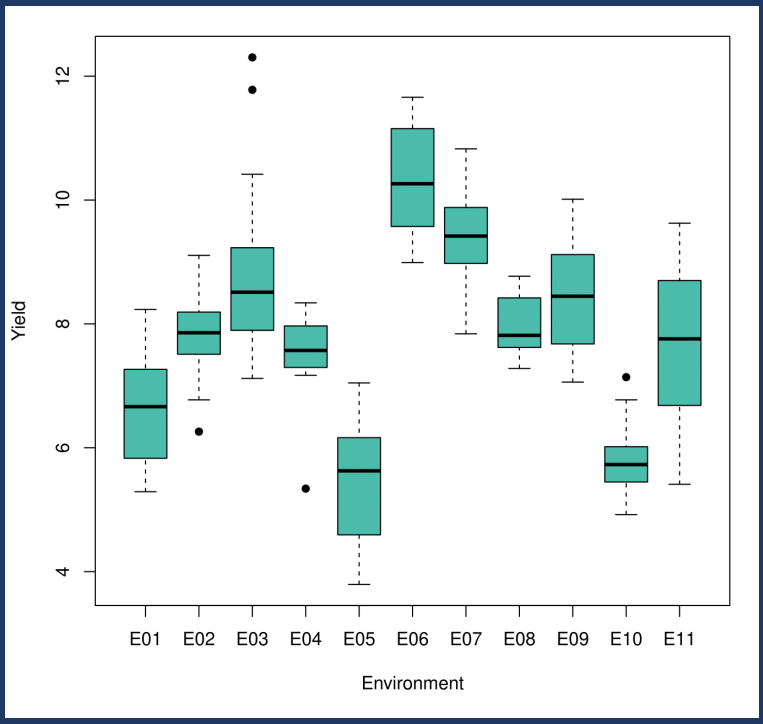
Stability Analysis in Plant Breeding Using PBSTAT-GE



• Willy Bayuardi Suwarno • Hajrial Aswidinnoor
• Sobir • Muhamad Syukur
• Arya Widura Ritonga • Trias Sitaresmi

Stability Analysis in Plant Breeding Using PBSTAT-GE

Willy Bayuardi Suwarno Hajrial Aswidinnoor
Sobir Muhamad Syukur
Arya Widura Ritonga Trias Sitaresmi

Table of contents

About the book

Stability Analysis in Plant Breeding Using PBSTAT-GE

Willy Bayuardi Suwarno*
Hajrial Aswidinnoor
Sobir
Muhamad Syukur
Arya Widura Ritonga
Department of Agronomy and Horticulture
Faculty of Agriculture, IPB University
Jl. Meranti Kampus IPB Dramaga
Bogor, Jawa Barat 16680, Indonesia
*bayuardi@gmail.com

Trias Sitaresmi
Research Center for Food Crops
National Research and Innovation Agency (BRIN)
Kawasan Sains dan Teknologi (KST) Soekarno
Jl. Raya Jakarta - Bogor Km. 46
Cibinong, Jawa Barat 16911, Indonesia

ISBN 979-8-89705-285-1
ISBN acquired, 31 December 2024
First revision, 10 September 2025

Publisher of record:
Primedia E-launch LLC
Dallas, Texas, USA

Preface

Alhamdulillah, praise and gratitude belong only to Allah SWT. May peace and blessings always be upon the Prophet Muhammad SAW and his family and followers until the end of time.

This book is written for plant breeders in particular, and researchers in the field of life sciences in general, who are interested in conducting genotype by environment interaction (GxE) analysis and stability analysis. The topic of stability analysis has been up for decades, and scientists have been exploring new methods until recently to understand more about the GxE interaction.

We would like to thank the various parties who have provided assistance and support in the development of PBSTAT-GE software and the writing of this book. We hope that this book can be useful for elucidating GxE in breeding programs.

1 Introduction

Genotype x environment (GxE) interaction is an important issue in plant breeding. The presence of GxE interaction can cause changes in genotype rankings between environments and make the estimated values of genetic parameters biased upward.

If the GxE interaction is significant, then a number of scientific questions may arise, including: Which genotypes have broad adaptation? Which genotypes are environment-specific? Stability analysis may be able to help answering these questions. This analysis can be regarded as a follow-up analysis when the combined analysis of variance (ANOVA) reveals the significance of the GxE interaction effect.

A number of software have been developed for GxE interaction and stability analyses for plant breeding, including GEA-R from The International Maize and Wheat Improvement Center (CIMMYT) and PBTools from the International Rice Research Institute (IRRI). In addition, several R packages can also be used for stability analysis, including 'metan' and 'agricolae'. Some of these software require installation, and not all of them are compatible with operating systems other than Microsoft Windows.

PBSTAT-GE is a software developed for GxE interaction and stability analysis. Unlike the software mentioned earlier, PBSTAT-GE is web-based so it does not require installation and only requires a web browser to access it. PBSTAT-GE can be accessed at www.pbstat.com.

The sample data set for analyses presented in this book was obtained from a multi-environment trial of new plant-type rice (Aswidinnoor et al., 2023). This data set consists of 14 rice genotypes evaluated in 11 environments in Indonesia, using a randomized complete block

design with three replications in each environment. Grain yield per plot was measured and then converted to ton per hectare at 15% moisture content. Data analyses were carried out using PBSTAT-GE version 3.5.1.

2 Data Preparation

PBSTAT-GE reads data files in .xlsx format (Excel workbook), so it is recommended that data be entered using Microsoft Excel. Table 2.1 presents a summary of the supported experimental design and data preparation.

Table 2.1: Summary of the supported experimental design and data preparation

Parameter	Value
Experimental design	Randomized complete block design in each environment
Minimum number of genotypes	5
Minimum number of environments	3
Minimum number of replications	2
Columns title in Excel	env, rep, geno (all lowercase), followed by trait names, e.g. yield, NPT, NFG
Number of rows in Excel	(no. of envs x no. of genos x no. of reps) + 1 title row
Missing data	Empty cell (not . or NA)
Maximum proportion of missing data	10%

Data entry needs to be done carefully so that there are no mistakes. After data entry is complete, recheck and make sure that the data entered is correct. Checking the results of data entry can be done

as follows. Click the Data > Filter menu in Excel, then click the arrow in each column title, one by one. Under the arrow will appear the unique values of the relevant column, making it easier for us to see whether there is any strange data or not. In the categorical variable columns such as env (environment), rep (replication), and geno (genotype), see if there is a typo, for example, replication 1 is accidentally written as 11. In the numeric variable column, see the range of the data (from the smallest to the largest). The presence of outlier data (too small or too large), or large gaps, or numbers that are read as letters, could be caused by incorrect data entry. Correct all errors if there are any so that all data is correct.

PBSTAT-GE is designed to analyze data from multi-environment experiments, involving the evaluation of at least five genotypes across a minimum of three environments, with at least two replications per environment. The set of genotypes tested in each environment must be the same. For example, if in environment 1 genotypes G01, G02, G03, … G14 are evaluated, then in environment 2 and so on G01, G02, G03, … G14 are also evaluated. The experiment in each environment uses a randomized complete block design. Other experimental designs are not currently supported. For a valid F test, the degrees of freedom (df) error in the ANOVA by environment as well as the df replication in environment and the df error in the combined ANOVA must be at least 6 (Gomez & Gomez, 1984).

PBSTAT-GE reads the column headings. The first three columns must be titled env, rep, geno (written in all lowercase letters), which contain information on the environment, replication, and genotype, respectively. The contents of the env and geno columns are recommended to be short and without spaces, for example in the form of codes E01, E02, E03, etc. for the environment and G01, G02, G03, etc. for the genotype. The contents of the rep column must be numbers (1, 2, 3, etc.). The following columns contain the observed variables, for example, yield, number of productive tillers, number of filled grains per panicle. The column titles of these variables must be made without spaces and begin with a letter (not a number), for example, Yield, NPT, and NFG.

The number of rows in the data worksheet must be the same as the number of experimental units in the entire experiment plus one row of column titles. For example, if the multilocation trial involves 14 lines, 11 environments, and 3 replicates per environment, then the number of experimental units is 4 x 11 x 3 = 462. The number of rows in the Excel worksheet is 462 rows of data + 1 row of column headers = 463. An inappropriate number of rows can cause the program to fail.

If there is missing data, then the relevant cell is blank (not filled with a dot, NA, or others). However, the env, rep, and geno information for the blank cell must remain on the worksheet. The proportion of missing data that can be accommodated by PBSTAT-GE is a maximum of 10%. Missing data will be estimated by the combined analysis linear model, and the analysis will be performed on the complete data set.

3 GxE means and correlation

3.1 GxE means

The GxE means table (Figure 3.1) shows the average genotype response by environment and the average genotype from all environments. At the bottom of the table is the LSD value at 0.05 level to compare the average of the tested genotype with that of the check variety. If the difference between the two averages is greater than the LSD value at 0.05 level, it is said to be significantly different. The CV value (%) shows the magnitude of the experimental error in percentage of the average. Rep p-value and G p-value in each environment are the p-values for the effects of replication and genotype from the F test by environment.

Rep p-value, G p-value, and GxE p-value in the Mean column are the p-values for the effects of replication in environment, genotype, and genotype x environment interaction from the combined analysis between environments. The Bartlett p-value shows the results of the homogeneity test of error variance using the Bartlett method, where the null hypothesis H0: error variances are homogenous and the alternative hypothesis H1: error variances are not homogenous. Here Bartlett p-value < 0.05 which indicates the error variance is not homogeneous. However, because the CV value is $< 20\%$ in all locations, the combined analysis can still be performed (Gomez and Gomez, 1984).

GxE means

Genotype	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10	E11	Mean
G01	5.83	9.11	11.78	8.34	4.34	10.24	9.64	7.93	7.34	5.45	8.07	8.01
G02	6.42	6.77	9.23	8.01	4.48	10.29	9.51	8.42	8.57	5.78	8.31	7.80
G03	6.82	8.29	8.12	7.71	6.43	11.55	9.11	7.68	8.32	5.91	7.75	7.97
G04	6.50	7.93	8.27	5.34	5.84	9.72	8.59	7.28	7.06	5.46	9.63	7.42
G05	8.23	8.05	8.64	8.04	6.53	11.34	9.88	8.77	9.24	6.39	8.70	8.53
G06	7.92	8.29	12.30	7.97	6.01	11.15	10.61	8.29	10.01	6.77	9.02	8.94
G07	5.63	6.26	10.42	7.50	6.16	9.57	8.98	7.68	8.02	5.68	5.80	7.43
G08	8.23	8.14	7.90	7.54	5.92	11.66	10.83	8.43	9.38	7.14	7.77	8.45
G09	7.27	8.19	7.77	7.22	7.05	9.32	10.16	8.63	9.12	6.02	8.75	8.14
G10	5.29	7.04	8.40	7.30	5.35	9.14	9.63	8.04	8.18	5.96	6.49	7.35
G11	5.90	7.51	7.82	7.17	3.79	9.81	9.33	7.62	7.46	5.19	6.83	7.13
G12	5.42	7.55	8.62	7.60	5.38	10.30	7.84	7.45	7.68	4.95	5.41	7.11
G13	6.88	7.79	8.90	7.66	4.59	8.99	8.51	7.70	8.70	4.92	6.68	7.39
G14	6.90	7.73	7.12	7.43	5.41	10.57	9.06	7.46	8.61	5.64	7.39	7.58
Mean	6.66	7.76	8.95	7.49	5.52	10.26	9.41	7.96	8.41	5.80	7.61	7.80
LSD 0.05	0.71	1.06	2.02	1.14	0.28	1.09	0.58	0.88	1.10	0.91	0.63	0.30
CV (%)	7.69	9.85	16.17	10.96	3.66	7.63	4.46	7.97	9.43	11.29	5.98	9.55
Rep p-value	0.16	0.01	0.01	0.22	0.24	0.15	0.00	0.00	0.90	0.00	0.20	0.00
G p-value	0.00	0.02	0.00	0.04	0.00	0.00	0.00	0.12	0.00	0.01	0.00	0.00
E p-value	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.00
GxE p-value	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.00
Bartlett p-value	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.00

Figure 3.1: GxE means

3.2 GxE boxplot

Below the GxE means table is a boxplot between the response variable (Y) vs. environment (Figure 3.2). This boxplot shows the distribution of plot values in each environment. The diversity in each environment is caused by the diversity between genotypes and the diversity between experimental plots. The lower edge of the box is the first quartile (Q1), the line in the middle of the box is the second quartile or the median (Q2), and the upper edge of the box is the third quartile (Q3). The dotted line connects the box with the minimum value line within the limits of Q1 - 1.5 * IQR and the maximum value within the limits of Q1 + 1.5 * IQR. Values outside these limits are considered outliers and are written in the form of black dots. From Figure 3.2 it is known that environments E02, E04, and E10 have relatively small variations within the environment, while E11 has relatively large variations.

3.3 GxE heatmap

GxE heatmap (Figure 3.3) visualizes the average genotype response in each environment. The average response is standardized by column (by environment) so that it has a mean of 0 and a standard deviation of 1. Therefore, the heatmap needs to be read by column (by environment). In one column, the darkest color indicates the highest response value, while the lightest color indicates the lowest response value.

The heatmap is complemented by two dendrograms based on Euclidean distance, namely the genotype dendrogram and the environment dendrogram. The genotype dendrogram (row) groups genotypes based on the similarity of their response patterns between environments, while the environment dendrogram (column) groups environments based on the similarity of their response patterns between genotypes. In the genotype dendrogram, it can be seen that G01 and G06 are grouped because they tend to have high results in most environments, and conversely G10 and G11 are grouped because they

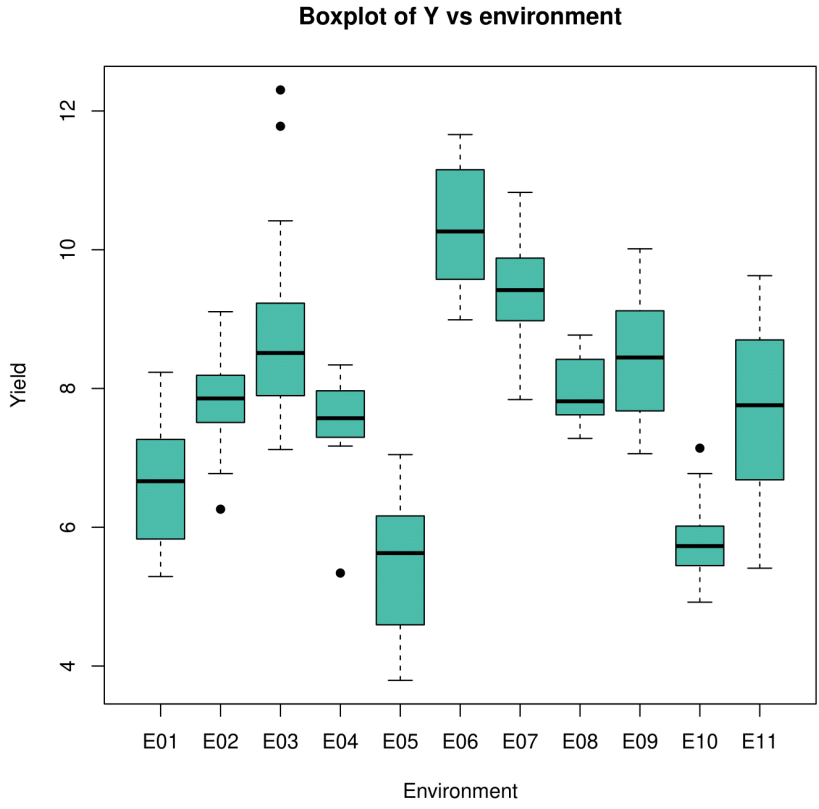


Figure 3.2: GxE boxplot

tend to have low results in most environments. In the environmental dendrogram, E05 and E10; E04 and E02; and E08 and E09 are located close together, indicating that each group has similar response patterns between genotypes.

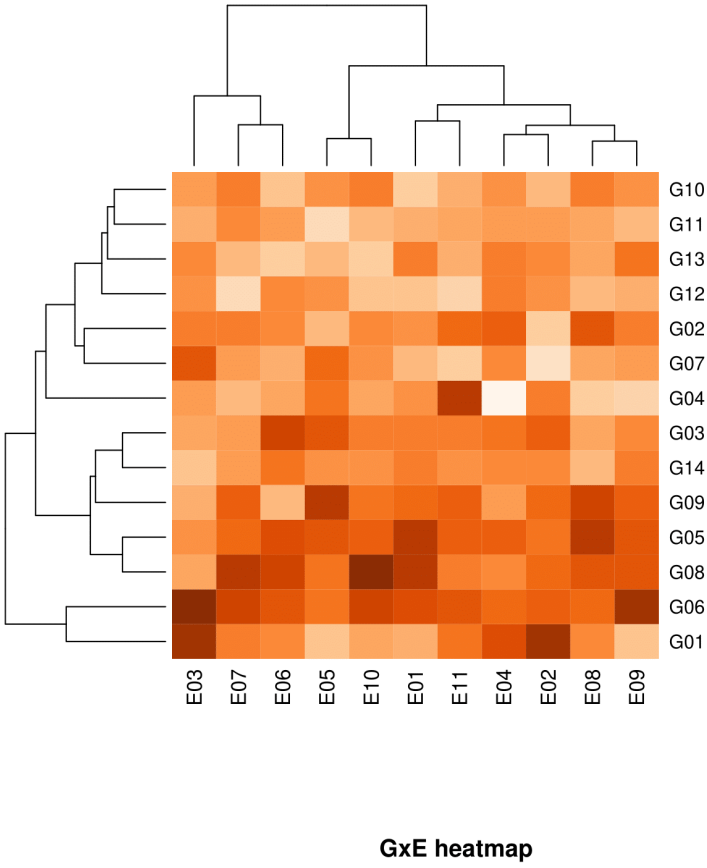


Figure 3.3: GxE heatmap

3.4 Spearman correlations among environments

Spearman correlations among environments (Figure 3.4) show the correlation of genotype ranks between pairs of environments. This

correlation also indicates the magnitude of the GxE interaction between two particular environments. If the correlation is large and positive, then the pattern of genotype ranks in one environment is similar to the pattern of genotype ranks in the other environment. In this case, the genotype that is a winner in one environment is likely to be a winner in the other environment, and vice versa, the genotype that is a loser in one environment is likely to be a loser in the other environment. In this case, the GxE interaction is not very large and the type of interaction may not be highly crossover. Conversely, a negative correlation indicates a large GxE interaction and is of the crossover type, where the patterns of genotype responses are opposite between the two environments. The genotype that is a winner in one environment is likely to be a loser in the other environment, and the genotype that is a loser in one environment is likely to be a winner in the other environment.

Spearman correlations among environments

	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10
E01										
E02	0.55*									
E03	-0.21	0.05								
E04	0.18	0.36	0.67**							
E05	0.54*	0.33	-0.15	-0.02						
E06	0.50	0.41	-0.08	0.44	0.30					
E07	0.53*	0.49	0.02	0.25	0.28	0.37				
E08	0.51	0.26	0.20	0.42	0.34	0.14	0.82**			
E09	0.80**	0.30	-0.01	0.30	0.50	0.38	0.59*	0.71**		
E10	0.57*	0.32	-0.04	0.14	0.64*	0.47	0.82**	0.72**	0.69**	
E11	0.61*	0.59*	0.04	0.13	0.33	0.25	0.53*	0.38	0.28	0.47

Figure 3.4: GxE correlation among environments

The results of the Spearman correlation analysis between environments can be in line with the environmental dendrogram in the GxE heatmap, or they can be different. Two adjacent environmental

groups in the dendrogram have significant positive Spearman correlation coefficient values: E05 and E10 ($r=0.64$, $p<0.05$); and E08 and E09 ($r=0.71$, $p<0.01$). In contrast, E04 and E02 which appear close together in the dendrogram have an insignificant rank correlation ($r=0.36$, $p>0.05$), and environments E07 and E08 which are far apart in the dendrogram have the largest Spearman correlation ($r=0.82$, $p<0.01$). This could be because the dendrogram is based on Euclidean distances calculated from the standardized actual response values, while Spearman correlation is based on ranking.

4 Analysis of variance

4.1 ANOVA by environment

Analysis of variance (ANOVA) by environment was conducted to study the effects of genotype and replication (block) on an observed response variable in each environment. As an example, the ANOVA results in environments 1 and 2 for yield trait are presented (Figure 4.1). In environments 1 and 2, the effects of genotype were highly significant ($p<0.01$) and significant ($p<0.05$) on yield, respectively, meaning that there was at least one pair of genotypes that had different yield in each environment. The effect of replication in environment 1 was not significant on yield ($p>0.05$) indicating that the variability between blocks was not large. In contrast, in environment 2, the effect of replication was highly significant ($p<0.01$) indicating that blocking was quite effective in controlling experimental errors.

4.2 Combined ANOVA across environments

A combined analysis of variance (ANOVA) between environments was conducted to study the effects of environment, replication within environment, genotype, and genotype x environment (GxE) interactions on the observed response variables (Figure 4.2). Genotype was assumed as a fixed factor while environment and replication within environment were assumed as random factors. Consequently:

- Genotype was tested against GxE interaction
- Replication in environment and GxE interaction were tested against experimental error

ANOVA by environment

```
$E01
Analysis of Variance Table

Response: yvar
            Df Sum Sq Mean Sq F value    Pr(>F)
Replication  2  1.016 0.50799  1.9376    0.1643
Genotype    13 38.580 2.96771 11.3194 1.338e-07 ***
Residuals   26  6.817 0.26218
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

$E02
Analysis of Variance Table

Response: yvar
            Df  Sum Sq Mean Sq F value   Pr(>F)
Replication  2  7.2829  3.6414  6.2268 0.006173 **
Genotype    13 20.0153  1.5396  2.6328 0.017241 *
Residuals   26 15.2048  0.5848
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4.1: ANOVA by environment

ANOVA results showed that all sources of variability had a highly significant effect on yield ($p<0.01$) (Figure 4.2). The highly significant GxE interaction indicated that the genotype effects changed between environments. To understand more about this GxE interaction, stability analyses can be performed.

Combined ANOVA across environments

```
Analysis of Variance Table

Response: yvar
                 Df Sum Sq Mean Sq F value    Pr(>F)
Environment (E)  10 880.43  88.043 37.1152 1.863e-11 ***
Replication/E    22  52.19   2.372  4.2695 3.554e-09 ***
Genotype (G)     13 134.23  10.325  5.5882 5.168e-08 ***
GxE             130 240.20   1.848  3.3255 < 2.2e-16 ***
Residuals       286 158.91   0.556
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4.2: Combined ANOVA across environments

4.3 Checking ANOVA assumptions

There are at least three assumptions of ANOVA, namely (1) errors are independent, (2) error variances are homogeneous, and (3) errors are normally distributed. The first assumption (errors are independent), can be considered fulfilled if the randomization and treatment placement were carried out correctly. The second and third assumptions can be checked using diagnostic plots (Figure 4.3).

The assumption of homogeneity of error variance can be checked with the residuals vs fitted values plot. If the distribution of points along the X-axis is relatively uniform (more or less the same width), and does not form a pattern (usually in the shape of a funnel; the greater the fitted value, the larger the variance), then the assumption of homogeneity of error variance can be considered fulfilled.

The assumption of normality of error is checked with a normal Q-Q plot. If the points are mostly on the line and there are no big deviations (often the deviations are at the tail and head), then the assumption of normality of error variance can be considered fulfilled.

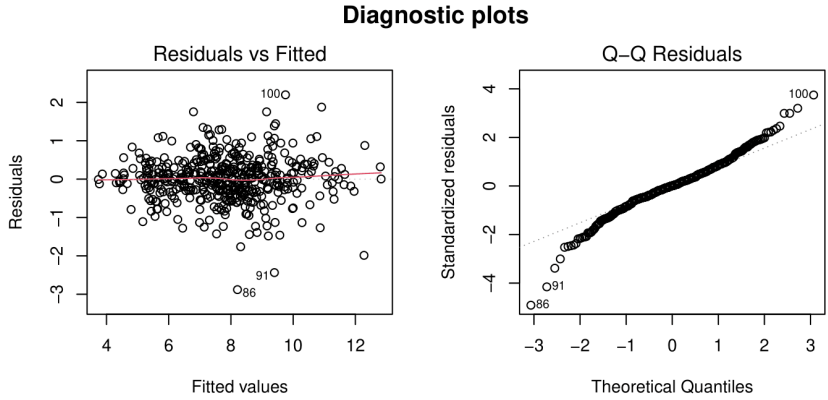


Figure 4.3: Diagnostic plots

The residual vs fitted values plot (Figure 4.3 left) shows that the widths of the points along the X-axis are not significantly different. The normal Q-Q plot (Figure 4.3 right) shows that most of the points fall on the line, although there are deviations at the tail and head. Both of these plots indicate that the assumptions of homogeneity of error variances and normality of error are met for the combined analysis of variance across environments.

5 Variance components and heritability

5.1 Estimates of variances

This analysis assumes a random effects model, i.e., all variables in the model are assumed random. Estimates of variance used the least square methods. Vg is the genotypic variance, Vge is the genotype x environment interaction variance, Ve is the error variance, Vp plot is the phenotypic variance on a plot basis, and Vp mean is the phenotypic variance on an entry-mean basis. MSE is the mean square error, MSGE is the mean square GxE, MSG is the mean square genotype, r is the number of replications, and t is the number of environments. A negative variance value may be interpreted as zero.

$$V_e = MSE$$

$$V_{ge} = \frac{MSGE - MSE}{r}$$

$$V_g = \frac{MSGE - MSE}{rt}$$

$$V_{p(plot)} = V_e + V_{ge} + V_g$$

$$V_{p(mean)} = \frac{V_e}{rt} + \frac{V_{ge}}{t} + V_g$$

5.2 Estimates of heritability

The heritability estimate was calculated according to Fehr (1991). Heritability indicates the proportion of phenotypic variance due to genetic variance. $h^2_{(plot)}$ is the broad-sense heritability on a plot basis and $h^2_{(mean)}$ is broad-sense heritability on an entry-mean basis. The usage of plot or mean basis heritability depends on the basis of selection. If the selection was based on plot values (i.e., selecting best plots), then the plot-basis heritability is relevant. On the other hand, if the selection was based on entry means (i.e., selecting best entries), then the mean-basis heritability is pertinent. A negative heritability value may be interpreted as zero.

$$h^2_{(plot)} = \frac{V_g}{V_{p(plot)}} \times 100\%$$

$$h^2_{(mean)} = \frac{V_g}{V_{p(mean)}} \times 100\%$$

6 Stability parameters

About forty parameters of stability are available in the current version of PBSTAT-GE which are divided into three parts. Here, parameters refer to measures, as the discussion encompasses both parametric and nonparametric approaches. For information on stability parameters and their calculations, see a review by Pour-Aboughadareh et al. (2022) and the manual of R packages 'metan' (Olivoto and Lucio, 2020).

Part 1 of stability parameters (Figure 6.1) comprises Y: Mean response; EVar: Environmental variance; W2: Ecovalence; b: Regression coefficient; b_sig: Significance for b (H0: b=1, 95% CI: 0.9 - 1.1); b_p: p-value for bi (H0: b=1); s2d: Deviation from regression; s2d_sig: Significance for s2d (H0: s2di=0); s2d_p: p-value for s2d (H0: s2di=0); D2: Genotypic stability; 2: Stability variance; R2: Coefficient of determination; CV: Coefficient of variation; GAI: Geometric adaptability index; POLAR: Power law residuals; aCV: Adjusted coefficient of variation; Wi_g, Wi_f, Wi_u: Genotypic confidence index for all, favorable, and unfavorable environments, respectively; Pi_a, Pi_f, Pi_u: Superiority indexes for all, favorable, and unfavorable environments, respectively.

Part 2 of stability parameters (Figure 6.2) contains ASTAB: AMMI based stability parameter; ASI: AMMI stability index; ASV: AMMI-stability value; AVAMGE: Sum across environments of absolute value of GEI modeled by AMMI; Da: Annicchiarico's D parameter; Dz: Zhang's D Parameter; EV: Sums of the averages of the squared eigenvector values; FA: Stability measure based on fitted AMMI model; MASI: Modified AMMI stability index; MASV: Modified AMMI stability value; SIPC: Sums of the absolute value of the IPC

Stability parameters (1)

Genotype	Y	EVar	W2	b	b_sig	b_p	s2d	s2d_sig	s2d_p	D2	sigma2	R2	CV	GAI	POLAR	aCV	Pi_a	Pi_f	Pi_u	Wi_g	Wi_f	Wi_u
G01	8.01	4.84	13.71	1.33	***	0.00	1.09	***	0.00	20.21	1.55	0.76	27.49	7.71	0.27	28.17	1.36	1.15	1.54	76.67	77.18	74.56
G02	7.80	3.05	3.21	1.15	**	0.00	0.12	ns	0.10	7.36	0.32	0.91	22.40	7.60	0.08	22.44	1.56	1.53	1.59	85.50	98.89	78.72
G03	7.97	2.25	3.42	0.96	ns	0.39	0.19	*	0.04	4.96	0.35	0.85	18.83	7.85	-0.06	19.22	1.50	2.45	0.71	90.51	85.81	95.88
G04	7.42	2.41	11.03	0.81	***	0.00	0.96	***	0.00	10.65	1.24	0.57	20.94	7.27	-0.02	20.09	2.44	3.60	1.47	73.05	84.14	70.16
G05	8.53	1.95	2.45	0.91	ns	0.07	0.07	ns	0.21	3.32	0.23	0.88	16.36	8.42	-0.13	17.71	0.82	1.50	0.24	98.04	96.60	100.85
G06	8.94	3.63	7.18	1.19	***	0.00	0.53	***	0.00	11.90	0.79	0.82	21.31	8.76	0.13	24.02	0.14	0.05	0.21	98.90	95.12	102.92
G07	7.43	2.83	8.53	0.97	ns	0.57	0.76	***	0.00	10.27	0.94	0.70	22.64	7.26	0.05	21.73	2.26	1.65	2.76	75.50	83.67	69.49
G08	8.45	2.67	6.25	0.99	ns	0.81	0.51	***	0.00	8.21	0.68	0.77	19.34	8.31	0.01	20.76	1.19	1.99	0.53	91.84	88.82	93.16
G09	8.14	1.45	6.90	0.68	***	0.00	0.35	**	0.00	4.78	0.75	0.67	14.81	8.05	-0.25	15.39	1.47	2.73	0.42	86.92	82.76	91.82
G10	7.35	2.17	3.15	0.94	ns	0.27	0.16	ns	0.06	4.51	0.32	0.86	20.06	7.21	-0.07	19.08	2.50	2.69	2.35	81.97	87.78	77.85
G11	7.13	2.95	2.24	1.15	**	0.00	0.01	ns	0.39	6.36	0.21	0.94	24.08	6.92	0.07	22.32	2.96	3.36	2.63	76.86	85.04	71.48
G12	7.11	2.73	5.27	1.02	ns	0.64	0.40	**	0.00	7.72	0.56	0.81	23.22	6.94	0.04	21.47	3.08	3.15	3.02	75.23	82.54	69.64
G13	7.39	2.28	3.30	0.97	ns	0.51	0.18	*	0.04	4.97	0.33	0.86	20.44	7.23	-0.04	19.54	2.30	2.70	1.97	81.63	84.89	78.30
G14	7.58	2.14	3.42	0.93	ns	0.17	0.18	*	0.04	4.60	0.35	0.85	19.32	7.45	-0.07	18.87	2.24	3.49	1.20	86.41	79.41	95.16

Figure 6.1: Stability parameters (1)

scores; Za: Absolute value of the relative contribution of IPCs to the interaction; WAAS: Weighted average of absolute scores (WAAS).

Stability parameters (2)

Genotype	ASTAB	ASI	ASV	AVAMGE	Da	Dz	EV	FA	MASI	MASV	SIPC	Za	WAAS
G01	2.82	0.56	2.63	9.36	3.69	0.80	0.11	13.60	0.56	2.98	3.36	0.37	0.84
G02	0.64	0.09	0.42	3.23	1.38	0.47	0.04	1.89	0.12	1.12	1.53	0.11	0.23
G03	0.96	0.18	0.87	4.78	1.81	0.56	0.05	3.29	0.20	1.48	2.02	0.17	0.36
G04	2.67	0.37	1.74	7.84	3.31	0.82	0.11	10.93	0.37	2.91	2.95	0.29	0.64
G05	0.45	0.23	1.07	3.74	1.45	0.34	0.02	2.11	0.23	1.18	1.13	0.12	0.28
G06	1.85	0.33	1.55	6.15	2.63	0.76	0.10	6.91	0.34	2.18	3.09	0.28	0.60
G07	2.11	0.35	1.66	7.27	2.90	0.76	0.10	8.43	0.37	2.28	3.09	0.30	0.65
G08	1.37	0.33	1.56	6.06	2.38	0.63	0.07	5.64	0.34	1.94	2.69	0.25	0.55
G09	1.81	0.33	1.55	6.83	2.61	0.74	0.09	6.81	0.34	2.08	2.86	0.26	0.56
G10	1.10	0.10	0.48	5.15	1.74	0.65	0.07	3.02	0.13	1.42	2.36	0.15	0.29
G11	0.77	0.02	0.08	3.88	1.47	0.53	0.05	2.18	0.09	1.08	1.73	0.10	0.18
G12	1.46	0.21	1.01	6.12	2.27	0.66	0.07	5.17	0.23	1.95	2.33	0.21	0.43
G13	1.39	0.11	0.50	5.57	1.80	0.80	0.11	3.24	0.12	1.34	2.17	0.14	0.27
G14	0.68	0.28	1.31	3.63	1.81	0.39	0.03	3.29	0.28	1.45	1.61	0.18	0.41

Figure 6.2: Stability parameters (2)

Part 3 of stability parameters (Figure 6.3) include YS: Yield and stability index; YS_sel: '+' selected genotypes having YS > mean of 2.29; TOP: Number of sites at which the genotype occurred in the top third of the ranks; S1, S2, S3, S6: Huhn nonparametric stability measures. S1: Mean of the absolute rank differences of a

genotype over environments, S2: Variance among the ranks over the environments, S3: Sum of the absolute deviations, S6: Relative sum of squares of rank for each genotype; Z1, Z2: Test statistics for S1 and S2, respectively; N1, N2, N3, N4: Thennarasu nonparametric stability measures. Chi-square tests of Huhn stability measures are shown in Figure 6.4.

Stability parameters (3)

Genotype	YS	YS_sel	TOP	S1	Z1	S2	Z2	S3	S6	N1	N2	N3	N4	BAI1
G01	3.00	+	3	6.25	4.73	28.65	6.49	26.16	5.20	4.45	0.64	0.72	0.88	0.44
G02	5.00	+	1	4.15	0.45	12.49	0.60	13.01	3.22	2.64	0.38	0.48	0.58	0.77
G03	6.00	+	3	4.22	0.33	13.29	0.37	9.68	3.11	2.64	0.38	0.54	0.65	0.85
G04	-5.00		1	5.27	0.72	19.42	0.42	28.81	6.03	3.55	0.35	0.44	0.55	0.66
G05	16.00	+	7	3.49	2.42	9.16	2.12	2.35	1.20	2.36	0.79	0.89	1.08	0.67
G06	9.00	+	7	5.13	0.43	18.87	0.29	1.78	1.13	3.64	1.82	1.47	1.82	0.52
G07	-4.00		1	5.13	0.43	18.29	0.18	21.41	5.15	3.55	0.35	0.43	0.54	0.70
G08	7.00	+	6	5.53	1.42	22.65	1.73	10.83	2.95	4.00	1.33	1.07	1.31	0.68
G09	5.00	+	4	5.31	0.81	20.22	0.66	20.78	4.29	3.73	0.93	0.76	0.94	0.74
G10	-1.00		0	4.98	0.21	17.80	0.10	17.63	5.29	3.27	0.33	0.42	0.52	0.87
G11	-1.00		0	3.82	1.24	10.65	1.32	7.91	3.86	2.64	0.24	0.28	0.34	0.76
G12	-10.00		0	5.53	1.42	22.25	1.52	20.75	6.50	4.00	0.36	0.42	0.52	0.80
G13	0.00		0	5.42	1.09	21.27	1.06	21.59	5.87	4.00	0.50	0.48	0.59	0.84
G14	2.00		0	3.71	1.59	10.09	1.60	12.56	3.65	2.45	0.27	0.34	0.42	0.83

Figure 6.3: Stability parameters (3)

Test of Huhn stability measures

Grand mean	Sum(Z1)	Sum(Z2)	E(S1)	E(S2)	Var(S1)	Var(S2)	Chi-sq table for Z1, Z2	Chi-sq table for Sum(Z1), Sum(Z2)
7.80	17.29	18.46	4.64	16.25	0.55	23.71	8.49	23.68

Figure 6.4: Test of Huhn stability measures

7 Stability ranks

Ranks of each stability parameter are shown in Figure 7.1, Figure 7.2, and Figure 7.3. Ranking 1 means the most stable and the highest ranking means the least stable for the parameter in question. Part 1 comprises rY: Rank of mean response; rEVar: Rank of environmental variance; rW2: Rank of ecovalence; rb: Rank of regression coefficient; rs2d: Rank of deviation from regression; rD2: Rank of genotypic stability; r 2: Rank of stability variance; rR2: Rank of coefficient of determination; rCV: Rank of coefficient of variation; rGAI: Rank of geometric adaptability index; rPOLAR: Rank of power law residuals; raCV: Rank of adjusted coefficient of variation.

Stability ranks (1)

Genotype	rY	rEVar	rW2	rb	rs2d	rD2	rsigma2	rR2	rCV	rGAI	rPOLAR	raCV	rPi_a	rPi_f	rPi_u	rWi_g	rWi_f	rWi_u
G01	5.00	14.00	14.00	14.00	14.00	14.00	14.00	11.00	14.00	6.00	14.00	14.00	4.00	2.00	8.00	11.00	14.00	10.00
G02	7.00	12.00	4.00	12.00	3.00	8.00	4.00	2.00	10.00	7.00	12.00	12.00	7.00	4.00	9.00	7.00	1.00	7.00
G03	6.00	5.00	6.00	6.00	7.00	5.00	6.00	6.00	3.00	5.00	5.00	5.00	6.00	7.00	5.00	4.00	6.00	3.00
G04	10.00	7.00	13.00	2.00	13.00	12.00	13.00	14.00	8.00	9.00	7.00	7.00	11.00	14.00	7.00	14.00	9.00	12.00
G05	2.00	2.00	2.00	3.00	2.00	1.00	2.00	3.00	2.00	2.00	2.00	2.00	2.00	3.00	2.00	2.00	2.00	2.00
G06	1.00	13.00	11.00	13.00	11.00	13.00	11.00	8.00	9.00	1.00	13.00	13.00	1.00	1.00	1.00	1.00	3.00	1.00
G07	9.00	10.00	12.00	8.00	12.00	11.00	12.00	12.00	11.00	10.00	10.00	10.00	9.00	5.00	13.00	12.00	10.00	14.00
G08	3.00	8.00	9.00	9.00	10.00	10.00	9.00	10.00	5.00	3.00	8.00	8.00	3.00	6.00	4.00	3.00	4.00	5.00
G09	4.00	1.00	10.00	1.00	8.00	4.00	10.00	13.00	1.00	4.00	1.00	1.00	5.00	10.00	3.00	5.00	11.00	6.00
G10	12.00	4.00	3.00	5.00	4.00	2.00	3.00	4.00	6.00	12.00	4.00	4.00	12.00	8.00	11.00	8.00	5.00	9.00
G11	13.00	11.00	1.00	11.00	1.00	7.00	1.00	1.00	13.00	14.00	11.00	11.00	13.00	12.00	12.00	10.00	7.00	11.00
G12	14.00	9.00	8.00	10.00	9.00	9.00	8.00	9.00	12.00	13.00	9.00	9.00	14.00	11.00	14.00	13.00	12.00	13.00
G13	11.00	6.00	5.00	7.00	5.00	6.00	5.00	5.00	7.00	11.00	6.00	6.00	10.00	9.00	10.00	9.00	8.00	8.00
G14	8.00	3.00	7.00	4.00	6.00	3.00	7.00	7.00	4.00	8.00	3.00	3.00	8.00	13.00	6.00	6.00	13.00	4.00

Figure 7.1: Stability ranks (1)

Part 2 contains rASTAB: Rank of AMMI based stability parameter; rASI: Rank of AMMI stability index; rASV: Rank of AMMI-stability value; rAVAMGE: Rank of sum across environments of absolute value of GEI modeled by AMMI; rDa: Rank of Annicchiarico's D parameter; rDz: Rank of Zhang's D Parameter; rEV: Rank of sums of

the averages of the squared eigenvector values; rFA: Rank of stability measure based on fitted AMMI model; rMASI: Rank of modified AMMI stability index; rMASV: Rank of modified AMMI stability value; rSIPC: Rank of sums of the absolute value of the IPC scores; rZa: Rank of absolute value of the relative contribution of IPCs to the interaction; rWAAS: Rank of weighted average of absolute scores (WAAS).

Stability ranks (2)

Genotype	rASTAB	rASI	rASV	rAVAMGE	rDa	rDz	rEV	rFA	rMASI	rMASV	rSIPC	rZa	rWAAS
G01	14.00	14.00	14.00	14.00	14.00	13.00	13.00	14.00	14.00	14.00	14.00	14.00	14.00
G02	2.00	2.00	2.00	1.00	1.00	3.00	3.00	1.00	2.00	2.00	2.00	2.00	2.00
G03	5.00	5.00	5.00	5.00	7.00	5.00	5.00	7.00	5.00	7.00	5.00	6.00	6.00
G04	13.00	13.00	13.00	13.00	13.00	14.00	14.00	13.00	13.00	13.00	11.00	12.00	12.00
G05	1.00	7.00	7.00	3.00	2.00	1.00	1.00	2.00	6.00	3.00	1.00	3.00	4.00
G06	11.00	9.00	9.00	10.00	11.00	10.00	10.00	11.00	10.00	11.00	13.00	11.00	11.00
G07	12.00	12.00	12.00	12.00	12.00	11.00	11.00	12.00	12.00	12.00	12.00	13.00	13.00
G08	7.00	11.00	11.00	8.00	9.00	6.00	6.00	9.00	9.00	8.00	9.00	9.00	9.00
G09	10.00	10.00	10.00	11.00	10.00	9.00	9.00	10.00	11.00	10.00	10.00	10.00	10.00
G10	6.00	3.00	3.00	6.00	4.00	7.00	7.00	4.00	4.00	5.00	8.00	5.00	5.00
G11	4.00	1.00	1.00	4.00	3.00	4.00	4.00	3.00	1.00	1.00	4.00	1.00	1.00
G12	9.00	6.00	6.00	9.00	8.00	8.00	8.00	8.00	7.00	9.00	7.00	8.00	8.00
G13	8.00	4.00	4.00	7.00	5.00	12.00	12.00	5.00	3.00	4.00	6.00	4.00	3.00
G14	3.00	8.00	8.00	2.00	6.00	2.00	2.00	6.00	8.00	6.00	3.00	7.00	7.00

Figure 7.2: Stability ranks (2)

Part 3 includes rYS: Rank of yield and stability index; rTOP: Rank of number of sites at which the genotype occurred in the top third of the ranks; rS1, rS2, rS3, rS4, rS6: Rank of Huhn nonparametric stability measures; rN1, rN2, rN3, rN4: Rank of Thennarasu nonparametric stability measures.

Stability ranks (3)

Genotype	rYS	rTOP	rS1	rS2	rS3	rS6	rN1	rN2	rN3	rN4	rBAI1
G01	7.00	5.50	14.00	14.00	13.00	10.00	14.00	10.00	10.00	10.00	14.00
G02	5.50	8.00	4.00	4.00	7.00	5.00	4.00	7.50	7.00	7.00	6.00
G03	4.00	5.50	5.00	5.00	4.00	4.00	4.00	7.50	9.00	9.00	2.00
G04	13.00	8.00	9.00	9.00	14.00	13.00	7.50	4.50	6.00	6.00	12.00
G05	1.00	1.50	1.00	1.00	2.00	2.00	1.00	11.00	12.00	12.00	11.00
G06	2.00	1.50	7.50	8.00	1.00	1.00	9.00	14.00	14.00	14.00	13.00
G07	12.00	8.00	7.50	7.00	11.00	9.00	7.50	4.50	5.00	5.00	9.00
G08	3.00	3.00	12.50	13.00	5.00	3.00	12.00	13.00	13.00	13.00	10.00
G09	5.50	4.00	10.00	10.00	10.00	8.00	10.00	12.00	11.00	11.00	8.00
G10	10.50	12.00	6.00	6.00	8.00	11.00	6.00	3.00	3.00	3.00	1.00
G11	10.50	12.00	3.00	3.00	3.00	7.00	4.00	1.00	1.00	1.00	7.00
G12	14.00	12.00	12.50	12.00	9.00	14.00	12.00	6.00	4.00	4.00	5.00
G13	9.00	12.00	11.00	11.00	12.00	12.00	12.00	9.00	8.00	8.00	3.00
G14	8.00	12.00	2.00	2.00	6.00	6.00	2.00	2.00	2.00	2.00	4.00

Figure 7.3: Stability ranks (3)

8 Ranks analyses

8.1 Spearman correlations among stability ranks

Spearman correlation analysis studies the relationship between the rankings of stability parameters (Figure 8.1). A positive and significant correlation between two stability parameters indicates that the rankings of the genotypes on both parameters are related in the same direction. This means that the winning genotype based on one parameter is also the winner based on the other parameter, and vice versa. This also indicates that if only the ranking of the genotypes matters, then only one of the two parameters can be chosen because it provides the same information. For example, Wricke's ecovalence has a perfect rank correlation (r=1.00, p<0.01) with Shukla stability variance, indicating that genotype rankings on both parameters are the same.

8.2 Principal component analysis

Principal component analysis was performed to study the relationship between genotypes and stability rankings (Figure 8.2). Red vectors represent stability parameters and black labels represent genotypes. The correlations of stability parameter rankings may be inferred from the vectors. Vectors with $< 90^{\circ}$ angles are positively correlated, 90° angles are uncorrelated, and $> 90^{\circ}$ angles are negatively correlated. For example, rTOP is positively correlated with rPi_a; rN2, rN3, and rN4 are positively correlated; however, the two groups are negatively correlated. Genotypes that are located in the same direction and pass the arrowhead have good rankings for the respective stability

Spearman correlations among stability ranks

	rY	rEVar	rW2	rb	rs2d	rD2	rsigma2	rR2	rCV	rGAI	rPOLAR
rY											
rEVar	0.04										
rW2	-0.27	0.32									
rb	-0.02	0.92**	0.10								
rs2d	-0.22	0.36	0.98**	0.16							
rD2	-0.09	0.83**	0.75**	0.63*	0.78**						
rsigma2	-0.27	0.32	1.00**	0.10	0.98**	0.75**					
rR2	-0.18	0.00	0.92**	-0.22	0.89**	0.52	0.92**				
rCV	0.44	0.88**	0.22	0.78**	0.27	0.70**	0.22	-0.01			
rGAI	0.99**	0.10	-0.27	0.05	-0.23	-0.06	-0.27	-0.20	0.50		
rPOLAR	0.04	1.00**	0.32	0.92**	0.36	0.83**	0.32	0.00	0.88**	0.10	
raCV	0.04	1.00**	0.32	0.92**	0.36	0.83**	0.32	0.00	0.88**	0.10	1.00**

Figure 8.1: Spearman correlations among stability ranks

parameters. For example, G06 is good based on TOP and G11 is good based on rN2, rN3, and rN4 parameters.

8.3 Heatmap visualization

Heatmap visualization for genotype and stability ranking also illustrates the relationship between the two (Figure 8.3). The data used for this heatmap visualization is standardized by column (stability parameter), so it needs to be read per column. The oldest color indicates the best ranking for the stability parameter in question, and the youngest color vice versa. For example, for the leftmost stability parameter, Wi_f, G02, and G05 have good rankings. For the rightmost stability parameter, sigma2, G11, and G05 have good rankings. The heatmap is complemented by dendrograms based on rows (genotypes) and columns (stability parameters). The row dendrogram indicates that G07 and G04 have similar stability ranking profiles, as do G10 and G13. The column dendrogram shows that rW2 (ranking of Wricke ecovalence) and rsigma2 (ranking of Shukla stability variance) are very close as discussed previously.

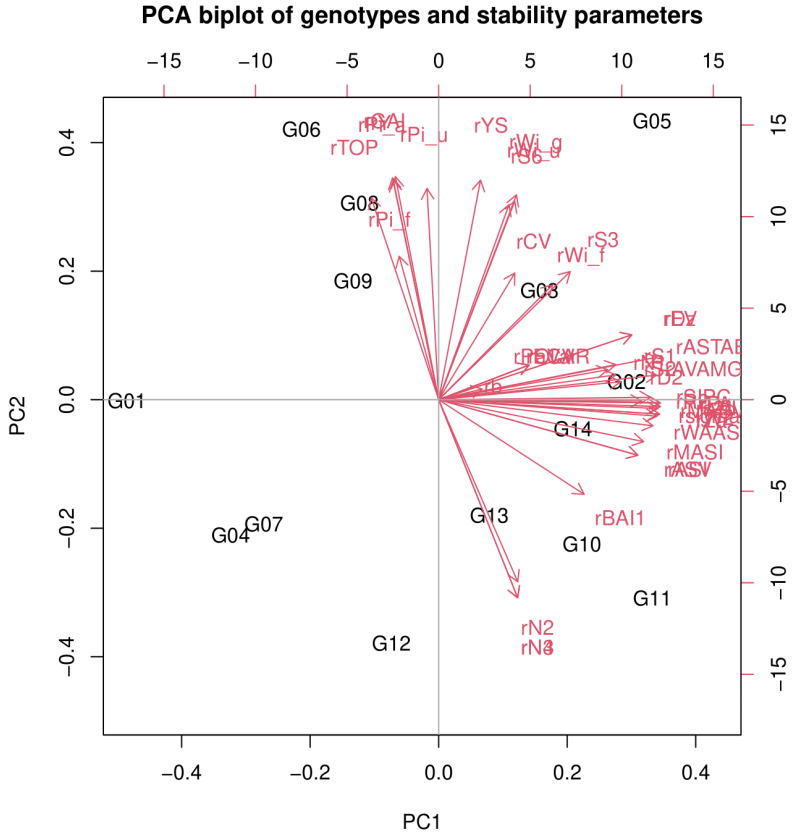


Figure 8.2: PCA of stability ranks

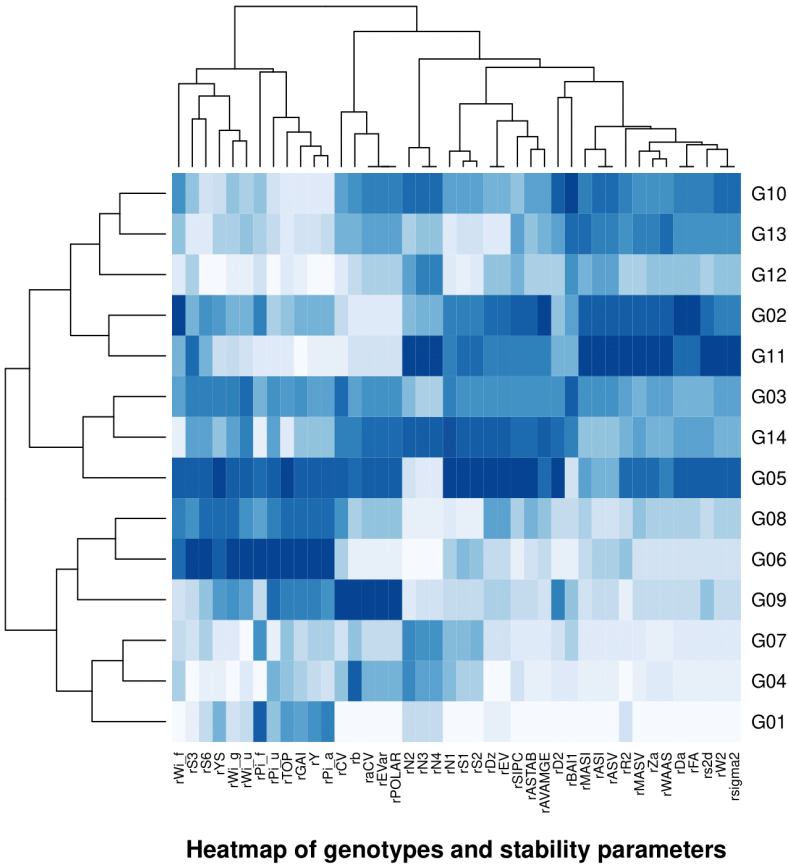


Figure 8.3: Heatmap of stability ranks

8.4 Cluster analysis

Cluster analysis for stability parameter ranking is performed to study the similarity of ranking between stability parameters (Figure 8.4). This cluster analysis is performed based on the dissimilarity coefficient calculated using the formula $1 - abs(r_{xy})$, where r_{xy} is the Spearman correlation coefficient between stability parameters x and y. The results of this dendrogram are similar to the stability parameter dendrogram in the heatmap, for example, rW2 (ranking of Wricke ecovalence) and rsigma2 (ranking of Shukla stability variance) are located very close to each other.

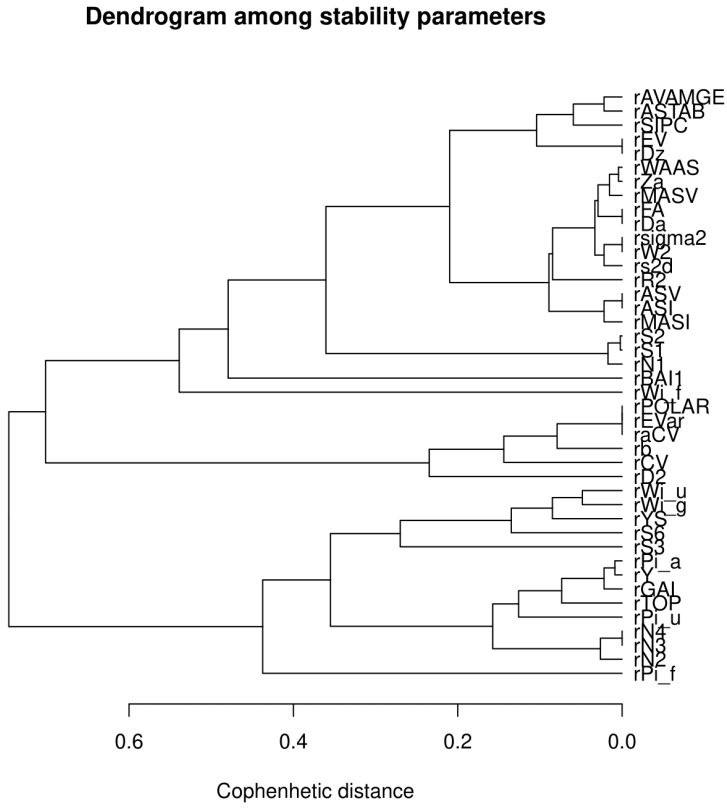


Figure 8.4: Dendrogram of stability ranks

9 Eberhart-Russell stability analysis

9.1 Analysis of variance

Eberhart and Russell (1966) proposed two stability parameters, namely the regression coefficient (b_i) and the regression deviation (s^2_{di}). A stable genotype has a regression coefficient value of $b_i = 1$ and a regression deviation of $s^2_{di} = 0$. The stability parameter $b_i = 1$ was previously described by Finlay and Wilkinson (1963).

The statistical test for the parameter s^2_{di} can be seen in the Eberhart-Russell ANOVA, in the 'Pooled deviations' partition section. There the regression deviation for each line is tested F with the null hypothesis $s^2_{di} = 0$ and the alternative hypothesis $s^2_{di} \neq 0$. The null hypothesis is rejected if the p-value < 0.05.

The Eberhart-Russell ANOVA is shown in Figure 9.1. The results of the F-tests on the regression deviation of genotypes G02, G05, G10, and G11 show insufficient evidence to reject H0, so H0 is accepted. This means that the regression deviations for these genotypes are not significantly different from 0. In other words, the four genotypes meet at least one Eberhart-Russell stability criterion ($s^2_{di} = 0$). The value of s^2_{di} and the results of the F test on $s^2_{di} = 0$ can also be seen in the stability parameters section.

Eberhart-Russel analysis of variance

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Total	153	418.29	2.734			
Genotype	13	44.74	3.442	5.9609	1.580e-08	***
Env + (Gen x Env)	140	373.54	2.668			
Env (linear)	1	293.48	293.478			
Gen x Env (linear)	13	7.32	0.563	0.9746	0.4799229	
Pooled deviations	126	72.75	0.577			
G01	9	11.45	1.273	6.8722	5.858e-09	***
G02	9	2.73	0.303	1.6369	0.1043778	
G03	9	3.38	0.376	2.0285	0.0362014	*
G04	9	10.29	1.144	6.1745	5.886e-08	***
G05	9	2.26	0.251	1.3570	0.2074799	
G06	9	6.39	0.710	3.8358	0.0001343	***
G07	9	8.51	0.946	5.1082	2.027e-06	***
G08	9	6.25	0.694	3.7482	0.0001786	***
G09	9	4.78	0.531	2.8653	0.0029831	**
G10	9	3.08	0.342	1.8463	0.0599651	.
G11	9	1.77	0.196	1.0599	0.3926766	
G12	9	5.26	0.585	3.1563	0.0011957	**
G13	9	3.28	0.364	1.9661	0.0431281	*
G14	9	3.32	0.369	1.9913	0.0401965	*
Pooled error	286	52.97	0.185			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure 9.1: Eberhart-Russell ANOVA

9.2 Plot of response and environmental index

The response and environmental index plot (Figure 9.2) visualizes the results of genotypes in each environment, where the environments are ranked by environmental index. The environmental index is calculated as the average response of all genotypes in that environment. The environment with the smallest index is the most 'infertile', located on the far left, while the environment with the largest index is the most 'fertile', located on the far right. From this plot, it can be seen that G09 has the highest average yield in the environment with the smallest index, while G06 is among those with the highest average yield in several environments with relatively small to relatively large indices.

9.3 Regression plot

The regression plot between the response value and environmental index is shown in the Figure 9.3. This plot is a linear regression plot for each genotype in the previous plot, with the equation $Y = a_i + b_i X$ where Y is the response value, a_i is the intercept of the i^{th} genotype, which is the value of Y at $X = 0$, and b is the regression coefficient of the i^{th} genotype. The average environmental regression line is visualized with the line $b = 1$ (see the legend at the bottom right), where the line is the line $Y = X$. The value of b_i and the result of its t-test against $b_i = 1$ are shown in the stability parameter section. Genotypes with a value of $b_i < 1$ are probably adaptive in marginal environments, genotypes with $b_i = 1$ have average stability, and genotypes with a value of $b_i > 1$ are probably adaptive in fertile environments. Genotypes with $b_i = 1$ and an average yield above the general mean show broad adaptation across all test environments, while genotypes with $b_i < 1$ and an average yield below the general mean are less adaptive across all environments. This plot shows that the position of the G6 genotype line (bi=1.19, p<0.01) is generally above the other genotypes. In this case, although the b_i value is

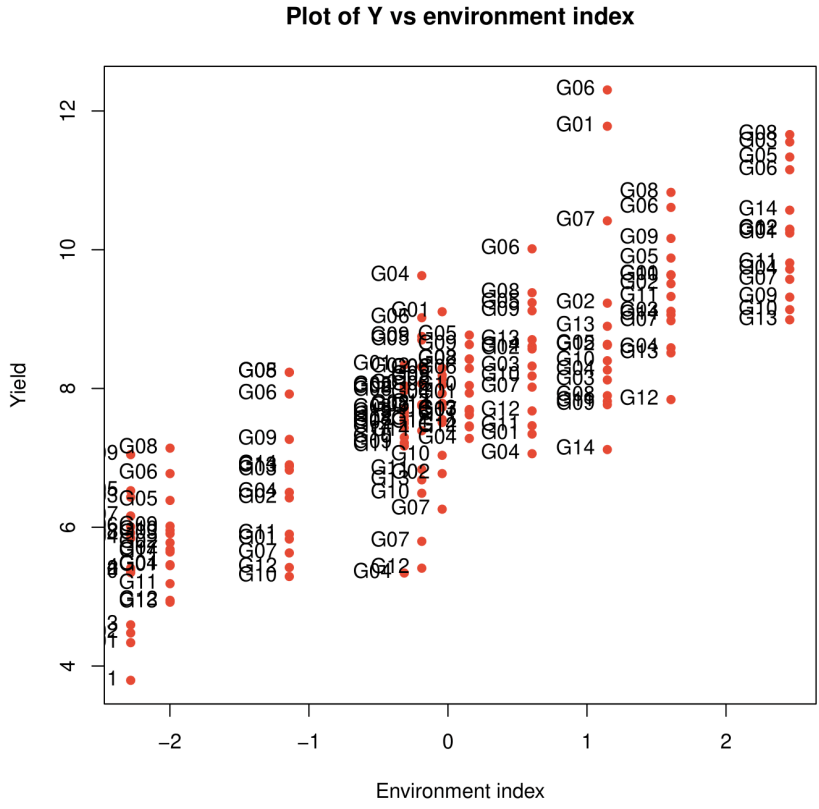


Figure 9.2: Plot of yield and environment index

significantly greater than 1, the genotype still appears to have broad adaptation potential.

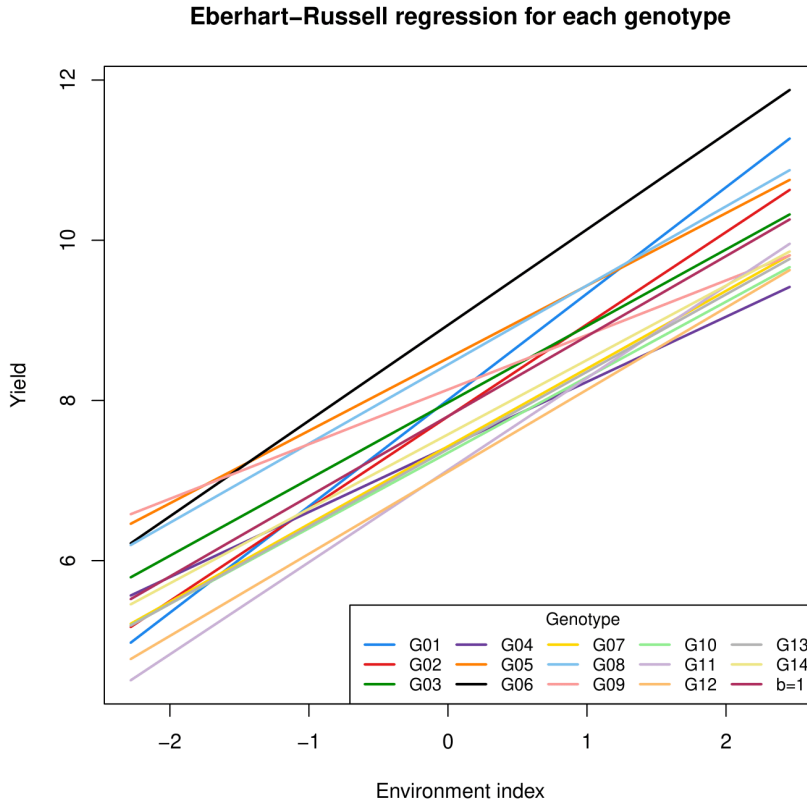


Figure 9.3: Eberhart-Russell regression plot

9.4 Plot of regression coefficient and response value

The plot of regression coefficient and response value (yield) (Figure 9.4) shows the position of each genotype based on its regression coefficient

and response value. The vertical gray line is $\bar{Y} \pm LSD_{0.05}$ and the horizontal gray line is the 95% confidence interval for $b_i = 1$. In Figure 9.4, it can be seen that the genotypes with b_i values that are not significantly different from 1 are G03, G05, G07, G08, G10, G12, G13, G14. Among these genotypes, the ones with high yields are G08 and G05. Of the two genotypes, G05 meets the two Eberhart-Russell stability criteria, namely $b_i = 1$ and $s^2_{di} = 0$.

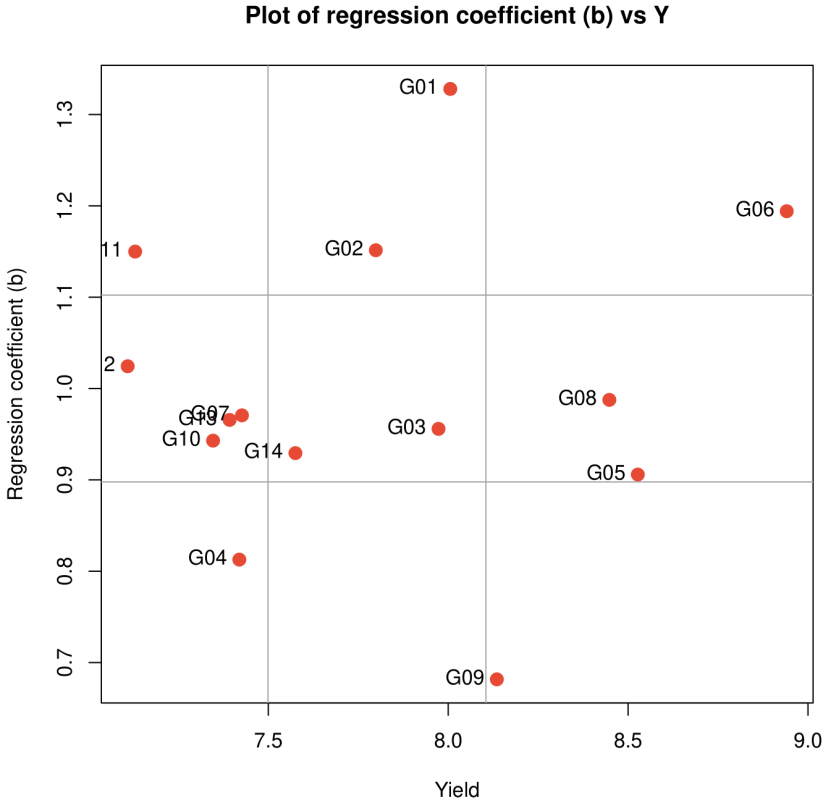


Figure 9.4: Plot of regression coefficient and yield

10 Francis-Kannenberg stability analysis

Francis and Kannenberg (1978) put forward the concept of stability based on the coefficient of variation (CV) and yield values. The CV mean line and the yield mean line divide the genotypes into 4 groups, namely: Group 1 (top left): high yield, small CV, Group 2 (top right): high yield, large CV, Group 3 (bottom left), low yield, small CV, and Group 4: low yield, large CV (bottom right) (Figure 10.1). According to Francis-Kannenberg, stable genotypes are those in Group 1. Thus, in our example, genotypes G03, G05, G08, and G09 are classified as stable based on this parameter.

From the results of their research on 15 corn hybrids, Finlay and Wilkinson (1978) stated that the genotypes in Group 1 had an average of $b_i = 1.00$ and s^2_{di} and σ^2_i that were not significantly different from 0. This indicates the possibility that a genotype that is stated to be stable based on Finlay and Wilkinson is also categorized as stable based on Eberhart and Russell (1966) and Shukla (1972).

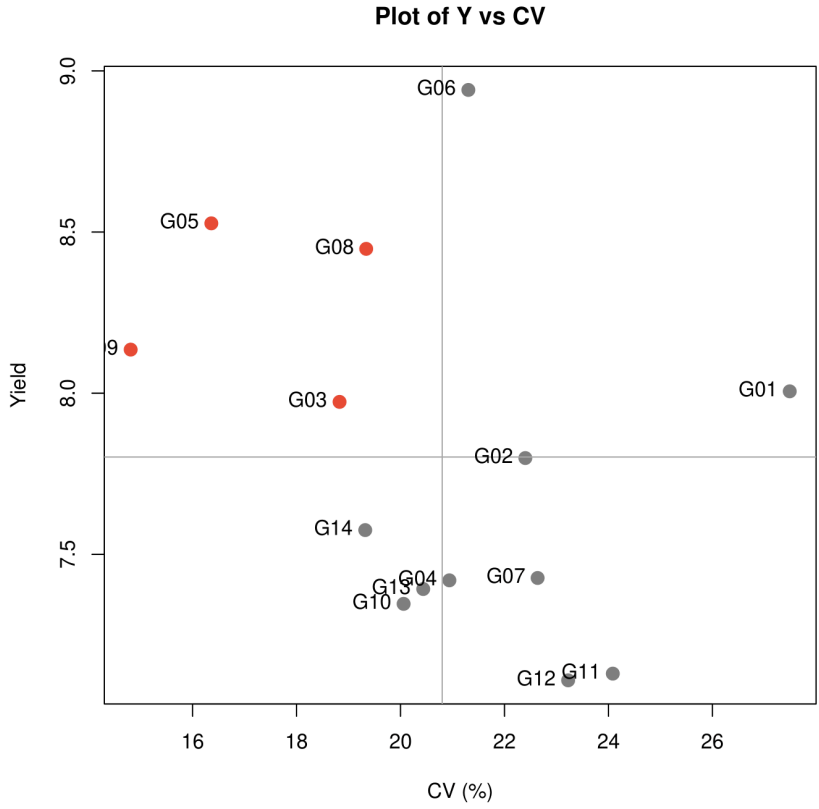


Figure 10.1: Plot of Y vs coefficient of variation

11 AMMI stability analysis

11.1 ANOVA and PCA

Additive main effect and multiplicative interaction (AMMI) is a multivariate stability analysis method to evaluate genotype stability in a number of environments. This method combines ANOVA to explain the additive effects of genotype and environment, and principal component analysis to explain the effects of genotype x environment (GxE) interactions (Gauch and Zobel, 1988). The results of the AMMI variance analysis for the example case of multilocation rice test data are shown in Figure 11.1.

11.2 AMMI biplot: PC1 and PC2

AMMI biplot PC1 and PC2 (Figure 11.2) depicts the interaction of genotype and environment that can be explained by two principal components. PC1 and PC2 explain 40.8% and 21.2% of the variability of GxE interactions, respectively, so the total variability of GxE explained by the two PCs is 62.0%. This shows that there is still 38.0% of GxE variability that cannot be explained by this biplot. However, this biplot may still be able to provide an overview of GxE interactions to some extent.

The coordinates of each environment are marked with green dots (boxes). The direction of the vector is from the center point (0,0) to that point. The coordinates of each genotype are marked with orange dots (circles). Genotypes located inside the ellipse around the center point are relatively stable genotypes (exhibit low GxE). Genotypes

Combined anova with AMMI analysis

	Df	Sum Sq	Percent	Accum	Mean Sq	F value	Pr(>F)	
Environment (E)	10	880.43			88.043	37.1152	1.863e-11	***
Replication/E	22	52.19			2.372	4.2695	3.554e-09	***
Genotype (G)	13	134.23			10.325	5.5882	5.168e-08	***
GxE	130	240.20			1.848	3.3255	< 2.2e-16	***
PC1	22	97.90	40.757	40.757	4.450	8.0091	< 2.2e-16	***
PC2	20	50.92	21.198	61.956	2.546	4.5822	1.845e-09	***
PC3	18	27.75	11.552	73.508	1.542	2.7746	0.0001915	***
PC4	16	23.52	9.793	83.301	1.470	2.6461	0.0006731	***
PC5	14	16.99	7.074	90.375	1.214	2.1845	0.0085040	**
PC6	12	12.44	5.177	95.553	1.036	1.8652	0.0383969	*
PC7	10	7.25	3.019	98.571	0.725	1.3050	0.2271341	
PC8	8	1.93	0.805	99.376	0.242	0.4350	0.8995621	
PC9	6	1.20	0.500	99.876	0.200	0.3602	0.9036268	
PC10	4	0.30	0.124	100.000	0.074	0.1340	0.9697770	
Residuals	286	158.91			0.556			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure 11.1: AMMI ANOVA

G11, G02, G13 appear to be included in this group. Genotypes located close to the end of the environmental vector are thought to be adaptive (have a positive interaction effect) with the environment. For example, G04 has a positive interaction ('suitable') with E11, and G12 with E04. However, the direction of the vectors E11 and E04 are opposite, so G04 has a negative interaction ('not suitable') with E04, as does G12 with E11. However, it should be noted that the unexplained variability by the PC1 and PC2 biplots is still quite large, so in this case, AMMI analysis is not recommended as the only basis for decision-making in selection.

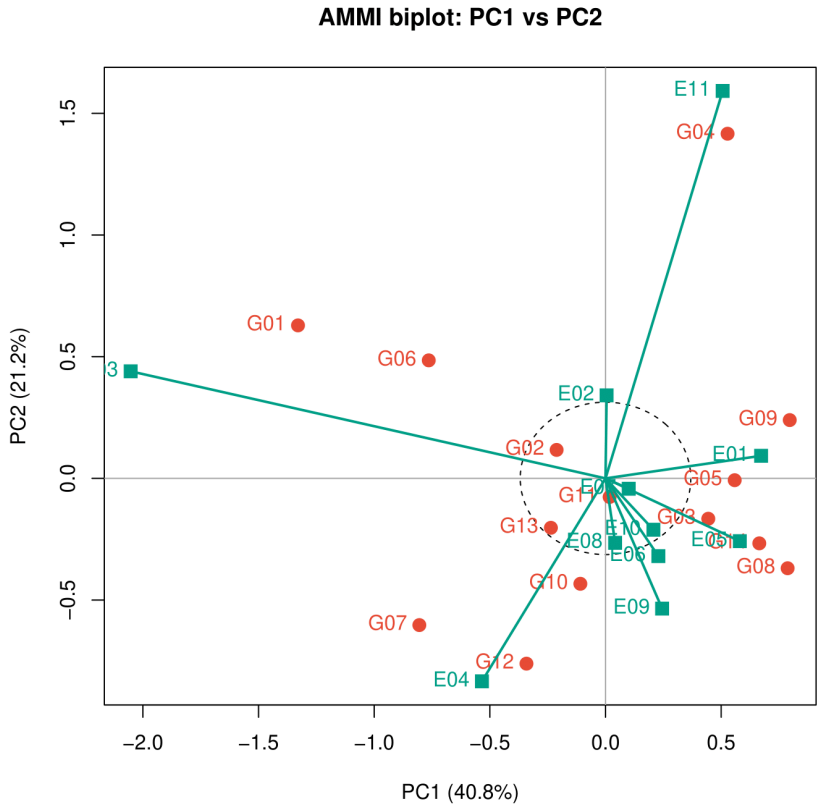


Figure 11.2: AMMI biplot: PC1 vs PC2

11.3 AMMI biplot: PC1 and Y

AMMI biplot PC1 and Y (Figure 11.3) shows the stability and response values of the genotypes. The horizontal line Y=0 shows stability, and the vertical line X=7.8 shows the average yield. The genotype with the highest yield is the one on the far right, and the most stable genotype is the one closest to the horizontal line Y=0. In this example, the ideal genotype (high yield and stable) based on AMMI analysis does not seem to exist, but the closest ones are G06 (high yield but less stable) and G05 (lower yield but more stable). From the environmental side, E06 is on the far right, which shows that the environment has a high average yield (7.80 tons/ha). When interpreting, we need to remember that PC1 only explains 40.8% of the GxE variability, so there is still a large proportion of variability that cannot be studied in this plot.

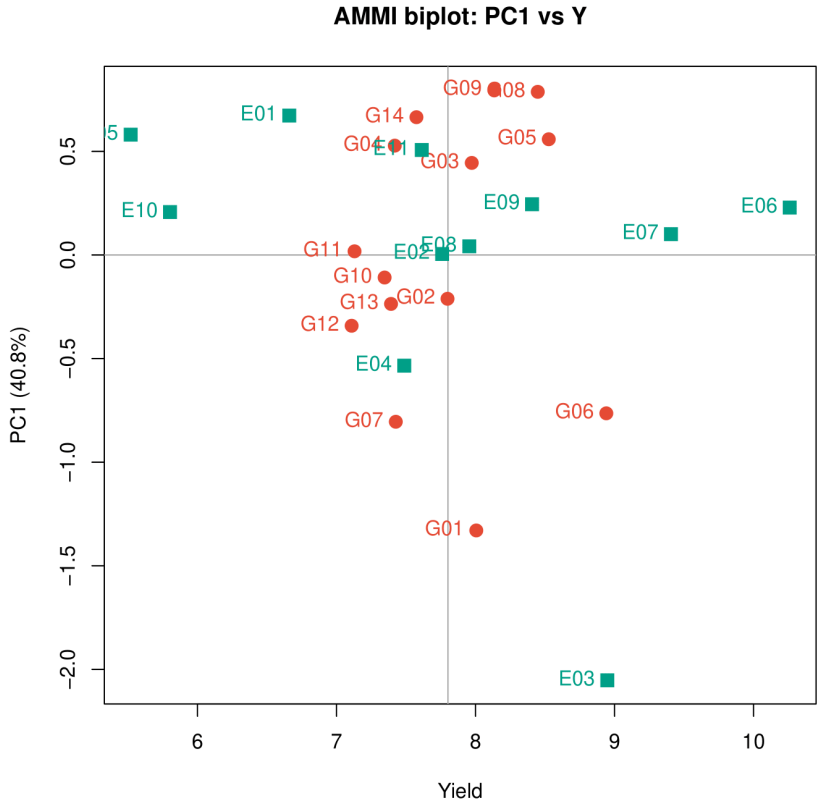


Figure 11.3: AMMI biplot: PC1 vs Yield

12 GGE stability analysis

12.1 GGE vs AMMI

The stability analysis of genotype and genotype by environment (GGE) is similar to AMMI, which uses principal component analysis (PCA). The difference is, that AMMI uses PCA to describe the influence of GxE, while GGE uses PCA to describe the influence of G+GxE. Therefore, the PC1 and PC2 biplots of AMMI can only describe stability/adaptability, while the PC1 and PC2 biplots of GGE can describe the average and stability/adaptability at the same time.

12.2 GGE biplot: Which-won-where

In this example, PC1 vs PC2 biplot explains 63.8% of G+GE variance. There is a remaining 36.8% unexplained G+GE variance, so the interpretation of this biplot should be taken with caution. Genotypes close to the center point are more stable than those far from the center point. The genotype yield is greater than the average if the angle of the genotype vector and the environment vector is $< 90^{\circ}$, the genotype yield is lower than the average if the angle of the genotype vector and the environment vector is $> 90^{\circ}$, and the genotype yield is close to the average if the angle of the genotype vector and the environment vector is around 90° (Pacheco et al, 2015).

The polygon is divided into sectors ('mega-environments') (Figure 12.1). The best genotype in each mega-environment is the one at the corner of the polygon. Genotype G06 is located at one of

the corners of the polygon in a sector that contains all environments except E01 and E05. This indicates that the genotype has a greater yield than the average in most environments. In contrast, the G12 genotype, which lies opposite to G06, has a lower yield than the average in most environments.

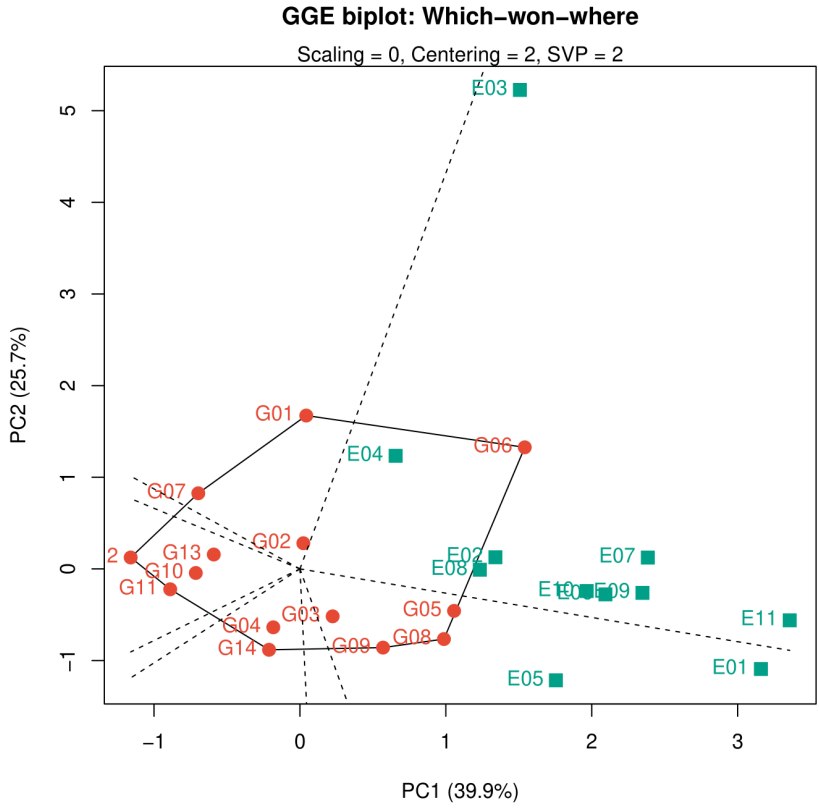


Figure 12.1: GGE biplot: Which-won-where

12.3 GGE biplot: Discriminativeness vs. representativeness

GGE biplot: Discriminativeness vs. representativeness (Figure 12.2) tells us about the environments. The angle between two environments shows the correlation between them. E03 is correlated with E04 (angle $< 90°$) and uncorrelated with E05 (angle $90°$). In this example, no two environments are negatively correlated (angle $> 90°$). Positive correlations between multiple locations indicate that the same information about the genotypes being tested can be obtained from a smaller number of environments. Negative correlations between locations indicate strong GxE, where a genotype that is adaptive at one location is likely not adaptive at the other location (Yan and Tinker, 2006).

The length of the vector indicates the ability of the environment to discriminate genotypes. Environments with long vectors are more discriminating (informative in distinguishing genotypes), while environments with short vectors are less discriminating. Environments that are consistently uninformative are not recommended for use as test environments. The circles help visualize the length of the vectors (Yan and Tinker, 2006).

The average environment axis (AEA) is depicted by a line through the center point and the average environment (AE) point. Environments with a more acute angle with AEA (for example E02) are more representative environments. An informative and representative test environment is an environment with a long vector and an acute angle with AEA (E11) which is considered good for selecting genotypes that are generally adapted (Yan and Tinker, 2006).

Informative but unrepresentative test environments (long vectors but angles with obtuse AEAs) are considered good for selecting environment-specific genotypes if the environment can be divided into 'mega-environments', or identifying unstable genotypes if the environment is a single 'mega-environment' (Yan and Tinker, 2006).

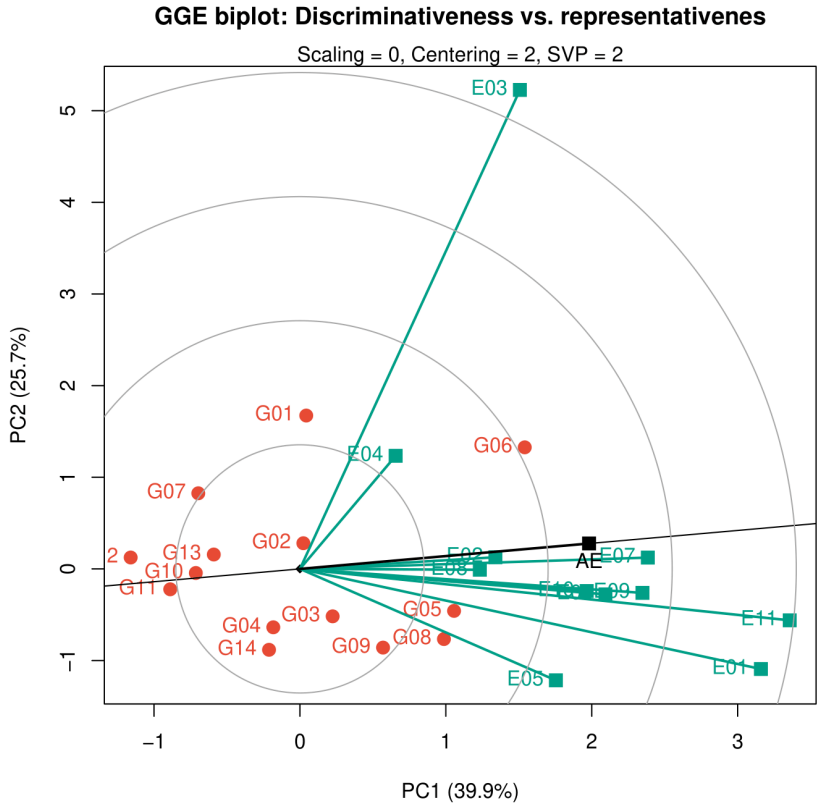


Figure 12.2: GGE biplot: Discriminativeness vs. representativeness

12.4 GGE biplot: Mean vs. stability

GGE biplot: Mean vs stability (Figure 12.3) projects the mean and stability of each genotype. The desired genotypes are those that are stable (close to the horizontal line) and high-yielding (on the right). In this example, G06 is closest to the ideal genotype.

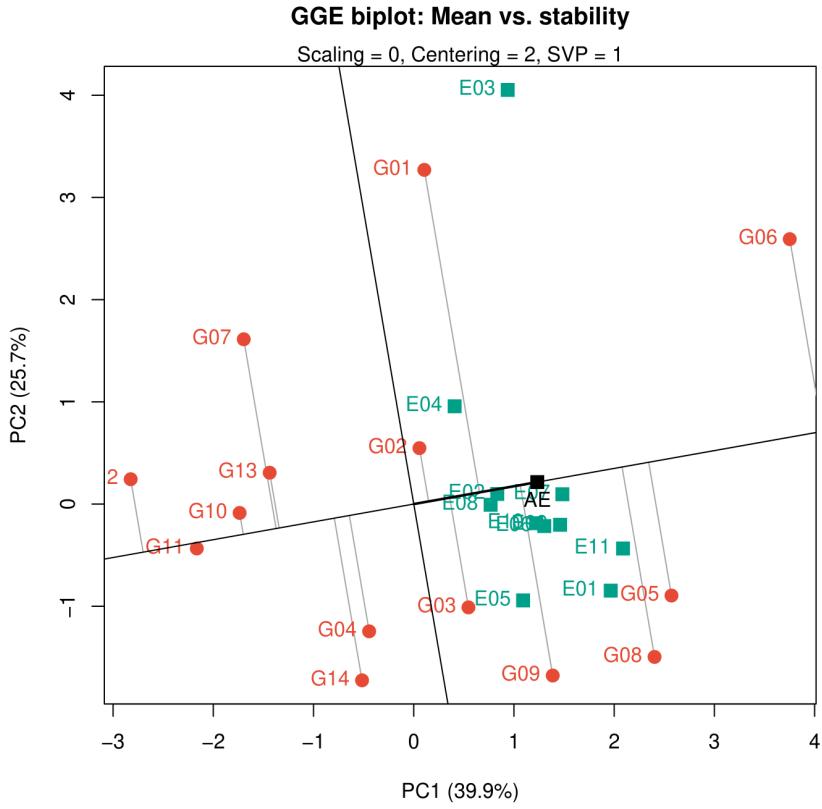


Figure 12.3: GGE biplot: Mean vs. stability

12.5 GGE biplot: Ranking genotypes

The ideal genotype (IG) is the most stable genotype and has the highest yield. This genotype is visualized as the IG point in the center of the circle (on the right, but not visible in Figure 12.4). The IG point comes from the projection of the genotype vector that has the highest average in AEA. The genotype closest to the IG point is the genotype closest to the ideal (Yan and Tinker, 2006).

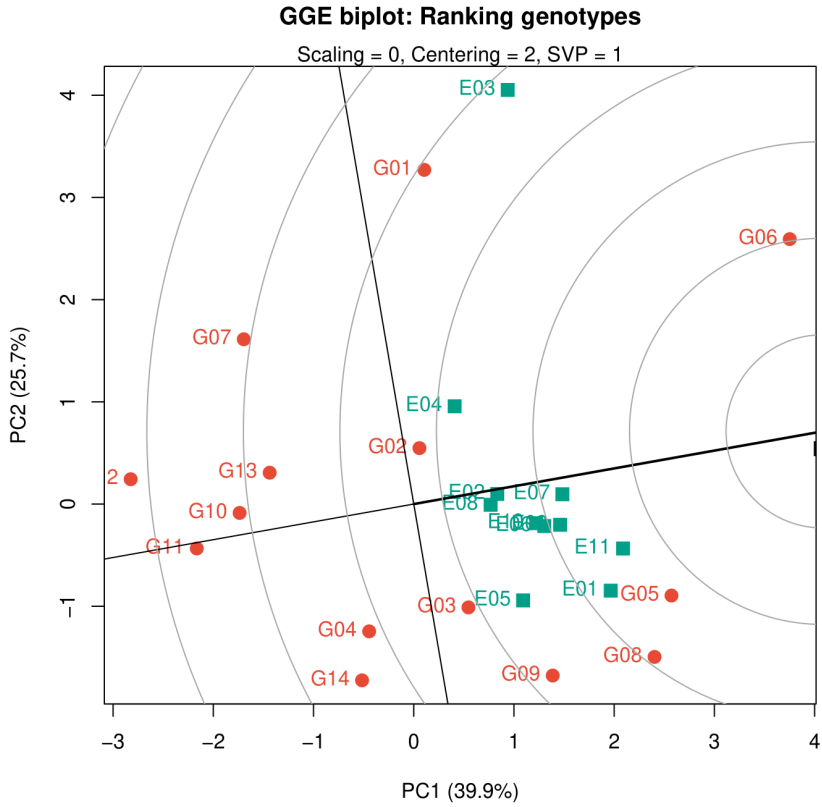


Figure 12.4: GGE biplot: Ranking genotypes

References

Aswidinnoor, H., Listiyanto, R., Rahim, S., Holidin, Setiyowati, H., Nindita, A., Ritonga, A. W., Marwiyah, S., & Suwarno, W. B. (2023). Stability analysis, agronomic performance, and grain quality of elite new plant type rice lines (*Oryza sativa* L.) developed for tropical lowland ecosystem. Frontiers in Sustainable Food Systems, 7, 1147611.

Eberhart, S. T., & Russell, W. (1966). Stability parameters for comparing varieties. *Crop Science*, *6*(1), 36-40.

Fehr, W. (1991). *Principles of cultivar development: Theory and technique.*

Finlay, K. W., & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, *14*(6), 742-754.

Francis, T. R., & Kannenberg, L. W. (1978). Yield stability studies in short-season maize. I. A descriptive method for grouping genotypes. *Canadian Journal of Plant Science*, *58*(4), 1029-1034.

Gauch, H. G., & Zobel, R. W. (1988). Predictive and postdictive success of statistical analyses of yield trials. *Theoretical and Applied Genetics*, *76*, 1-10.

Gomez, K. A., & Gomez, A. A. (1984). *Statistical procedures for agricultural research*. John Wiley & Sons.

Olivoto, T., & Lúcio, A. D. C. (2020). metan: An R package for multi-environment trial analysis. *Methods in Ecology and Evolution*, *11*(6), 783-789.

Pacheco, Á., Vargas, M., Alvarado, G., Rodríguez, F., Crossa, J., & Burgueño, J. (2015). *GEA-R (Genotype x Environment Analysis with R for Windows)*. (Version 4.1). [Computer software]. CIMMYT.

Pour-Aboughadareh, A., Khalili, M., Poczai, P., & Olivoto, T. (2022). Stability indices to deciphering the genotype-by-environment interaction (GEI) effect: An applicable review for use in plant breeding programs. *Plants*, *11*(3), 414.

Yan, W., & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian Journal of Plant Science*, *86*(3), 623-645.

ISBN 979-889705285-1

Primedia eLaunch LLC