Package 'geneticae'

July 22, 2025

Title Statistical Tools for the Analysis of Multi Environment
Agronomic Trials

Version 0.4.0

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>). 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>), and also imputation techniques
specifically developed for this kind of data ('Arciniegas-Alarcón' 2014,
<doi: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, missMDA, calibrate, graphics, reshape2,
matrixStats, tidyr, prettydoc, rlang

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

Maintainer Julia Angelini <jangelini_93@hotmail.com>

Repository CRAN

Date/Publication 2022-07-20 15:40:06 UTC

Contents

GGEmodel   *Site Regression model*

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. There are different functions in R to fit the SREG model, such as the GGEModel from the GGEBiplots package. However, this function has the following improvements:

- Includes recently published robust versions of the SREG model (Angelini et al., 2022).

- It can be used for data from trials with repetitions (there is no need to calculate means before-hand).

- Other variables not used in the analysis can be present in the dataset.

Usage

```
GGEmodel(
  Data,
  genotype = "gen",
  environment = "env",
  response = "yield",
  rep = NULL,
  model = "SREG",
  SVP = "symmetrical"
)
```

Arguments

`Data`	dataframe with genotypes, environments, repetitions (if any) and the phenotypic trait of interest. 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.
`model`	method for fitting the SREG model: '"SREG"','"CovSREG"','"hSREG"' or '"ppSREG"' (see References). Defaults to '"SREG"'.
`SVP`	method for singular value partitioning. Either '"row"', '"column"', or '"symmetrical"'. Defaults to '"symmetrical"'.

Details

A linear model by robust regression using an M estimator proposed by Huber (1964, 1973) fitted by iterated re-weighted least squares, in combination with three robust SVD/PCA procedures, resulted in a total of three robust SREG alternatives. The robust SVD/PCA considered were:

- CovSREG: robust PCA that is obtained by replacing the classical estimates of location and covariance by their robust analogues using Minimum Regularized Covariance Determinant (MRCD) approach;

- hSREG: robust PCA method that tries to combine the advantages of both approaches, PCA based on a robust covariance matrix and based on projection pursuit;

- ppSREG: robust PCA that uses the projection pursuit and directly calculates the robust estimates of the eigenvalues and eigenvectors without going through robust covariance estimation. It is a very attractive method for bigdata situations, which are very common in METs (a few genotypes tested in a large number of environments), as the principal components can be calculated sequentially.

Value

A list of class `GGE_Model` containing:

`model`	SREG model version.
`coordgenotype`	plotting coordinates for each genotype in every component.
`coordenviroment`	
	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.
SVP	name of SVP method.

A biplot of class `ggplot`

References

Julia Angelini, Gabriela Faviere, Eugenia Bortolotto, Gerardo Domingo Lucio Cervigni & Marta Beatriz Quaglino (2022) Handling outliers in multi-environment trial data analysis: in the direction of robust SREG model, Journal of Crop Improvement, DOI: 10.1080/15427528.2022.2051217

Examples

```
library(geneticae)

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

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

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 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.

- "Biplot": Basic biplot.
- "Selected Environmen"t: Ranking of cultivars based on their performance in any given environment.
- "Selected Genotype": Ranking of environments based on the performance of any given cultivar.
- "Relationship Among Environments".
- "Comparison of Genotype".
- "Which Won Where/What": Identifying the 'best' cultivar in each environment.
- "Discrimination vs. representativeness": Evaluating the environments based on both discriminating ability and representativeness.
- "Ranking Environments": Ranking environments with respect to the ideal environment.
- "Mean vs. stability": Evaluating cultivars based on both average yield and stability.
- "Ranking Genotypes": Ranking genotypes with respect to the ideal genotype.

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)
GGEPlot(GGE1)

# Data with replication
data(plrv)
GGE2 <- GGEmodel(plrv, genotype = "Genotype", environment = "Locality",
                 response = "Yield", rep = "Rep")
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,
  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. 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", "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.
change.factor	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.
simplified.model	
	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, simplified.model=FALSE). 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 (simplified.model=TRUE). 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.
scale	boolean. By default TRUE leading to a same weight for each variable
method	"Regularized" by default or "EM"
row.w	row weights (by default, a vector of 1 for uniform row weights)
coeff.ridge	1 by default to perform the regularized imputePCA algorithm; useful only if method="Regularized". 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
seed	integer, by default seed = NULL implies that missing values are initially imputed by the mean of each variable. Other values leads to a random initialization
nb.init	integer corresponding to the number of random initializations; the first initialization is the initialization with the mean imputation
Winf	peso inferior
Wsup	peso superior

Details

Often, multi-environment experiments are unbalanced because several genotypes are not tested in some environments. Several methodologies have been proposed in order to solve this lack of balance caused by missing values, some of which are included in this function:

- EM-AMMI: an iterative scheme built round the above procedure is used to obtain AMMI imputations from the EM algorithm. The additive parameters are initially set by computing the grand mean, genotype means and environment means obtained from the observed data. The residuals for the observed cells are initialized as the cell mean minus the genotype mean minus the environment mean plus the grand mean, and interactions for the missing positions are initially set to zero. The initial multiplicative parameters are obtained from the SVD of this matrix of residuals, and the missing values are filled by the appropriate AMMI estimates. In subsequent iterations, the usual AMMI procedure is applied to the completed matrix and the missing values are updated by the corresponding AMMI estimates. The arguments used for this method are:initial.values, precision, maxiter, change.factor and simplified.model

- Gabriel: combines regression and lower-rank approximation using SVD. This method initially replaces the missing cells by arbitrary values, and subsequently the imputations are refined through an iterative scheme that defines a partition of the matrix for each missing value in turn and uses a linear regression of columns (or rows) to obtain the new imputation. The arguments used for this method is only the dataframe.

- WGabriel: is a a modification of Gabriel method that uses weights chosen by cross-validation. The arguments used for this method are Winf and Wsup.

- EM-PCA: impute the missing entries of a mixed data using the iterative PCA algorithm. The algorithm first consists imputing missing values with initial values. The second step of the iterative PCA algorithm is to perform PCA on the completed dataset to estimate the parameters. Then, it imputes the missing values with the reconstruction formulae of order nPC (the fitted matrix computed with nPC components for the scores and loadings). These steps of estimation of the parameters via PCA and imputation of the missing values using the fitted matrix are iterate until convergence. The arguments used for this methods are: nPC, scale, method, row.w, coeff.ridge, precision, seed, nb.init and maxiter

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.

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-byenvironment 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-AMMI")
```

---

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. 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

To overcome the problem of data contamination with outlying observations, Rodrigues, Monteiro and Lourenco (2015) propose a robust AMMI model based on the M-Huber estimator and in robusts SVD/PCA procedures. Several SVD/PC methods were considered, briefly described below, thus conveying a total of five robust AMMI candidate models:

- R-AMMI: uses the L1 norm instead of the more usual least squares L2 norm, to compute a robust approximation to the SVD of a rectangular matrix.

- H-AMMI: Combines projection-pursuit and robust covariance estimation techniques to compute the robust loadings. It is most adequate for high-dimensional data.

- G-AMMI: Uses projection-pursuit to compute PCA estimators. The optimization is done via the grid search algorithm in the plane instead of the p-dimensional space.

- L-AMMI: The idea behind this approach is to perform classical PCA on the data but projected onto a unit sphere. When the data are elliptically distributed the estimates of the eigenvectors are consistent

- PP-AMMI: Uses projection-pursuit calculating the robust eigenvalues and eigenvectors without going through robust covariance estimation. The principal components can be sequentially computed and thus this method is very appealing when few genotypes are evaluated under a wide range of environmental and/or experimental conditions.

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