Genotype x Environmental Analysis with R for Windows

Biometrics and Statistics Unit
Angela Pacheco

CIMMYT,Int.

23-24 Junio 2015
In agricultural experimentation, a large number of genotypes are normally tested over a wide range of environments.

Advanced breeding materials must be evaluated in multiple locations for more than one year.

The occurrence of the GEI effect further complicates the selection of superior genotypes for a target population of environments.
The presence of the GEI confirms particular genotypes being superior in particular environments.

Several statistical methods were developed for resolved that kind of problems, for example, AMMI, SREG, PLS, Stability analysis and Factorial regression.
GEA-R description

http://hdl.handle.net/11529/10203

GEA-R is a set of R programs that performs statistical analyses to calculate AMMI, SREG, PLS, Stability Analysis, Factorial Regression, in order to make boxplots and histograms. GEA-R contains a graphical JAVA interface that helps the user to easily choose input files, which analyses to implement, and which variables to analyze.
GEA-R requirements and installation

- Windows operating system.
- For install, you only need to follow the instructions.
- When you install GEA-R, you will can start the program directly of the programs window.
GEA-R requirements and installation

- Or you will search the folder in C:\GEA-R, the folder contains the following files:

- When using GEA-R for the first time, you must load (just once) the R-version that GEA-R will use to perform the analysis, or in case that you don’t have installed R, GEA-R installed automatically.
GEA-R requirements and installation

The R-version is loaded by clicking on ‘Load R’. GEA-R will display a list of the R-versions detected as installed, usually in the Program Files folder. To look for installed R-versions located in another folder, click on the ‘Browse’ button.
GEA-R requirements and installation

Once the R-version is loaded, GEA-R will verify if all packages exist and will automatically try to install them if they do not.

When the R-version is loaded and the packages are ready, the next step is to open the input file to be analyzed.
Whenever you need help, you can click in "Help" button, and GEA-R show you two options: "Manual", in it you find step by step an explanation for do the analyses; and "About...", in it you find the citation and license for GEA-R.
Preparing for the analysis
Preparing for the analysis
Additive main effect and multiplicative interaction analysis (AMMI)
AMMI

- Method for analyzing GEI to identify patterns of interaction and reduce background noise.
- Combines conventional ANOVA with principal component analysis.
- May provide more reliable estimates of genotype performance than the mean across sites.
AMMI

- Biplots help to visualize relationships among genotypes and environments; show both main and interaction effects.
- Enables you to identify target breeding environments and to choose representative testing sites in those environments.
- Enables you to select varieties with good adaptation to target breeding environments.
AMMI Model

\[ Y_{ijk} = \mu + G_i + E_j + \sum \lambda_k \alpha_{ik} \gamma_{jk} + d_{ij} + e_{ijk} \]

- \( \mu \) the grand mean.
- \( G_i \) the genotype deviations from the grand mean.
- \( E_j \) the environment deviations from the grand mean.
- \( \lambda_k \) kth eigenvalue.
- \( \alpha_{ik} \) principal component score for the \( i^{th} \) genotype for the \( k^{th} \) principal component axis.
- \( \gamma_{jk} \) principal component score for the \( j^{th} \) environment for the \( k^{th} \) principal component axis.
- \( d_{ij} \) residual GEI not explained by model.
- \( e_{ijk} \) residual model.
AMMI: General Interpretation

- Usually need only the first 2 PC axes to adequately explain the data.
- Genotypes that occur close to particular environments on the biplot show specific adaptation to those environments.
- A genotype that falls near the center of the biplot (small PC1 and PC2 values) may have broader adaptation.
AMMI: General Interpretation

- Acute angles indicate positive correlation, with vectors in exactly the same directions representing a correlation of 1.
- Obtuse angles represent negative correlations, with opposite directions indicating a correlation of -1.
- Perpendicularly of directions indicates a correlation of 0.
Sites Regression Model (SREG)
This model is recommended when the environments are the main source of variation in relation to the contributions of the genotypes and the GEI with respect to the total variability.

The biplot graphics is used to visualize patterns and interactions without environmental effects.
In addition, as a difference with AMMI model, this technique allows the detection of GEI in terms of the crossover effect resulting from great changes in the ranking of the genotypes across the environments.
SREG Model

\[ Y_{ijk} = \mu + E_j + \sum \lambda_k \alpha_{ik} \gamma_{jk} + d_{ij} + e_{ijk} \]

- \( \mu \): the grand mean.
- \( E_j \): the environment deviations from the grand mean.
- \( \lambda_k \): kth eigenvalue.
- \( \alpha_{ik} \): principal component score for the \( i^{th} \) genotype for the \( k^{th} \) principal component axis.
- \( \gamma_{jk} \): principal component score for the \( j^{th} \) environment for the \( k^{th} \) principal component axis.
- \( d_{ij} \): residual GEI not explained by model.
- \( e_{ijk} \): residual model.
SREG: General Interpretation

- Usually the first principal component represents responses of the genotypes that are proportional to the environments, which are associated with the GEI without change of the range.
- The second principal component provides information about cultivation locations that are not proportional to the environments, indicating that those are responsible of the GEI crossover interaction.
SREG: General Interpretation

This technique allows the determination of mega-environments, which mean parts of the cultivation area of a species that show homogeneous environmental conditions, where the performance of certain genotypes is similar through the years.
Partial Least Square (PLS)
PLS

- Is a model for to incorporate external variables (environmental or genotypic).
- Is convenient if you have more variables than observations and there is high collinearity among variables.
- Can be used for to determinate which of these external variables influence in the GEI.
Factorial Regression (FR)
Also helps us to incorporate environmental and genotypic covariates and allow us to choose which covariate provides greater proportion of the variability in the GEI.

FR is easier to interpret than PLS, but may give misleading results when there are correlations among the explanatory variables in the model.
Similar to stepwise multiple regression, where additional variables are added to the model in sequence until sufficient variability due to GEI can be explained.
GEA-R

Types of Analyses

FR Model

\[ Y_{ij} = \mu + G_i + E_j + \sum_{g=1}^{Gc} x_{ig} \xi_{ig} + e_{ij} \]

- \( \mu \) the grand mean.
- \( G_i \) and \( E_j \) the genotype and environment deviations from the grand mean.
- \( x_{ig} \) genotypic covariates
- \( \xi_{ig} \) the environmental factor
- \( e_{ij} \) residual model.
- \( Gc \) number of genotypic covariates (Gc<genotypes)
Stability Analysis

Analyses Options
- Model: Stability Analysis
- Plots Format: PDF (*.pdf), PNG (*.png), Metafile (*.wmf)

Response Variables
- Select one or more: YLD
- Selected

Locations
- Select one or more: Batan, Toluca, Obregon, AguaFria, Tlatizapan, Celaya, Tuxtlas, Armecha
- Selected
## Stability parameters

<table>
<thead>
<tr>
<th>Stability parameter</th>
<th>Symbol</th>
<th>Increased stability if:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variation coefficient</td>
<td>CV</td>
<td>small values</td>
</tr>
<tr>
<td>Shuckla's variance</td>
<td>$\sigma^2$</td>
<td>small values</td>
</tr>
<tr>
<td>Perkins and Jinks</td>
<td>B</td>
<td>values close to 1</td>
</tr>
<tr>
<td>Perkins and Jinks</td>
<td>D</td>
<td>small values</td>
</tr>
<tr>
<td>Eberhart and Russell (regression coefficient)</td>
<td>b</td>
<td>values close to 1</td>
</tr>
<tr>
<td>Eberhart and Russell (mean square deviation)</td>
<td>$S_d^2$</td>
<td>small values</td>
</tr>
<tr>
<td>Determination coefficient</td>
<td>$R^2$</td>
<td>big values</td>
</tr>
<tr>
<td>Wricke's ecovalence</td>
<td>W</td>
<td>small values</td>
</tr>
<tr>
<td>Superiority measure</td>
<td>P</td>
<td>small values</td>
</tr>
<tr>
<td>Average absolute rank difference of genotype on the environment</td>
<td>$S^{(1)}$</td>
<td>small values</td>
</tr>
<tr>
<td>Variance ranges of environments</td>
<td>$S^{(2)}$</td>
<td>small values</td>
</tr>
</tbody>
</table>
The GGEBiplotGUI package provides a graphical user interface for the construction of, interaction with, and manipulation of GGE biplots in R. Some of the functions of the package are:

- Ranking the cultivars based on their performance in any given environment.
- Ranking the environments based on the relative performance of any given cultivar.
Comparing the performance of any pair of cultivars in different environments.
Identifying the best cultivar in each environment.
Grouping the environments based on the best cultivars.
GEA-R

Types of Analyses

GGEBiplot

- Evaluating the cultivars based on both average yield and stability.
- Evaluating the environments based on both discriminating ability and representativeness.
It’s ready?

Successful Analysis!

Outputs were saved in:
'C:\Users\rapacheco\Desktop\GEA-R\GxE-1.1\Output\Analysis1'
It’s ready?

- If you have problems with your data set.
## Summary Results

<table>
<thead>
<tr>
<th>AMMI</th>
<th>SREG</th>
<th>PLS</th>
<th>Stability Analysis</th>
<th>Factorial Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Score</td>
<td>Final Scores</td>
<td>Biplot PLS Final Scores</td>
<td>Stability Coefficients Plot CV</td>
<td>ANOVA for general model</td>
</tr>
<tr>
<td>Gollob Test</td>
<td>Gollob Test</td>
<td></td>
<td>Plot Eberhart&amp;Russel Coefficient</td>
<td>ANOVA for Best model</td>
</tr>
<tr>
<td>Biplot PC1 vs PC2</td>
<td>Biplot PC1 vs PC2</td>
<td></td>
<td></td>
<td>Summary of steps for factorial regression</td>
</tr>
<tr>
<td>Biplot PC1 vs PC3</td>
<td>Biplot PC1 vs PC3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biplot PC2 vs PC3</td>
<td>Biplot PC2 vs PC3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biplot PC1 vs Yield</td>
<td>Biplot PC1 vs Yield</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biplot 3D</td>
<td>Biplot 3D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Biplots with polygon</td>
<td>All Biplots with polygon</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
THANK YOU!!