Package ‘MATTOOLS’

February 19, 2015

Type Package

Title Modern Calibration Functions for the Modern Analog Technique (MAT)

Version 1.1

Date 2012-07-23

Author Dr. M. Sawada.

Maintainer steven mosher <`moshersteven@gmail.com`>

Depends R (>= 2.15.0)

Description This package in includes functions for receiver operating characteristic (ROC) analyses as well as Monte Carlo simulation. It includes specific graphical functions for interpreting the output of these techniques.

License GPL (>= 2)

LazyLoad yes

LazyData TRUE

LazyDataCompression bzip2

Repository CRAN

Date/Publication 2012-08-19 05:21:39

NeedsCompilation no

R topics documented:

`mattools-package` *Modern Calibration Functions for the Modern Analog Technique*

Description

This package in includes functions for receiver operating characteristic (ROC) analyses as well as Monte Carlo simulation. It includes specific graphical functions for interpreting the output of these techniques.

Details

Package:	mattools
Type:	Package
Version:	1.1
Date:	2012-07-23
License:	GPL 2.0

Author(s)

Dr. M. Sawada.

Maintainer: steven mosher <moshersteven@gmail.com>

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

`euclidean.compx.f` *Decomposition of direction to x-vector component*

Description

Internal function used by mat.dissim.

Usage

```
euclidean.compx.f(modsite, fossite)
```

Arguments

`modsite`	Vector: containing the x and y Euclidean values for a point.
`fossite`	Vector: containing the x and y Euclidean values for another point.

Value

A numeric value representing the x-vector component decomposition.

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

See Also as `euclidean.compy.f`, `euclidean.direction.f`,`euclidean.distance.f`,`mat.dissim`,`great.circle.distanc`

`euclidean.compy.f` *Decomposition of direction to y-vector component*

Description

Internal function used by mat.dissim.

Usage

```
euclidean.compy.f(modsite, fossite)
```

Arguments

`modsite`	Vector: containing the x and y Euclidean values for a point.
`fossite`	Vector: containing the x and y Euclidean values for another point.

Value

A numeric value representing the x-vector component decomposition.

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

See Also as `euclidean.compx.f`, `euclidean.direction.f`,`euclidean.distance.f`,`mat.dissim`,`great.circle.distan`

`euclidean.direction.f` *Euclidean direction on a plane*

Description

This function is ancillary to mat.dissim and gives the direction from point A to B on a plane.

Usage

```
euclidean.direction.f(modsite, fossite)
```

Arguments

`modsite`	Vector: containing the x and y Euclidean values for a point.
`fossite`	Vector: containing the x and y Euclidean values for another point.

Details

Provides the quadrat correct direction from fossite to modsite.

Value

A numeric single value representing the direction between the two points.

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

See Also as `euclidean.compx.f`, `euclidean.compy.f`,`euclidean.distance.f`,`mat.dissim`,`great.circle.distance.f`

`euclidean.distance.f` *Euclidean distance function*

Description

This function is ancillary to mat.dissim and gives the distance from point A to B on a plane in the units of x and y.

Usage

```
euclidean.distance.f( fossite,modsite)
```

Arguments

`modsite`	Vector: containing the x and y Euclidean values for a point.
`fossite`	Vector: containing the x and y Euclidean values for another point.

Details

Provides the quadrat correct direction from fossite to modsite.

Value

A numeric single value representing the direction between the two points.

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

See Also as `euclidean.compx.f`, `euclidean.compy.f`,`euclidean.direction.f`,`mat.dissim`,`great.circle.distance.`

`great.circle.distance.f`

Computes great circle distance

Description

This is an ancillary function to mat.dissim that calculates distance in meters on a sphere between two points of longitude/latitude

Usage

```
great.circle.distance.f(cLonLat, oLonLat)
```

Arguments

`cLonLat`	Vector: with two values, one of longitude and the other latitude.
`oLonLat`	Vector: with two values, one of longitude and the other latitude.

Details

The calculation is based on a sphere of 6374 km diamater. The user can change this for other spheres or earth models.

Value

A single value of distance in meters.

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

See Also as `euclidean.compx.f`, `euclidean.compy.f`,`euclidean.direction.f`,`mat.dissim`,`euclidean.distance.f`

`mat.dissim` *Dissimilarity object for MAT reconstructions*

Description

Creates the dissimilarity object between either a modern dataset and itself or a modern and fossil dataset. This list object contains matrices and parameters that are used in both optimizations of critical choices as well as environmental reconstructions.

Usage

```
mat.dissim(inFossil, inModern, llMod = c(), modTaxa = c(), llFoss = c(), fosTaxa = c(), numAnalogs = 1
```

Arguments

`inFossil`	Dataframe: Fossil (or Modern) Calibration Dataset: a file containing field names in the first row of the fossil/modern calibration dataset where each subsequent row containing a site/row identifier (Sample ID), coordinates in either a planar/projected x,y system or as Longitude and Latitude in decimal degrees, dd, and taxon counts followed by the modern environmental variables (Mod.Env 1,Mod.Env n) that will be used for modern training and/or paleoenvironmental reconstruction. The final and optional field would contain, for each row, a nominal code representing the biological zone to which each row/site belongs.
`inModern`	Dataframe: Modern Calibration Dataset: a file containing field names in the first row of the modern calibration dataset where each subsequent row containing a site/row identifier (Sample ID), coordinates in either a planar/projected x,y system or as Longitude and Latitude in decimal degrees, dd, and taxon counts followed by the modern environmental variables (Mod.Env 1,Mod.Env n) that will be used for modern training and/or paleoenvironmental reconstruction. The final and optional field would contain, for each row, a nominal code representing the biological zone to which each row/site belongs.
`llMod`	Vector: with two items, the number of the column containing the x-coordinate for each sample and the number of the column containing the y-coordinate for each sample within the inModern dataset.
`modTaxa`	Vector: with two items, the number of the column containing the first taxon for analysis for each sample and the number of the column containing the last taxon for each sample within the inModern dataset.
`llFoss`	Vector: see argument llMod above for explanation but with reference to the inFossil argument
`fosTaxa`	Vector: see argument modTaxa above for explanation but with reference to the inFossil argument
`numAnalogs`	Numeric: a single number > 1 that specifies the number of modern analogs to use in the reconstruction.
`counts`	Logical: True (default) then the program assumes that your inFossil AND inModern datasets are taxon counts and so will automatically convert them to proportions.

`sitenames`	The column with the sample site names within the inFossil dataset.
`dist.method`	the distance method used

Details

When count = True, the default, the program will automatically attempt to convert the taxon counts to proportions. If there are all zero counts for one or more rows in the inModern or inFossil datasets the program will terminate. It is a good idea to use rowSums() function on the columns of the datasets before using mat.dissim to ensure that there are no zero-count sample sites.

Value

LIST with the following components:

`x`	Vector: x-coordinates of fossil sample sites.
`y`	Vector: y-coordinates of fossil sample sites.
`sqdist`	Matrix: with n rows, each containing m dissimilarity values in m columns. The first row is the best analog for each inFossil sample, the second row is the 2nd best analog, etc...The number of rows n is equal to the number of analogs specified by the argument numAnalogs. The number of columns, m, are equal to the number of sample sites (rows) of the inFossil argument.
`position`	Matrix: with n rows, each containing m values in m columns. Each value specifies the row number within the inModern dataframe argument that is the sample site's analog. The number of rows n is equal to the number of analogs specified by the argument numAnalogs. The number of columns, m, are equal to the number of sample sites (rows) of the inFossil argument. The order is the same as the above sqdist component.
`distance`	Matrix: with n rows, each containing m geographic distance values in m columns. The number of rows n is equal to the number of analogs specified by the argument numAnalogs. The number of columns, m, are equal to the number of sample sites (rows) of the inFossil argument. The order is the same as the above sqdist component.
`direction`	Matrix: with n rows, each containing m azimuth's to the corresponding analog in m columns. The number of rows n is equal to the number of analogs specified by the argument numAnalogs. The number of columns, m, are equal to the number of sample sites (rows) of the inFossil argument. The order is the same as the above sqdist component.
`xcomponent`	Matrix: with n rows, each containing m x component distances for creation of a vector with the ycomponent component in m columns. The number of rows n is equal to the number of analogs specified by the argument numAnalogs. The number of columns, m, are equal to the number of sample sites (rows) of the inFossil argument. The order is the same as the above sqdist component.
`ycomponent`	Matrix: with n rows, each containing m y component distances for creation of a vector with the xcomponent component in m columns. The number of rows n is equal to the number of analogs specified by the argument numAnalogs. The number of columns, m, are equal to the number of sample sites (rows) of the inFossil argument. The order is the same as the above sqdist component.

`inModern`	Name of the inModern dataset used.
`inFossil`	Name of the inFossil dataset used.
`llmod`	The vector specifying the value assigned to the argument llMod.
`modTaxa`	The vector specifying the value assigned to the argument modTaxa.
`counts`	The value of the counts argument.

Note

Utilizes MATTOOLS functions: "great.circle.distance.f","spherical.direction.f","euclidean.distance.f","euclidean.direction.f

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

See Also as `euclidean.compx.f`, `euclidean.compy.f`,`euclidean.direction.f`,`great.circle.distance.f`,`euclidean`

`mat.fossavg` *Fossil sample site environmental reconstruction*

Description

After the mat.dissim object is created this function can be used to provide the environmental reconstruction for each fossil site.

Usage

```
mat.fossavg(dObj, modEnvCol, fossEnvCol = 0, fossCols = 0, cutoff = 0, distance = -1, wmethod = "none"
```

Arguments

`dObj`	List: output from application of mat.dissim
`modEnvCol`	Numeric: the column of the modern dataset (inModern in mat.dissim) that contains the environmental variable of interest.
`fossEnvCol`	Numeric: Not currently implemented. The column of the fossil dataset (inFossil in mat.dissim) that contains the environmental variable of interest.
`fossCols`	Vector: the column in the fossil dataset that contains the modern environmental variable measured at the fossil sites.

`cutoff`	Numeric: the critical value determined either through application of mat.mc or mat.roc or any arbritrary value
`distance`	Numeric: the maximum distance a modern sample site can be from each fossil site and still be considered as a potential analog.
`wmethod`	Character: either "equal.wt" (default) or "inv.dissim", "inv.dist","inv.rank", "single" or "none".
`numanalogs`	Numeric: the number of analogs to use in the reconstruction.

Details

The argument numanalogs must be less than or equal to the number of analogs chosen in the application of mat.dissim. If this is not the case then an error will be returned. This function computes the weighted absolute deviation from the weighted mean of data. The absolute deviation from the mean is defined for each sample as as the sum of the individual weights multiplied by the absolute deviation of each value from the mean divided by the sum of the weights. With a cutoff value, say 0.15, all dissimilarity values strictly greater than this critical threshold are masked out. Standard Deviations (weighted/unweighted) are only reported if there are greater than 1 analogs retained after the application of a critical threshold of dissimilarity. The argument wmethod will determine the size of the output. With "none" there will be as many columns output as there are number of analogs, that is, for each analog the modern environmental variable is reconstructed and each is output in the output table. The argument fossEnvCol is not currently implemnted in this version but is there as a placeholder for inclusion of potential anomaly calculations.

Value

A dataframe containing the fossil sites and the reconstructed variable(s) (a function of weighting.

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

See Also `mat.dissim`

`mat.jump` *Optimize the alpha value*

Description

The MATTOOLS function mat.jump optimizes the value of for a given number of analogs and modern dataset.

Usage

```
mat.jump(dObj, inModern, envColumn, cutoff = 0, calib = T)
```

Arguments

`dObj`	List: A Dissimilarity Object from the mat.dissim function.
`inModern`	Dataframe: The modern dataset with column(s) representing environmental values at each modern sample site.
`envColumn`	Numeric: The number of the column with the environmental values to be reconstructed.
`cutoff`	Numeric: The critical value of dissimilarity based on a priori knowledge or determined via mat.roc or mat.mc functions
`calib`	Logical: T for modern calibration (default). If F then it is assumed that a set of fossil assemblages are being tested.

Details

With a cutoff > 0, only pairwise complete observations are used.

Value

A LIST with the following components:

`alpha`	Numeric: optimal alpha value.
`alphacor`	Datatable: two columns, alpha and correlation.

...

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

`mat.jumprecon`, `mat.plotjump`,`mat.dissim`,

`mat.jumprecon` *Modified Jump Method (MJM) Application*

Description

Once the dissimilarity object, consisting of the n closest dissimilarity scores for each fossil sample, has been created, then either the MJM method can be applied to the reconstruction using the function mat.jumprecon

Usage

```
mat.jumprecon(dObj, modEnvCol, fossEnvCol, cutoff = 0, alpha = NULL, distance = -1, numanalogs = lengt
```

Arguments

`dObj`	List: any object created with mat.dissim
`modEnvCol`	Numeric: the column within the modern dataset that contains the environmental value undergoing reconstruction.
`fossEnvCol`	Numeric: the column within the fossil dataset that contains the environmental value undergoing reconstruction
`cutoff`	Numeric: the critical value of dissimilarity to be applied.
`alpha`	Numeric: the value of alpha determined by application of mat.jump
`distance`	Numeric: a geographic distance threshold (in meters) to be applied in the further exclusion of sites.
`numanalogs`	Numeric: the number of analogs to retain in the MJM.

Value

A datatable containing the results of the application of the MJM. Some values may be NA when they fail to meet the criterion.

Author(s)

Dr. M. Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

`mat.plotjump`, `mat.jump`,`mat.dissim`,

`mat.mc` *Monte Carlo Simulation for Determining Critical Threasholds of Dissimilarity*

Description

Undertakes a Monte-Carlo simulation of pairwise modern spectra to critical thresholds of dissimilarity and associated probabilities.

Usage

```
mat.mc(inModern, modTaxa = c(NULL, NULL), probs = c(0.05, 0.025, 0.01, 0.001), freqint = seq(0, 2, 0.0
```

Arguments

`inModern`	Modern Calibration Dataset: a file containing field names in the first row of the modern calibration dataset where each subsequent row containing a site/row identifier (Site ID), coordinates in either a planar/projected x,y system or as Longitude and Latitude in decimal degrees, and taxon counts followed by the modern environmental variables (Mod.Env 1,Mod.Env n) that will be used for modern training and/or paleoenvironmental reconstruction. The final and optional field would contain, for each row, a nominal code representing the biological zone to which each row/site belongs.
`modTaxa`	A vector containing the column numbers containing the modern taxon counts or proportions within inModern
`probs`	A vector of significance levels of 0.1, 0.05 etc. at which to return the dissimilarity values. The valid range is between 0 and 1 for this vector.
`freqint`	A vector or sequence that spans the range of the dissimiliarity value being analyzed, for example, for squared-chord distance which can range between 0 and 2, a valid sequence would be seq(0,2,0.02).
`sampleSize`	A single number that determins how many samples are compared for the Monte-Carlo simulation. Defaults to the number of samples in the input inModern dataset.
`method`	Either "sawada" or "allpairs" are implemented; "sawada" is the default method of comparison whereby pairwise comparisons are taken randomly with replacement from the inModern dataset sampleSize times. Alternatively, for all pairwise comparisons use "allpairs" as the argument value.
`withReplace`	A logical value specific to the "sawada" comparison method that defaults to TRUE and allows for random pairwise comparisons with replacement from the inModern dataset.
`counts`	A logical value that describes whether the taxon datset of inModern is in the form of raw counts or proportions. If this artument is FALSE then the comparison method assumes the dataset is proportions.

Details

The default arguments to mat.mc are set for working with the squared-chord distance SCD dissimilarity measure and if another dissimilarity measure is chosen the user will want to change these defaults, particularly the evaluation sequence in the argument freqint. When choosing a comparison method, either Monte-Carlo with the default method "sawada" or all pairwise comparisons with the method "bartlein", consider that this latter method is computationally more intensive and for large datasets may fail due to memory limitations. This method is primarily intended for smaller datasets of a few hundred samples at most.

Value

A LIST with the following components:

`sqdist`	a vector of dissimilarity values of sampleSize
`cumcurve`	a vector of cumulative frequencies at the sequence defined by the freqint argument
`cutoffs`	a list object with associated Monte Carlo probabilities in the component cutoffs under cutoffs$x which are from the probs argument and cutoffs$y which contain the associated dissimilarity values for each probability.
`method`	a string describing the method chosen
`samplesize`	the sample size from the argument sampleSize
`replacement`	the value from the argument withReplace
`probabilities`	the value probabilities from the argument probs
`wascounts`	the value of the argument count

Author(s)

Dr. M. Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

`mat.plot.mc`,`mat.dissim`

`mat.plot.mc` *Plots output from mat.mc*

Description

Plots three panels A) the distribution of dissimilarity values from Monte Carlo pairwise simulation from the output of the function mat.mc; B) Cumulative distribution function across range of dissimilarity with blue indicating the user specified tail of significance; C) Zoom view of the user specified tail of significance with corresponding critical thresholds indicated by intersecting colored lines with the numbers on these indicating the threshold (x-axis) and corresponding Monte-Carlo p-value on the y-axis.

Usage

```
mat.plot.mc(mcObj)
```

Arguments

`mcObj`	LIST: output from application of mat.mc

Value

A graphic device with plots.

Author(s)

Dr. M. Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

`mat.mc`,

`mat.plot.recon` *Plot output from mat.fossavg*

Description

Plots the output from the function mat.fossavg.

Usage

```
mat.plot.recon(inrec, colAgDp = 7, inCritVal)
```

Arguments

`inrec`	List: from application of mat.fossavg
`colAgDp`	Numeric: the column with the age or depth for each level
`inCritVal`	List: object must come from application of mat.mc

Details

This function is intended as an example plotting function for output of MATTOOLS analyses. Users are encouraged to study this function and modify it as they please.

Value

A graphics device with two graphs showing the reconstructed variable and the dissimilarity value with probabilities associated with non-analogs.

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

`mat.fossavg`,

`mat.plotjump` *Plot Jump Analysis*

Description

Plots the result of mat.jump for visualization of optimal alpha value.

Usage

```
mat.plotjump(inJumpObj)
```

Arguments

`inJumpObj`	List: any output from applying mat.jump function.

Details

This is another simple plotting function useful for study and modification by users.

Value

A graphical device with a plot.

Author(s)

Professor Michael Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

`mat.jumprecon`, `mat.jump`,`mat.dissim`,

`mat.plotroc` *Plot output of ROC analysis*

Description

The output of mat.roc can be plotted using the function mat.plotroc. Plots output from either mat.roc or mat.roc.allpair.

Usage

```
mat.plotroc(inRocObj)
```

Arguments

`inRocObj`	List: any object created by the function mat.roc

Value

Numerous graphic devices showing output of zonal comparisons in the ROC application.

Author(s)

Dr. M. Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

`mat.roc`, `mat.roc.allpair`,`mat.mc`,

`mat.roc` *ROC analysis for the Modern Analog Technique (MAT)*

Description

Undertakes Receiver Operating Curve (ROC) analysis on a modern dataset when each row/site has a nominal class defining biological zone membership. This function undertakes ROC on a within-zone by out-zone basis.

Usage

```
mat.roc(inModern, modTaxa = c(), colClasses = NULL, numAnalogs = 2, rocEvalSeq = seq(0, 2, 0.05), coun
```

Arguments

`inModern`	Dataframe: Modern Calibration Dataset: a file containing field names in the first row of the modern calibration dataset where each subsequent row containing a site/row identifier (Sample ID), coordinates in either a planar/projected x,y system or as Longitude and Latitude in decimal degrees, dd, and taxon counts followed by the modern environmental variables (Mod.Env 1,Mod.Env n) that will be used for modern training and/or paleoenvironmental reconstruction. The final and optional field would contain, for each row, a nominal code representing the biological zone to which each row/site belongs.
`modTaxa`	Vector: With two items, the number of the column containing the first taxon for analysis for each sample and the number of the column containing the last taxon for each sample within the inModern dataset.
`colClasses`	Numeric: The column of inModern that has the zonal affiliations for each sample.
`numAnalogs`	Numeric: a single number > 1 that specifies the number of modern analogs to use in the reconstruction.
`rocEvalSeq`	Vector: a numeric vector specifying the sequence over which the ROC analyses will be done. Generally, this sequence will range from the minimum value of the dissimilarity index to the maximum value of the dissimilarity index divided by some interval. For example, the default is set for squared-chord distance as rocEvalSeq=seq(0,2,0.05). This should be changed for other implementations of dissimilarity metrics.
`counts`	Logical: True (default) then the program assumes that your inFossil AND inModern datasets are taxon counts and so will automatically convert them to proportions.
`aucmethod`	Character: either "trap" for the trapezoidal integration or "wilcox" for the Mann-Whitney-Wilcoxon statistic.

Details

This function uses helper functions called mattools.roc, mat.ROCcalc and polyarea. The standard error of the AUC, SEAUC, can be calculated from the nonparametric Mann-Whitney-Wilcoxon rank-sum statistic or alternatively by the method of Hanley and McNeil (1982) which is computationally faster for large datasets and is the default method for calculating SEAUC in MATTOOLS.

Value

A LIST with the following components:

`Zone1`	ROC results for first zone
`ZoneN`	ROC results for last zone
`Overall`	ROC results for all zones lumped.

Author(s)

Dr. M. Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

`mat.roc.allpair`, `mat.plotroc`,`mat.mc`,

`mat.roc.allpair` *ROC Analysis by Pairwise Comparisons*

Description

Undertakes Receiver Operating Curve (ROC) analysis on a modern dataset when each row/site has a nominal class defining biological zone membership. This function undertakes ROC on a within-zone by out-zone basis. This function undertakes pairwise comparisons of zones.

Usage

```
mat.roc.allpair(inModern, modTaxa = c(), colClasses = NULL, numAnalogs = 2, rocEvalSeq = seq(0, 2, 0.0
```

Arguments

`inModern`	Dataframe: Modern Calibration Dataset: a file containing field names in the first row of the modern calibration dataset where each subsequent row containing a site/row identifier (Sample ID), coordinates in either a planar/projected x,y system or as Longitude and Latitude in decimal degrees, dd, and taxon counts followed by the modern environmental variables (Mod.Env 1,Mod.Env n) that will be used for modern training and/or paleoenvironmental reconstruction. The final and optional field would contain, for each row, a nominal code representing the biological zone to which each row/site belongs.
`modTaxa`	Vector: With two items, the number of the column containing the first taxon for analysis for each sample and the number of the column containing the last taxon for each sample within the inModern dataset.
`colClasses`	Numeric: The column of inModern that has the zonal affiliations for each sample.
`numAnalogs`	Numeric: a single number > 1 that specifies the number of modern analogs to use in the reconstruction.
`rocEvalSeq`	Vector: a numeric vector specifying the sequence over which the ROC analyses will be done. Generally, this sequence will range from the minimum value of the dissimilarity index to the maximum value of the dissimilarity index divided by some interval. For example, the default is set for squared-chord distance as rocEvalSeq=seq(0,2,0.05). This should be changed for other implementations of dissimilarity metrics.

`counts`	Logical: True (default) then the program assumes that your inFossil AND inModern datasets are taxon counts and so will automatically convert them to proportions.
`aucmethod`	Character: either "trap" for the trapezoidal integration or "wilcox" for the Mann-Whitney-Wilcoxon statistic.

Value

A LIST with the following components:

`Zone1`	ROC results for first zone
`ZoneN`	ROC results for last zone
`Overall`	ROC results for all zones lumped.

Author(s)

Dr. M. Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

See Also

`mat.roc`, `mat.plotroc`,`mat.mc`,

`mat.roc.up` *ROC Analysis*

Description

TBD

Usage

```
mat.roc.up(mattools.roclist, rocEvalSeq = seq(0, 2, 0.01))
```

Arguments

`mattools.roclist`	TBD
`rocEvalSeq`	TBD

Details

TBD

Value

TBD

Author(s)

Dr. M. Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

`mat.ROCcalc` *ROC calculation*

Description

Needs docs

Usage

```
mat.ROCcalc(truth, data, evalseq, method = "trap")
```

Arguments

`truth`	TBD
`data`	TBD
`evalseq`	TBD
`method`	TBD

Details

TBD

Value

TBD

Author(s)

Dr. M. Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

`mattools.roc` *tools*

Description

TBD

Usage

```
mattools.roc(inFossil, inModern, modTaxa = c(), fosTaxa = modTaxa, numAnalogs = NULL)
```

Arguments

`inFossil`	TBD
`inModern`	TBD
`modTaxa`	TBD
`fosTaxa`	TBD
`numAnalogs`	TBD

Details

TBD

Value

TBD

Author(s)

Dr. M. Sawada

References

SAWADA, M. 2006. An open-source implementation of the modern analog technique (MAT) within the R computing environment. Computers and Geosciences, Vol 32(6):818-833, DOI information: 10.1016/j.cageo.2005.10.008

`modpoll` *Sample Modern Pollen Dataset*

Description

This data is a subset of those described in Whitmore, J., K. Gajewski, M. Sawada, J.W. Williams, B. Shuman, P.J. Bartlein, T. Minckley, A.E. Viau, T. Webb III, S Shafer, P. Anderson and Brubaker, L. (2005). "Modern Pollen Data from North America and Greenland for Multi-Scale Paleoenvironmental Applications." Quaternary Science Reviews 24(16-17): 1828-1848.

Usage

```
data(modpoll)
```

Format

A data frame with 2191 observations on the following 116 variables.

`Site` a numeric vector

`SITENAME` a character vector

`LONDD` a numeric vector

`LATDD` a numeric vector

`ELEVATION` a numeric vector

`DEPTH` a numeric vector

`AGEBP` a numeric vector

`CALBP` a numeric vector

`POLLSUM` a numeric vector

`ABIES` a numeric vector

`ACER` a numeric vector

`ALNUSX` a numeric vector

`AMBROSIA` a numeric vector

`ANACARDI` a numeric vector

`APIACEAE` a numeric vector

`AQUIFOLI` a numeric vector

`ARCEUTHOBI` a numeric vector

`ARECACEAE` a numeric vector

`ARMERIA` a numeric vector

`ARTEMISIA` a numeric vector

`ASTERX` a numeric vector

`ASTERCICHO` a numeric vector

`BETULA` a numeric vector

`BORAGINACEAE` a numeric vector

`BOTRYCHIUM` a numeric vector

`BRASSICACEAE` a numeric vector

`CACTACEAE` a numeric vector

`CAMPANULACEA` a numeric vector

`CAPRIFOLIA` a numeric vector

`CARYA` a numeric vector

`CARYOPHYLL` a numeric vector

`CASTANEA` a numeric vector

`CEANOTHUS` a numeric vector

`CELTIS` a numeric vector

`CEPHALANTH` a numeric vector

`CERCOCAR` a numeric vector

`CHENOAMX` a numeric vector

`CHRYSOLEP` a numeric vector

`CORNUS` a numeric vector

`CORYLUS` a numeric vector

`CUPRESSA` a numeric vector

`CYPERACE` a numeric vector

`DODECATH` a numeric vector

`DRYAS` a numeric vector

`ELAEAGNX` a numeric vector

`EPHEDRA` a numeric vector

`EQUISETU` a numeric vector

`ERICALES` a numeric vector

`EUPHORB` a numeric vector

`FABACEAE` a numeric vector

`FAGUS` a numeric vector

`FRAXINUX` a numeric vector

`IVA` a numeric vector

`JUGLANSX` a numeric vector

`KOENISLD` a numeric vector

`LAMIACEAE` a numeric vector

`LARIXPSEU` a numeric vector

`LARREA` a numeric vector

`LILIACEAE` a numeric vector

`LIQUIDAM` a numeric vector

`LIRIODEN` a numeric vector

`LYCOANNO` a numeric vector

`LYCOCLAV` a numeric vector

`LYCOCOMP` a numeric vector

`LYCOSELA` a numeric vector

`LYCOPODX` a numeric vector

`MAGNOLIACE` a numeric vector

`MALVACEAE` a numeric vector

`MORACEAX` a numeric vector

`MYRICACX` a numeric vector

`NYSSA` a numeric vector

`ONAGRACEAE` a numeric vector

`OSMUNDACEA` a numeric vector

`OSTRYCAR` a numeric vector

`PAPAVERX` a numeric vector

`PEDICULARI` a numeric vector

`PICEAX` a numeric vector

`PINUSX` a numeric vector

`PLANTAGINX` a numeric vector

`PLATANUS` a numeric vector

`POACEAE` a numeric vector

`POLEMONI` a numeric vector

`POLYGONAX` a numeric vector

`POLYGBIS` a numeric vector

`POLYGONMX` a numeric vector

`POLYGVIV` a numeric vector

`POLYPOD` a numeric vector

`POPULUS` a numeric vector

`POTENTILLA` a numeric vector

`PROSOPIS` a numeric vector

`PTERIDIUM` a numeric vector

`QUERCUS` a numeric vector

`RANUNCUL` a numeric vector

`RHAMNACEAE` a numeric vector

`ROSACEAX` a numeric vector

`RUBIACEAE` a numeric vector

`RUBUS` a numeric vector

`RUMEX` a numeric vector

`SALIX` a numeric vector

`SANGUI` a numeric vector

`SARCOBATUS` a numeric vector

`SXFRAGAX` a numeric vector

`SCROPHUL` a numeric vector

`SELAGINE` a numeric vector

`SHEPHERDIA` a numeric vector

`SPHAGNUM` a numeric vector

`TAXODIUM` a numeric vector

`TAXUS` a numeric vector

`THALICTRUM` a numeric vector

`TILIA` a numeric vector

`TSUGAX` a numeric vector

`ULMUS` a numeric vector

`URTICACX` a numeric vector

`tjul` a numeric vector

`tave` a numeric vector

`vegzone` a factor with levels `Arctic tundra` `Central taiga` `Deserts` `Forest-steppes` `Forest-tundra` `Grasslands` `Hemixerophytic forests` `Humid forests` `Mesophytic forests` `Mountain vegetation` `Northern taiga` `Oceanic meadows` `Polar deserts` `Seasonally wet forests and wood` `Semideserts` `Southern taiga` `Southern tundra` `Subtaiga` `Typical tundra` `Wet savannas`

Details

The full dataset is available at www.lpc.uottawa.ca

Source

Whitmore, J., K. Gajewski, M. Sawada, J.W. Williams, B. Shuman, P.J. Bartlein, T. Minckley, A.E. Viau, T. Webb III, S Shafer, P. Anderson and Brubaker, L. (2005). "Modern Pollen Data from North America and Greenland for Multi-Scale Paleoenvironmental Applications." Quaternary Science Reviews 24(16-17): 1828-1848.

Examples

```
 data(modpoll)
## maybe str(modpoll) ; plot(modpoll) ...
```

`zagoskin` *Sample Fossil Pollen Dataset*

Description

This pollen dataset contains fossil pollen from Ager, T. W. (2003). "Late Quaternary Vegetation and Climate History of the Central Bering Land Bridge from St. Michael Island, Western Alaska." Quaternary Research 60: 19-32.

Usage

```
data(zagoskin)
```

Format

A data frame with 151 observations on the following 113 variables.

`Site` a numeric vector

`SITENAME` a factor with levels `Zagoskin L`

`LONDD` a numeric vector

`LATDD` a numeric vector

`ELEVATION` a numeric vector

`DEPTH` a numeric vector

`AGEBP` a numeric vector

`CALBP` a numeric vector

`POLLSUM` a numeric vector

`ABIES` a numeric vector

`ACER` a numeric vector

`ALNUSX` a numeric vector

`AMBROSIA` a numeric vector

`ANACARDI` a numeric vector

`APIACEAE` a numeric vector

`AQUIFOLI` a numeric vector

`ARCEUTHOBI` a numeric vector

`ARECACEAE` a numeric vector

`ARMERIA` a numeric vector

`ARTEMISIA` a numeric vector

`ASTERX` a numeric vector

`ASTERCICHO` a numeric vector

`BETULA` a numeric vector

`BORAGINACEAE`	a numeric vector
`BOTRYCHIUM`	a numeric vector
`BRASSICACEAE`	a numeric vector
`CACTACEAE`	a numeric vector
`CAMPANULACEA`	a numeric vector
`CAPRIFOLIA`	a numeric vector
`CARYA`	a numeric vector
`CARYOPHYLL`	a numeric vector
`CASTANEA`	a numeric vector
`CEANOTHUS`	a numeric vector
`CELTIS`	a numeric vector
`CEPHALANTH`	a numeric vector
`CERCOCAR`	a numeric vector
`CHENOAMX`	a numeric vector
`CHRYSOLEP`	a numeric vector
`CORNUS`	a numeric vector
`CORYLUS`	a numeric vector
`CUPRESSA`	a numeric vector
`CYPERACE`	a numeric vector
`DODECATH`	a numeric vector
`DRYAS`	a numeric vector
`ELAEAGNX`	a numeric vector
`EPHEDRA`	a numeric vector
`EQUISETU`	a numeric vector
`ERICALES`	a numeric vector
`EUPHORB`	a numeric vector
`FABACEAE`	a numeric vector
`FAGUS`	a numeric vector
`FRAXINUX`	a numeric vector
`IVA`	a numeric vector
`JUGLANSX`	a numeric vector
`KOENISLD`	a numeric vector
`LAMIACEAE`	a numeric vector
`LARIXPSEU`	a numeric vector
`LARREA`	a numeric vector
`LILIACEAE`	a numeric vector
`LIQUIDAM`	a numeric vector

`LIRIODEN` a numeric vector
`LYCOANNO` a numeric vector
`LYCOCLAV` a numeric vector
`LYCOCOMP` a numeric vector
`LYCOSELA` a numeric vector
`LYCOPODX` a numeric vector
`MAGNOLIACE` a numeric vector
`MALVACEAE` a numeric vector
`MORACEAX` a numeric vector
`MYRICACX` a numeric vector
`NYSSA` a numeric vector
`ONAGRACEAE` a numeric vector
`OSMUNDACEA` a numeric vector
`OSTRYCAR` a numeric vector
`PAPAVERX` a numeric vector
`PEDICULARI` a numeric vector
`PICEAX` a numeric vector
`PINUSX` a numeric vector
`PLANTAGINX` a numeric vector
`PLATANUS` a numeric vector
`POACEAE` a numeric vector
`POLEMONI` a numeric vector
`POLYGONAX` a numeric vector
`POLYGBIS` a numeric vector
`POLYGONMX` a numeric vector
`POLYGVIV` a numeric vector
`POLYPOD` a numeric vector
`POPULUS` a numeric vector
`POTENTILLA` a numeric vector
`PROSOPIS` a numeric vector
`PTERIDIUM` a numeric vector
`QUERCUS` a numeric vector
`RANUNCUL` a numeric vector
`RHAMNACEAE` a numeric vector
`ROSACEAX` a numeric vector
`RUBIACEAE` a numeric vector
`RUBUS` a numeric vector

`RUMEX` a numeric vector

`SALIX` a numeric vector

`SANGUI` a numeric vector

`SARCOBATUS` a numeric vector

`SXFRAGAX` a numeric vector

`SCROPHUL` a numeric vector

`SELAGINE` a numeric vector

`SHEPHERDIA` a numeric vector

`SPHAGNUM` a numeric vector

`TAXODIUM` a numeric vector

`TAXUS` a numeric vector

`THALICTRUM` a numeric vector

`TILIA` a numeric vector

`TSUGAX` a numeric vector

`ULMUS` a numeric vector

`URTICACX` a numeric vector

Details

The dataset was modified from that available at the NGDC Paleoclimate homepage, specifically the North American Pollen Database (NAPD) maintained by Dr. Eric Grimm of the Illinois State Museum.

Source

http://www.ncdc.noaa.gov/paleo/napd.html

Examples

```
data(zagoskin)
## maybe str(zagoskin) ; plot(zagoskin) ...
```

Index