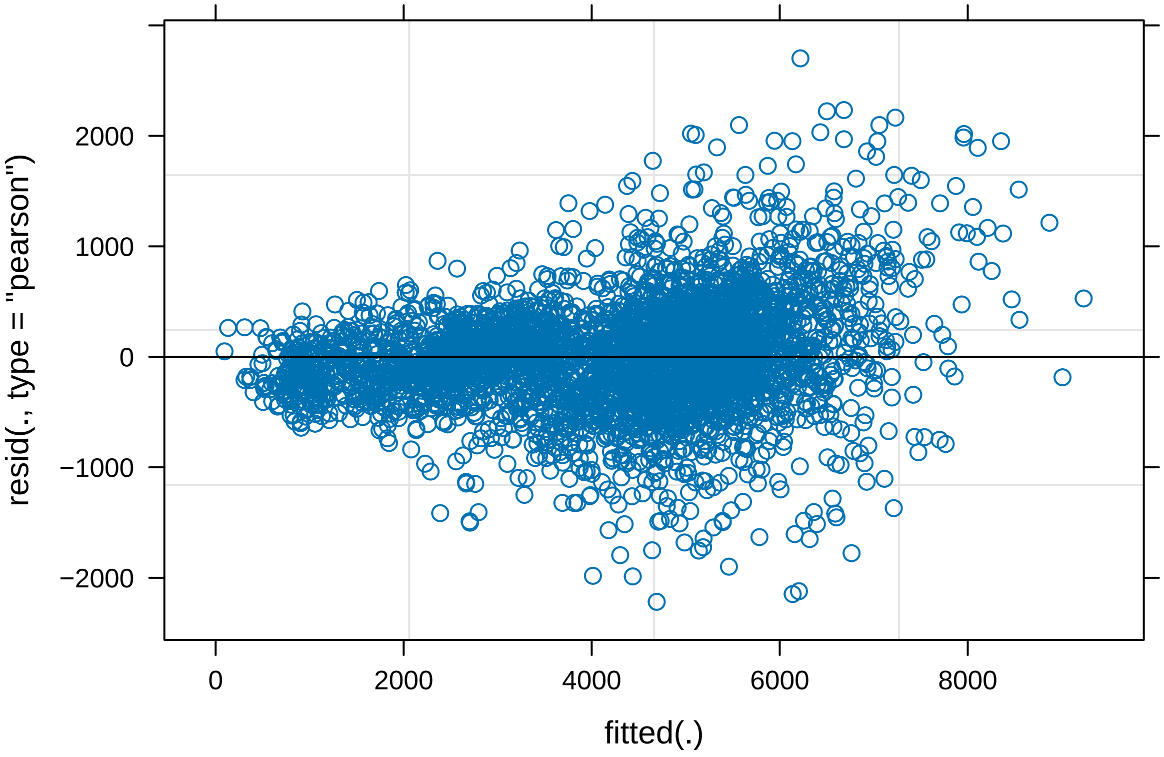
```
Linear mixed model fit by REML ['lmerMod']
Formula: Yield ~ Rep + (1 | Genotype) + (1 | Environment:Genotype) + +(1 |
    Environment:Rep:Block)
   Data: demo.data

REML criterion at convergence: 65988.1
```

```
Scaled residuals:
    Min      1Q  Median      3Q     Max 
-3.4691 -0.4245 -0.0182  0.4002  4.2264 

Random effects:
 Groups                Name        Variance Std.Dev.
 Environment:Genotype  (Intercept)  656050   810.0  
 Genotype              (Intercept)  209564   457.8  
 Environment:Rep:Block (Intercept) 1877587  1370.3  
 Residual                           408575   639.2  
Number of obs: 3964, groups:  
Environment:Genotype, 1988; Genotype, 200; Environment:Rep:Block, 96

Fixed effects:
            Estimate Std. Error t value
(Intercept)  4236.75     201.76  20.999
Rep2          -79.83     280.44  -0.285

Correlation of Fixed Effects:
     (Intr)
Rep2 -0.695
```

Plot of model

With the plot function model we will get the residuals vs fitted values

```
>   plot(Model3.lme4)
```



Extract the variance components

```
>   Ve<- data.frame (VarCorr(Model3.lme4))
>   Ve
```

grp	var1	var2	vcov	sdcor
Environment:Genotype	(Intercept)	NA	656049.7	809.9690
Genotype	(Intercept)	NA	209563.8	457.7814
Environment:Rep:Block	(Intercept)	NA	1877587.3	1370.2508
Residual	NA	NA	408575.1	639.1988

Heritability

- Here will estimate the combined heritability based on **Cullis et al.2006**

```
>   std.err<-se.ranef(Model3.lme4)$Genotype
>   v_BLUP<- mean(std.err)
> # Heritability/Reliability
>   h2<- (1-((v_BLUP)^2/(Ve[2,4]*2)))*100
>   h2
```

```
[1] 85.11792
```

BLUPs for Random Effects

```
>   BLUPs<-data.frame(BLUps.GY=coef(Model3.lme4)$Genotype[,1])
>   head(BLUPs)
```

BLUps.GY
4098.055
3766.068
4629.123
4728.669
4755.460
4306.043

Accounting Spatial Varaibility

```
> library(statgenSTA)
> TD_STA <- createTD(demo.data,
+                   trial = "Environment",
+                   genotype = "Genotype",
+                   rowCoord = "Row",
+                   colCoord = "Column",
+                   repId = "Rep",
+                   subBlock = "Block",
+                   trDesign = "res.ibd")
>
> ## Create layout plot with variety labels
> #plot(TD_STA,
>        # plotType = "layout",
>         #showGeno = TRUE)
> ## Model specification (using engine = "SpATS")
> sta_model_SpATS <- fitTD(TD = TD_STA,
+                     traits = "Yield",
+                     design = "res.ibd",
+                     what = "fixed",
+                     spatial = TRUE,
+                     engine = "SpATS")
>
> #plot(sta_model_SpATS,
>    #  plotType = "spatial",
>     # traits = 'Yield')
>
> ## Extract all available statistics from the fitted model.
>   extr <- extractSTA(sta_model_SpATS)
> ## Extract only the BLUEs from the fitted model.
>   BLUEs <- extractSTA(sta_model_SpATS,
+                     what = "BLUEs")
```

For any suggestions or comments, please feel to reach at waseem.hussain@cgiar.org; m.anumalla@cgiar.org; m.catolos@cgiar.org

---