

# Package ‘geneticae’

February 9, 2022

**Title** Statistical Tools for the Analysis of Multi Environment  
Agronomic Trials

**Version** 0.3.0

**Description** Data from multi environment agronomic trials, which are often carried out by plant breeders, can be analyzed with the tools offered by this package such as the Additive Main effects and Multiplicative Interaction model or 'AMMI' ('Gauch' 1992, ISBN:9780444892409) and the Site Regression model or 'SREG' ('Cornelius' 1996, <[doi:10.1201/9780367802226](https://doi.org/10.1201/9780367802226)>). Since these methods present a poor performance under the presence of outliers and missing values, this package includes robust versions of the 'AMMI' model ('Rodrigues' 2016, <[doi:10.1093/bioinformatics/btv533](https://doi.org/10.1093/bioinformatics/btv533)>), and also imputation techniques specifically developed for this kind of data ('Arciniegas-Alarcón' 2014, <[doi:10.2478/bile-2014-0006](https://doi.org/10.2478/bile-2014-0006)>).

**License** GPL-2

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.2

**Imports** stats, GGEbiplots, ggforce, ggplot2, scales, MASS, pcaMethods, rrcov, dplyr, bcv, missMDA, calibrate, graphics, reshape2, matrixStats, tidyr, prettydoc

**Suggests** agridat, spelling, knitr, rmarkdown, testthat

**VignetteBuilder** knitr

**Depends** R (>= 2.12.0)

**URL** <https://jangelini.github.io/geneticae/>

**BugReports** <https://github.com/jangelini/geneticae/issues>

**Language** en-US

**NeedsCompilation** no

**Author** Julia Angelini [aut, cre] (<<https://orcid.org/0000-0002-5815-1771>>),  
Marcos Prunello [aut] (<<https://orcid.org/0000-0002-9611-527X>>),  
Gerardo Cervigni [aut]

**Maintainer** Julia Angelini <jangelini\_93@hotmail.com>

**Repository** CRAN

**Date/Publication** 2022-02-09 22:30:13 UTC

## R topics documented:

GGEmodel . . . . .	2
GGEPlot . . . . .	4
imputation . . . . .	6
plrv . . . . .	9
rAMMI . . . . .	10
<b>Index</b>	<b>13</b>

---

GGEmodel	<i>Site Regression model</i>
----------	------------------------------

---

## Description

The Site Regression model (also called genotype + genotype-by-environment (GGE) model) is a powerful tool for effective analysis and interpretation of data from multi-environment trials in breeding programs. This function is a wrapper for [GGEmodel](#) from the [GGEbiplots package](#), with the following enhancements:

- It can be used for data from trials with repetitions (there is no need to calculate means beforehand)
- Other variables not used in the analysis can be present in the dataset.
- Variables can be in any order in the dataset.

## Usage

```
GGEmodel(
  Data,
  genotype = "gen",
  environment = "env",
  response = "yield",
  rep = NULL,
  centering = "tester",
  scaling = "none",
  SVP = "symmetrical"
)
```

**Arguments**

Data	dataframe with genotypes, environments, repetitions (if any) and the phenotypic trait of interest. Variables can be in any order, even additional variables that will not be used in the model may be present in the data.
genotype	column name for genotypes.
environment	column name for environments.
response	column name for the phenotypic trait.
rep	column name for replications. If this argument is NULL, there are no replications in the data. Defaults to NULL.
centering	centering method. Either "tester" for tester centered (G+GE), "global" for global centered (E+G+GE), "double" for double centered (GE) or "none" for no centering. Defaults to "tester".
scaling	scaling method. Either "sd" for standard deviation or "none" for no scaling. Defaults to "none".
SVP	method for singular value partitioning. Either "row", "column", "dual" or "symmetrical". Defaults to "symmetrical".

**Value**

A list of class GGE\_Model containing:

coordgenotype	plotting coordinates for each genotype in every component.
coordenvironment	plotting coordinates for each environment in every component.
eigenvalues	vector of eigenvalues for each component.
vartotal	overall variance.
varexpl	percentage of variance explained by each component.
labelgen	genotype names.
labelenv	environment names.
axes	axis labels.
Data	scaled and centered input data.
centering	name of centering method.
scaling	name of scaling method.
SVP	name of SVP method.

**References**

- Sam Dumble (2017). GGEBiplots: GGE Biplots with 'ggplot2'. R package version 0.1.1. <https://CRAN.R-project.org/package=GGEBiplots>
- Yan W, Kang M (2003). *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists*. CRC Press.
- Yan W, Kang M (2002). *Singular-Value Partitioning in Biplot Analysis of Multienvironment Trial Data*. Agronomy Journal, 94, 990-996. doi: [10.2134/agronj2002.0990](https://doi.org/10.2134/agronj2002.0990)

## Examples

```
library(geneticae)
# Data without replication
library(agridat)
data(yan.winterwheat)
GGE1 <- GGEmodel(yan.winterwheat, genotype = "gen", environment = "env",
                response = "yield", centering = "tester")

# Data with replication
data(plrv)
GGE2 <- GGEmodel(plrv, genotype = "Genotype", environment = "Locality",
                response = "Yield", rep = "Rep", centering = "tester")
```

---

GGEPlot

*GGE biplots with ggplot2*

---

## Description

GGE biplots are used for visual examination of the relationships between test environments, genotypes, and genotype-by-environment interactions. ‘GGEPlot()’ produces a biplot as an object of class ‘ggplot’, using the output of the `GGEmodel` function. Several types of biplots are offered which focus on different aspects of the analysis. Customization options are also included. This function is a modification of `GGEPlot` from the [hrefhttps://CRAN.R-project.org/package=GGEbiplots](https://CRAN.R-project.org/package=GGEbiplots) package.

## Usage

```
GGEPlot(
  GGEmodel,
  type = "Biplot",
  d1 = 1,
  d2 = 2,
  selectedE = NA,
  selectedG = NA,
  selectedG1 = NA,
  selectedG2 = NA,
  colGen = "gray47",
  colEnv = "darkred",
  colSegment = "gray30",
  colHull = "gray30",
  sizeGen = 4,
  sizeEnv = 4,
  largeSize = 4.5,
  axis_expand = 1.2,
  axislabels = TRUE,
```

```

axes = TRUE,
limits = TRUE,
titles = TRUE,
footnote = TRUE
)

```

## Arguments

GGEModel	An object of class GGEModel.
type	type of biplot to produce. <ul style="list-style-type: none"> <li>• "Biplot": Basic biplot.</li> <li>• "Selected Environment": Ranking of cultivars based on their performance in any given environment.</li> <li>• "Selected Genotype": Ranking of environments based on the performance of any given cultivar.</li> <li>• "Relationship Among Environments".</li> <li>• "Comparison of Genotype".</li> <li>• "Which Won Where/What": Identifying the 'best' cultivar in each environment.</li> <li>• "Discrimination vs. representativeness": Evaluating the environments based on both discriminating ability and representativeness.</li> <li>• "Ranking Environments": Ranking environments with respect to the ideal environment.</li> <li>• "Mean vs. stability": Evaluating cultivars based on both average yield and stability.</li> <li>• "Ranking Genotypes": Ranking genotypes with respect to the ideal genotype.</li> </ul>
d1	PCA component to plot on x axis. Defaults to 1.
d2	PCA component to plot on y axis. Defaults to 2.
selectedE	name of the environment to evaluate when 'type="Selected Environment"'.
selectedG	name of the genotype to evaluate when 'type="Selected Genotype"'.
selectedG1,	name of the genotype to compare to 'selectedG2' when 'type="Comparison of Genotype"'.
selectedG2,	name of the genotype to compare to 'selectedG1' when 'type="Comparison of Genotype"'.
colGen	genotype attributes colour. Defaults to "gray47".
colEnv	environment attributes colour. Defaults to "darkred".
colSegment	segment or circle lines colour. Defaults to "gray30".
colHull	hull colour when 'type="Which Won Where/What"'. Defaults to "gray30".
sizeGen	genotype labels text size. Defaults to 4.
sizeEnv	environment labels text size. Defaults to 4.
largeSize	larger labels text size to use for two selected genotypes in 'type="Comparison of Genotype"', and for the outermost genotypes in 'type="Which Won Where/What"'. Defaults to 4.5.

axis_expand	multiplication factor to expand the axis limits by to enable fitting of labels. Defaults to 1.2.
axislabels	logical, if this argument is 'TRUE' labels for axes are included. Defaults to 'TRUE'.
axes	logical, if this argument is 'TRUE' x and y axes going through the origin are drawn. Defaults to 'TRUE'.
limits	logical, if this argument is 'TRUE' the axes are re-scaled. Defaults to 'TRUE'.
titles	logical, if this argument is 'TRUE' a plot title is included. Defaults to 'TRUE'.
footnote	logical, if this argument is 'TRUE' a footnote is included. Defaults to 'TRUE'.

**Value**

A biplot of class ggplot

**References**

Yan W, Kang M (2003). *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists*. CRC Press.

Sam Dumble (2017). GGEBiplots: GGE Biplots with 'ggplot2'. R package version 0.1.1. <https://CRAN.R-project.org/package=GGEBiplots>

**Examples**

```
library(geneticae)

# Data without replication
library(agridat)
data(yan.winterwheat)
GGE1 <- GGEmodel(yan.winterwheat, centering = "tester")
GGEPlot(GGE1)

# Data with replication
data(plrv)
GGE2 <- GGEmodel(plrv, genotype = "Genotype", environment = "Locality",
                 response = "Yield", rep = "Rep", centering = "tester")
GGEPlot(GGE2)
```

---

imputation

*Imputation of missing cells in two-way data sets*


---

**Description**

Missing values are not allowed by the AMMI or GGE methods. This function provides several methods to impute missing observations in data from multi-environment trials and to subsequently adjust the mentioned methods.

**Usage**

```

imputation(
  Data,
  genotype = "gen",
  environment = "env",
  response = "yield",
  rep = NULL,
  type = "EM-AMMI",
  nPC = 2,
  initial.values = NA,
  precision = 0.01,
  maxiter = 1000,
  change.factor = 1,
  simplified.model = FALSE,
  k = min(nrow(Data), ncol(Data)),
  scale = TRUE,
  method = "EM",
  row.w = NULL,
  coeff.ridge = 1,
  seed = NULL,
  nb.init = 1,
  Winf = 0.8,
  Wsup = 1
)

```

**Arguments**

Data	dataframe containing genotypes, environments, repetitions (if any) and the phenotypic trait of interest. Columns can be in any order in the dataset and other variables that will not be used in the analysis can be present.
genotype	column name containing genotypes.
environment	column name containing environments.
response	column name containing the phenotypic trait.
rep	column name containing replications. If this argument is NULL, there are no replications available in the data. Defaults to NULL.
type	imputation method. Either "EM-AMMI", "EM-SVD", "Gabriel", "WGabriel", "EM-PCA". Defaults to "EM-AMMI".
nPC	number of components used to predict the missing values. Default to 2.
initial.values	initial values of the missing cells. It can be a single value or a vector of length equal to the number of missing cells (starting from the missing values in the first column). If omitted, the initial values will be obtained by the main effects from the corresponding model, that is, by the grand mean of the observed data increased (or decreased) by row and column main effects.
precision	threshold for assessing convergence.
maxiter	maximum number of iteration for the algorithm.

<code>change.factor</code>	When ‘change.factor’ is equal to 1, the previous approximation is changed with the new values of missing cells (standard EM-AMMI algorithm). However, when ‘change.factor’ less than 1, then the new approximations are computed and the values of missing cells are changed in the direction of this new approximation but the change is smaller. It could be useful if the changes are cyclic and thus convergence could not be reached. Usually, this argument should not affect the final outcome (that is, the imputed values) as compared to the default value of ‘change.factor’ = 1.
<code>simplified.model</code>	the AMMI model contains the general mean, effects of rows, columns and interaction terms. So the EM-AMMI algorithm in step 2 calculates the current effects of rows and columns; these effects change from iteration to iteration because the empty (at the outset) cells in each iteration are filled with different values. In step 3 EM-AMMI uses those effects to re-estimate cells marked as missed (as default, <code>simplified.model=FALSE</code> ). It is, however, possible that this procedure will not converge. Thus the user is offered a simplified EM-AMMI procedure that calculates the general mean and effects of rows and columns only in the first iteration and in next iterations uses these values ( <code>simplified.model=TRUE</code> ). In this simplified procedure the initial values affect the outcome (whilst EM-AMMI results usually do not depend on initial values). For the simplified procedure the number of iterations to convergence is usually smaller and, furthermore, convergence will be reached even in some cases where the regular procedure fails. If the regular procedure does not converge for the standard initial values, the simplified model can be used to determine a better set of initial values.
<code>k</code>	rank of the SVD approximation.
<code>scale</code>	boolean. By default TRUE leading to a same weight for each variable
<code>method</code>	"Regularized" by default or "EM"
<code>row.w</code>	row weights (by default, a vector of 1 for uniform row weights)
<code>coeff.ridge</code>	1 by default to perform the regularized imputePCA algorithm; useful only if <code>method="Regularized"</code> . Other regularization terms can be implemented by setting the value to less than 1 in order to regularized less (to get closer to the results of the EM method)
<code>seed</code>	integer, by default <code>seed = NULL</code> implies that missing values are initially imputed by the mean of each variable. Other values leads to a random initialization
<code>nb.init</code>	integer corresponding to the number of random initializations; the first initialization is the initialization with the mean imputation
<code>Winf</code>	peso inferior
<code>Wsup</code>	peso superior

**Value**

imputed data matrix

**References**

Paderewski, J. (2013). An R function for imputation of missing cells in two-way data sets by EM-AMMI algorithm. *Communications in Biometry and Crop Science* 8, 60–69.



Julie Josse, Francois Husson (2016). missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *Journal of Statistical Software* 70, 1-31.

Arciniegas-Alarcón S., García-Peña M., Dias C.T.S., Krzanowski W.J. (2010). *An alternative methodology for imputing missing data in trials with genotype-by-environment interaction*. *Biometrical Letters* 47, 1–14.

Perry P.O. (2015). bcv: Cross-Validation for the SVD (Bi-Cross-Validation). R package version 1.0.1.

Arciniegas-Alarcón S., García-Peña M., Krzanowski W.J., Dias C.T.S. (2014). *An alternative methodology for imputing missing data in trials with genotype-by-environment interaction: some new aspects*. *Biometrical Letters* 51, 75-88.

## Examples

```
library(geneticae)
# Data without replications
library(agridat)
data(yan.winterwheat)

# generating missing values
yan.winterwheat[1,3]<-NA
yan.winterwheat[3,3]<-NA
yan.winterwheat[2,3]<-NA

imputation(yan.winterwheat, genotype = "gen", environment = "env",
           response = "yield", type = "EM-AMMI")

# Data with replications
data(plrv)
plrv[1,3] <- NA
plrv[3,3] <- NA
plrv[2,3] <- NA
imputation(plrv, genotype = "Genotype", environment = "Locality",
           response = "Yield", rep = "Rep", type = "EM-SVD")
```

---

plrv

*Clones from the PLRV population*

---

## Description

resistance study to PLRV (Patato Leaf Roll Virus) causing leaf curl. 28 genotypes were experimented at 6 locations in Peru. Each clone was evaluated three times in each environment, and yield, plant weight and plot were registered.

## Usage

```
data(plrv)
```

**Format**

Data frame with 504 observations and 6 variables (genotype, locality, repetition, weightPlant, weightPlot and yield).

**References**

Felipe de Mendiburu (2020). agricolae: Statistical Procedures for Agricultural Research. R package version 1.3-2. <https://CRAN.R-project.org/package=agricolae>

**Examples**

```
library(geneticae)
data(plrv)
str(plrv)
```

---

rAMMI

*AMMI biplots with ggplot2*

---

**Description**

Produces classical or robust AMMI biplot as an object of class 'ggplot', with options for customization.

**Usage**

```
rAMMI(  
  Data,  
  genotype = "gen",  
  environment = "env",  
  response = "Y",  
  rep = NULL,  
  Ncomp = 2,  
  type = "AMMI",  
  colGen = "gray47",  
  colEnv = "darkred",  
  sizeGen = 4,  
  sizeEnv = 4,  
  titles = TRUE,  
  footnote = TRUE,  
  axis_expand = 1.2,  
  limits = TRUE,  
  axes = TRUE,  
  axislabels = TRUE  
)
```

**Arguments**

Data	a dataframe with genotypes, environments, repetitions (if any) and the phenotypic trait of interest. There is no restriction on the order in which these variables should be presented in the dataframe, and also other variables that will not be used in the analysis can be included.
genotype	column name containing genotypes.
environment	column name containing environments.
response	column name containing the phenotypic trait of interest.
rep	column name containing replications. If this argument is 'NULL' (default), replications are not considered for the analysis.
Ncomp	number of principal components that will be used in the analysis.
type	method for fitting the AMMI model: "AMMI", "rAMMI", "hAMMI", "gAMMI", "lAMMI" or "ppAMMI" (see References). Defaults to "AMMI".
colGen	genotype attributes colour. Defaults to "gray".
colEnv	environment attributes colour. Defaults to "darkred".
sizeGen	genotype labels text size. Defaults to 4.
sizeEnv	environment labels text size. Defaults to 4.
titles	logical, if this argument is 'TRUE' a plot title is generated. Defaults to 'TRUE'.
footnote	logical, if this argument is 'TRUE' a footnote is generated. Defaults to 'TRUE'.
axis_expand	multiplication factor to expand the axis limits by to enable fitting of labels. Defaults to 1.2.
limits	logical. If 'TRUE' axes are automatically rescaled. Defaults to 'TRUE'.
axes	logical, if this argument is 'TRUE' axes passing through the origin are drawn. Defaults to 'TRUE'.
axislabels	logical, if this argument is 'TRUE' labels axes are included. Defaults to 'TRUE'.

**Details**

Comentar cuál es cada método.

**Value**

A biplot of class ggplot

**References**

Rodrigues P.C., Monteiro A., Lourenco V.M. (2015). *A robust AMMI model for the analysis of genotype-by-environment data*. *Bioinformatics* 32, 58–66.

**Examples**

```
library(geneticae)
# Data without replication
library(agridat)
data(yan.winterwheat)
BIP_AMMI <- rAMMI(yan.winterwheat, genotype = "gen", environment = "env",
                 response = "yield", type = "AMMI")
BIP_AMMI

# Data with replication
data(plrv)
BIP_AMMI2 <- rAMMI(plrv, genotype = "Genotype", environment = "Locality",
                  response="Yield", rep = "Rep", type = "AMMI")
BIP_AMMI2
```

# Index

- \* **Biplot**
  - GGEPlot, 4
- \* **GGE**
  - GGEPlot, 4
- \* **datasets**
  - plrv, 9
  
- GGEModel, 2
- GGEmodel, 2, 4
- GGEPlot, 4, 4
  
- imputation, 6
  
- plrv, 9
  
- rAMMI, 10