Package 'geneticae'

April 16, 2026

Title Statistical Tools for the Analysis of Multi Environment Agronomic Trials

Version 1.0.0

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 7.3.3

Imports stats, ggforce, ggplot2, scales, MASS, pcaMethods, rrcov, dplyr, missMDA, tidyr, rlang, corpcor

Suggests agridat, spelling, knitr, rmarkdown, patchwork, testthat

VignetteBuilder knitr

Depends R (>= 2.12.0)

URL https://jangelini.github.io/geneticae/, https://github.com/jangelini/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 2026-04-16 17:50:02 UTC

Contents

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

Description

Missing values are not allowed by the AMMI, GGE or SREG 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", "EM-GGE", "EM-SREG", "EM-bSREG", "Gabriel", "Eigenvector", "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.
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 (standard EM). Smaller values can help convergence if changes are cyclic.
simplified.model	logical. If TRUE, calculates effects only in the first iteration to speed up convergence or help in cases where the regular procedure fails.
scale	boolean. By default TRUE for "EM-PCA".
method	"Regularized" or "EM" for "EM-PCA".
row.w	row weights for "EM-PCA".
coeff.ridge	ridge coefficient for "EM-PCA".
seed	integer for random initialization in "EM-PCA".
nb.init	number of random initializations for "EM-PCA".
Winf	lower weight for WGabriel.
Wsup	upper weight for WGabriel.

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

- EM-GGE: Iterative SVD-based imputation focusing on G+GE.
- EM-SREG: Iterative algorithm using the Sites Regression model. Supports variants like standard SVD and Bayesian PCA (EM-bSREG).
- 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

A matrix of the imputed data.

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.

Yan, W. (2013). *Biplot analysis of incomplete two-way data*. Crop Science, 53(1), 48-57. doi:10.2135/cropsci2012.05.0301

Arciniegas-Alarcón, S., García-Peña, M., Krzanowski, W., & Dias, C. T. S. (2014b). *An alternative methodology for imputing missing data in trials with genotype-by-environment interaction: some new aspects*. Biometrical Letters, 51(2), 75-88. doi:10.2478/bile20140006

Angelini, J., Cervigni, G. D. L., & Quaglino, M. B. (2024). *New imputation methodologies for genotype-by-environment data: an extensive study of properties of estimators*. Euphytica, 220(6), 92. doi:10.1007/s1068102403344z

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)
```

`rAMMIModel` *Robust AMMI Model*

Description

Fits a classical or robust Additive Main effects and Multiplicative Interaction (AMMI) model.

Usage

```
rAMMIModel(
  Data,
  genotype = "gen",
  environment = "env",
  response = "Y",
  rep = NULL,
  Ncomp = 2,
  type = "AMMI"
)
```

Arguments

`Data`	a dataframe with genotypes, environments and the phenotypic trait.
`genotype`	column name containing genotypes.
`environment`	column name containing environments.
`response`	column name containing the phenotypic trait.
`rep`	column name containing replications. If provided, means are calculated.
`Ncomp`	number of principal components to retain.
`type`	method for fitting: ‘"AMMI"‘, ‘"rAMMI"‘, ‘"hAMMI"‘, ‘"gAMMI"‘, ‘"lAMMI"‘ or ‘"ppAMMI"‘.

`rAMMIPlot` *AMMI Biplots with ggplot2*

Description

Produces a biplot for objects of class 'AMMI'.

Usage

```
rAMMIPlot(
  model_res,
  colGen = "gray47",
  colEnv = "darkred",
  sizeGen = 6,
  sizeEnv = 6,
  titles = TRUE,
  footnote = TRUE,
  axis_expand = 1.2,
  limits = TRUE,
  axes = TRUE,
  axislabels = TRUE
)
```

Arguments

`model_res`	an object of class 'AMMI' from AMMIModel.
`colGen`	genotype colour. Defaults to "gray47".
`colEnv`	environment colour. Defaults to "darkred".
`sizeGen`	genotype text size.
`sizeEnv`	environment text size.
`titles`	logical, show plot title.
`footnote`	logical, show footnote with explained variance.
`axis_expand`	expansion factor for axis limits.
`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'

`rSREGModel` *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, 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 beforehand).
- Other variables not used in the analysis can be present in the dataset.

Usage

```
rSREGModel(
  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 <- rSREGModel(yan.winterwheat, genotype="gen", environment="env", response="yield")

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

---

rSREGPlot *GGE biplots with* **ggplot2**

Description

GGE biplots are used for visual examination of the relationships between test environments, genotypes, and genotype-by-environment interactions. 'rSREGPlot()' produces a biplot as an object of class 'ggplot', using the output of the rSREGModel 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 the 'rSREGPlot' function from the GGEBiplots package.

Usage

```
rSREGPlot(
  rSREGModel,
  type = "Biplot",
  d1 = 1,
  d2 = 2,
  selectedE = NA,
  selectedG = NA,
  selectedG1 = NA,
  selectedG2 = NA,
  colGen = "gray47",
  colEnv = "darkred",
  colSegment = "gray30",
  colHull = "gray30",
  sizeGen = 6,
  sizeEnv = 6,
  largeSize = 4.5,
  axis_expand = 1.2,
  axislabels = TRUE,
  axes = TRUE,
  limits = TRUE,
```

```
  titles = TRUE,
  footnote = TRUE
)
```

Arguments

rSREGModel	An object of class rSREGModel.
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 <- rSREGModel(yan.winterwheat)
rSREGPlot(GGE1)

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

Index