

Package ‘codep’

January 25, 2017

Version 0.6-5

Date 2017-01-24

Type Package

Title Multiscale Codependence Analysis

Author Guillaume Guenard and Pierre Legendre, with contributions from Bertrand Pages

Maintainer Guillaume Guenard <guillaume.guenard@gmail.com>

Description Computation of Multiscale Codependence Analysis and spatial eigenvector maps, as an additional feature. Early development version.

Depends R (>= 3.0.0), grDevices, graphics, stats

Suggests vegan

License GPL-3

LazyLoad yes

NeedsCompilation yes

Repository CRAN

Date/Publication 2017-01-25 09:02:46

R topics documented:

codep-package	2
cthreshold	3
Doubs	4
eigenmap	5
eigenmap-class	9
gcd	10
MCA	12
mca-class	16
minpermute	18
Mite	19
Product-distribution	21
Salmon	23

Index	25
--------------	-----------

codep-package

Functions and methods to perform Multiscale Codependence Analysis (MCA)

Description

Multiscale Codependence Analysis (MCA) consists in assessing the coherence of pairs of variables in space (or time) using the product of their correlation coefficients with series of spatial (or temporal) eigenfunctions. That product, which is positive or negative when variables show similar or opposing trends, respectively, are called codependence coefficients. These eigenfunctions are obtained in three steps: 1) a distance matrix calculated from the locations of samples in space (or the organisation of the sampling schedule). 2) from that distance matrix, a matrix of spatial weights is obtained; the same matrix as to calculate Moran's autocorrelation index, hence the name, and 3) the spatial weight matrix is eigenvalue-decomposed after centering each rows and columns of the spatial weight matrix.

The statistical significance of the codependence coefficients is tested using parametric or permutational testing of a τ statistic. The τ statistic is the product of the two Student's t statistics obtained from each of the two variables with a given eigenfunction. The τ statistic can take both positive and negative values, thereby allowing one to perform one-directional or two-directional testing. For multiple response variables, testing is performed using the ϕ statistic instead. That statistics is the distribution of the product of two Fisher-Snedocor F statistics (see [Product-distribution](#) for details).

Details

Package:	codep
Type:	Package
Version:	0.6-2
Date:	2015-11-10
License:	Copyleft
LazyLoad:	yes

Function [MCA](#) performs Multiscale Codependence Analysis (MCA).

Functions [test.mca](#) and [permute.mca](#) handle parametric permutational testing of the codependence coefficients, respectively.

Methods are provided to print and plot [mca-class](#) objects ([print.mca](#) and [plot.mca](#), respectively) as well as summary ([summary.mca](#)), fitted values ([fitted.mca](#)), residuals ([residuals.mca](#)), and to make predictions ([predict.mca](#)).

Function [eigenmap](#) calculates spatial eigenvector maps following the approach outlined in Dray et al. (2006), and which are necessary to calculate [MCA](#). It returns a [eigenmap-class](#) object. The package also features methods to print ([print.eigenmap](#)) and plot ([plot.eigenmap](#)) these objects. Function [eigenmap.score](#) can be used to make predictions for spatial models built from the eigenfunctions of [eigenmap](#) using distances between one or more target locations and the sampled locations for which the spatial eigenvector map was built.

The package also features an exemplary dataset [Salmon](#) containing 76 sampling site positions along a 1520 m river segment as well as functions [cthreshold](#) and [minpermute](#), which calculates the testwise type I error rate threshold corresponding to a given familywise threshold and the minimal number of permutations needed for testing Multiscale Codependence Analysis given the alpha threshold, respectively.

Author(s)

Guillaume Guénard, Département des sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

- Dray, S.; Legendre, P. and Peres-Neto, P. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecol. Modelling* 196: 483-493
- Guénard, G., Legendre, P., Boisclair, D., and Bilodeau, M. 2010. Multiscale codependence analysis: an integrated approach to analyse relationships across scales. *Ecology* 91: 2952-2964

See Also

Legendre, P. and Legendre, L. 2012. *Numerical Ecology*, 3rd English edition. Elsevier Science B.V., Amsterdam, The Neatherlands.

cthreshold	<i>Familywise type I error rate</i>
------------	-------------------------------------

Description

Function to calculate the testwise type I error rate threshold corresponding to a give familywise threshold.

Usage

```
cthreshold(alpha, nbtest)
```

Arguments

alpha	The familywise type I error threshold.
nbtest	The number of tests performed.

Details

Type I error rate inflation occurs when a single hypothesis is tested indirectly using inferences about two or more (*i.e.* a family of) sub-hypotheses. In such situation, the probability of type I error (*i.e.* the probability of incorrectly rejecting the null hypothesis) of the single, familywise, hypothesis is higher than the lowest, testwise, probabilities. As a consequence, the rejection of null hypothesis for one or more individual tests does not warrant that the correct decision (whether to reject the the null hypothesis on a familywise basis) was taken properly. This function allows to obtain correct, familywise, alpha thresholds in the context of multiple testing. It is base on the Sidak inequality.

Value

The threshold that have to be used for individual tests.

Author(s)

Guillaume Guénard, Département des sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

Sidak, Z. 1967. Rectangular Confidence Regions for Means of Multivariate Normal Distributions J. Am. Stat. Assoc. 62: 626-633

Wright, P. S. 1992. Adjusted p-values for simultaneous inference. Biometrics 48: 1005-1013

See Also

Legendre, P. and Legendre, L. 1998. Numerical Ecology. Elsevier Science B.V., Amsterdam, The Neatherlands. p. 18

Examples

```
# For a familywise threshold of 5% with 5 tests:
cthreshold(c(0.05),5) # The corrected threshold for each test is 0.01020622
```

Doubs

The Doubs fish data

Description

Fish community composition of the Doubs River, France.

Usage

Doubs

Format

Contains three matrices: Doubs.fish the abundance of 27 fish species.

Doubs.env 9 environmental variables (all quantitative).

Doubs.geo geographic information of the samples.

Details

Values in `Doubs.fish` are counts of individuals of each of 27 species observed in a set of 30 sites located along the 453 km long Doubs River, France (see Verneaux 1973 for further details about fishing methods and effort).

`Doubs.env` contains 11 quantitative variables, namely the slope (`slo` 1/1000) and mean minimum discharge (`flo` m³/s) of the river, the pH of the water, its harness (Calcium concentration; `har`; mg/L), phosphate (`pho`; mg/L), nitrate (`nit`; mg/L), and ammonium (`amm`; mg/L), concentration as well as its dissolved oxygen (`oxy`; mg/L) and biological oxygen demand (`bdo`; mg/L).

`Doubs.geo` contains geographical information. `Lon`, the longitude and `Lat`, the latitude of the sample (degree) as well as `DFS`, its distance from the source of the river (km) and `Alt`, altitude (m above sea level).

Source

Verneaux, 1973

References

Verneaux J. 1973. - Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie. Thèse d'état, Besançon. 257 p.)

Verneaux, J.; Schmitt, V.; Verneaux, V. & Prouteau, C. 2003. Benthic insects and fish of the Doubs River system: typological traits and the development of a species continuum in a theoretically extrapolated watercourse. *Hydrobiologia* 490: 60-74

See Also

Borcard, D.; Gillet, F. & Legendre, P. 2011. *Numerical Ecology* with R. Springer, New-York, NY, USA.

Examples

```
data(Doubs)
summary(Doubs.fish)
summary(Doubs.env)
summary(Doubs.geo)
```

eigenmap

Spatial eigenvector maps

Description

Function to calculate spatial eigenvector maps of a set of locations in a space with an arbitrary number of dimension.

Usage

```
eigenmap(x, opt.coord=NA, weighting=Wf.sqrd, boundaries, wpar, select=.Machine$double.eps^0.5)
Wf.sqrd(D)
Wf.binary(D, boundaries)
Wf.PCNM(D, boundaries)
Wf.Drayf1(D, boundaries)
Wf.Drayf2(D, boundaries, wpar=1)
Wf.Drayf3(D, boundaries, wpar=1)
eigenmap.score(object, target)
```

Arguments

x	A set of coordinates defined in one (numeric vector) or many (a coordinate x dimension matrix) dimensions or, alternatively, a distance matrix provided by dist . Coordinates are treated as cartesian coordinates and the distances between them are assumed to be Euclidean.
opt.coord	Coordinates to be used when a distance matrix is provided as x. Used for plotting purposes.
weighting	<p>The function to obtain the edge weighting matrix. That function must have the raw distances as a first parameter, optionally a second parameter named <i>boundaries</i> giving the boundaries of the within which locations are regarded as neighbour and a third parameter named <i>wpar</i> containing any other weighting function parameter.</p> <p>Wf.binary (default value) the spatial weighting matrix is simply the connectivity matrix,</p> <p>Wf.PCNM is $a_{i,j} = 1 - (d_{i,j} / (4 * boundaries_2))^2$</p> <p>Wf.Drayf1 is $a_{i,j} = 1 - (d_{i,j} / d_{max})$ where d_{max} is the distance between the two most distant locations in the set,</p> <p>Wf.Drayf2 is $a_{i,j} = 1 - (d_{i,j} / d_{max})^{wpar}$,</p> <p>Wf.Drayf3 is $a_{i,j} = 1 / d_{i,j}^{wpar}$, and</p> <p>Wf.sqrd consists in taking $w_{i,j} = -0.5 * d_{i,j}$ and does not involve any truncation. Functions Wf.Drayf1, Wf.Drayf2, and Wf.Drayf3 were proposed by Dray et al. (2006) and function PCNM was proposed by Legendre and Legendre (2012). The Wf.sqrd weighting approach is equivalent to submitting the elementwise square-root of the distance matrix to a principal coordinate analysis. That option is not much documented in the ecological literature, but is actually equivalent, for evenly spaced transect or surfaces (square or rectangle), to using the basis functions of type II discrete cosine basis transforms.</p>
boundaries	(optional) Threshold values (minimum and maximum) used to obtain the connectivity matrix. Pairs of location whose distance to one another are between these values are considered as neighbours ($b_{i,j} = 1$) whereas values located below the minimum and above the maximum are considered as equivalent or distant, respectively ($b_{i,j} = 0$ in both cases). Defaults are 0 for the minimum value and NA for the maximum. Values NA indicates the function to take the minimum value that allow every locations to form a single cluster following single linkage clustering as a maximum value (obtained from hclust . Ignored when <i>weighting</i> ="sqrd".

wpar	Weighting function parameters.
select	The smallest absolute eigenvalue for eigenfunctions to be considered as a suitable predictive variables. Default value depends on one's particular computer and is set to the square-root of <code>.Machine\$double.eps</code>
D	A distance matrix.
object	A <code>eigenmap-class</code> object.
target	A set of distances between the sampling locations (passed to <code>eigenmap</code> using <code>x</code>) and the target locations where spatially-explicit predictions are to be made.

Details

Spatial eigenvector maps are sets of eigenfunctions obtained from the locations of the observations in a structuring framework, e.g., space, time, or in a graph. It is obtained by eigenvalue decomposition of a spatial weighting matrix, computed as described in Dray et al. (2006) and Legendre & Legendre (2012, Section 14.2). That square matrix is Gower-centred before eigen-decomposition. The spatial weighting matrix is the Hadamard product of a connectivity matrix **B** and an edge weighting matrix **A**. The function described herein handles user-chosen truncation parameters to calculate **B** and provides a default approach to estimate these parameters should they be missing. It also offers four different ways of computing **A** through parameters `weighting` and `wpar`.

It is noteworthy that in the present implementation, matrix **B** is not obtained using a minimum spanning tree as suggested by Dray et al. (2006) but using a simpler approach whereby every distances within a user-defined truncation interval are taken as neighbour.

Functions `Wf.sqrd`, `Wf.binary`, `Wf.PCNM`, `Wf.Drayf1`, `Wf.Drayf2`, and `Wf.Drayf3` are not intended to be called as is but through `eigenmap` (and within `eigenmap.score`). Other, user-defined, function can be used by `eigenmap` and should be visible to if one wants to call `eigenmap.score` to obtain predictors.

For `eigenmap.score`, the distances between sampling locations and the targets locations must be of the same type as those that had been passed to `eigenmap`. If cartesian coordinates were passed to `x`, the distances to target must be Euclidean.

Value

`eigenmap` returns a `eigenmap-class` object and `eigenmap.score` returns a the scores on for each target locations

Author(s)

Guillaume Guénard, Département des sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

- Borcard, D. and Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol. Model.* 153: 51-68
- Dray, S.; Legendre, P. and Peres-Neto, P. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecol. Modelling* 196: 483-493

Legendre, P. and Legendre, L. 2012. Numerical Ecology, 3rd English edition. Elsevier Science B.V., Amsterdam, The Neatherlands.

See Also

[MCA eigenmap-class](#)

Examples

```
#
### Example 1: A linear transect.
#
data(Salmon)
#
## No boundaries provided for a function that requires them: a warning is issued
map <- eigenmap(x=Salmon[, "Position"], weighting=Wf.binary)
map # plot(map)
#
## Boundaries are provided: the function is happy
map <- eigenmap(x=Salmon[, "Position"], weighting=Wf.binary, boundaries=c(0,20))
map # plot(map)
#
map <- eigenmap(x=Salmon[, "Position"], weighting=Wf.Drayf1, boundaries=c(0,20))
map # plot(map)
#
map <- eigenmap(x=Salmon[, "Position"], weighting=Wf.Drayf2, boundaries=c(0,20))
map # plot(map)
#
map <- eigenmap(x=Salmon[, "Position"], weighting=Wf.Drayf3, boundaries=c(0,20), wpar=2)
map # plot(map)
#
map <- eigenmap(x=Salmon[, "Position"], weighting=Wf.PCNM, boundaries=c(0,20))
map # plot(map)
#
map <- eigenmap(x=Salmon[, "Position"], weighting=Wf.sqrd)
map # plot(map)
#
### Example 2: Using predictor scores
#
smpl <- c(4,7,10,14,34,56,61,64) # A sample to discard
map <- eigenmap(x=Salmon[-smpl, "Position"], weighting=Wf.sqrd)
scr <- eigenmap.score(object=map, target=as.matrix(dist(Salmon[, "Position"]))[-smpl])
all(round(scr[-smpl,] - map$U, 10) == 0) # Scores of sampling points are the eigenvectors
scr[smpl,]
#
wh <- 5L # You can try with other vectors.
plot(map$U[,wh]~Salmon[-smpl, "Position"], ylab = expression(U[5]),
      xlab = "Position along transect")
points(y=scr[smpl,wh], x=Salmon[smpl, "Position"], pch=21, bg="black")
#
map <- eigenmap(x=Salmon[-smpl, "Position"], weighting=Wf.binary, boundaries=c(0,20))
scr <- eigenmap.score(object=map, target=as.matrix(dist(Salmon[, "Position"]))[smpl, -smpl])
#
```



```

wh <- 1L # You can try with other vectors.
plot(map$U[,wh]~Salmon[-smp1,"Position"], ylab = expression(U[1]),
      xlab = "Position along transect (m)")
points(y=scr[,wh],x=Salmon[smp1,"Position"],pch=21,bg="black")
#
map <- eigenmap(x=Salmon[-smp1,"Position"],weighting=Wf.PCNM,boundaries=c(0,100))
scr <- eigenmap.score(object=map,target=as.matrix(dist(Salmon[, "Position"]))[smp1,-smp1])
#
wh <- 1L # You can try with other vectors.
plot(map$U[,wh]~Salmon[-smp1,"Position"], ylab = expression(U[1]),
      xlab = "Position along transect (m)")
points(y=scr[,wh],x=Salmon[smp1,"Position"],pch=21,bg="black")
#
### Example 3: A unevenly sampled surface.
#
data(Mite)
map <- eigenmap(x=as.matrix(mite.geo),weighting=Wf.sqrd)
map # plot(map)
#

```

eigenmap-class

Class and methods for spatial eigenvector maps

Description

Methods to handle spatial eigenvector maps of a set of locations in a space with an arbitrary number of dimension.

Usage

```

## S3 method for class 'eigenmap'
print(x, ...)
## S3 method for class 'eigenmap'
plot(x, ...)

```

Arguments

x	An object of eigenmap-class
...	Further parameters to be passed to other functions or methods (currently ignored).

Details

The print method provides the number of the number of orthonormal variables (i.e. basis functions), the number of observations these functions are spanning, and their associated eigenvalues.

The plot method provides a plot of the eigenvalues and offers the possibility to plot the values of variables for 1- or 2-dimensional sets of coordinates. `plot.eigenmap` opens the default graphical device driver, i.e., X11, windows, or quartz and recurses through variable with a left mouse click on the graphical window. A right mouse click interrupts recursing on X11 and windows (Mac OS X users should hit *Esc* on the quartz graphical device driver (Mac OS X users)).

Value

[eigenmap-class](#) objects contain:

coordinates	a matrix of coordinates,
truncate	the interval within which pairs of sites are considered as neighbour,
D	the distance matrix,
weighting	the weighting function that had been used,
wpar	the weighting function parameter that had been used,
lambda	a vector of the eigenvalues obtain from the computation of the eigenvector map, and
U	a matrix of the eigenvectors defining the eigenvector map.

Author(s)

Guillaume Guénard, Département des sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

Dray, S.; Legendre, P. and Peres-Neto, P. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecol. Modelling* 196: 483-493

See Also

[MCA eigenmap](#)

gcd

Great circle distances

Description

Functions to calculate great circle distances among sets of points with known longitude and latitudes.

Usage

```
gcd.slc(ss, radius = 6371)
gcd.hf(ss, radius = 6371)
gcd.vife(ss, a = 6378137, b = 6356752.314245, f = 1/298.257223563)
```

Arguments

<code>ss</code>	A two-columns data.frame or matrix of the geographic coordinates (in degree.decimals; latitude and then longitude) of the locations between which the geodesic distances are being calculated.
<code>radius</code>	Mean earth radius (mean length of parallels) in km. Default: 6371 km
<code>a</code>	Length (in m) of major axis of the ellipsoid (radius at equator). Default: 6378137 m (i.e. that for WGS-84).
<code>b</code>	Length (in m) of minor axis of the ellipsoid (radius at the poles). Default: 6356752.314245 m (i.e. that for WGS-84).
<code>f</code>	Flattening of the ellipsoid. Default: 1/298.257223563 (i.e. that for WGS-84).

Details

The calculation of spatial eigenvector maps requires a distance matrix. The euclidean distance is appropriate when a cartesian plan can be reasonably assumed. The latter can be calculated with function `dist`. The great circle distance is appropriate when the sampling points are located on a spheroid (e.g. planet Earth). Function `gcd.slc` uses the spherical law of cosines, which is the fastest of the three approaches and performs relatively well for distances above 1 m. Function `gcd.hf` uses the the Haversine formula, which is more accurate for smaller distances (bellow 1 m). Finally, functions `gcd.vife` uses the Vincenty inverse formula for ellipsoids, which is an iterative approach that take substantially more computation time than the latter two methods has precision up to 0.5 mm with exact longitudes and latitudes.

Value

An object of class "dist" (see `dist` for details) that contains the distances (in km) between the locations (rows of `ss`).

Author(s)

Guillaume Guénard, Departement des sciences biologiques, Universite de Montréal, Montréal, Quebec, Canada.

References

- Mario Pineda-Krch, URL: <http://www.r-bloggers.com/great-circle-distance-calculations-in-r/>
- Vincenty, T. (1975) Closed formulas for the direct and reverse geodetic problems. J. Geodesy 51(3): 241-342

See Also

[eigenmap-class](#)

Examples

```
#
# Calculating the distances between Canada's capital cities:
CapitalCitiesOfCanada <-
  matrix(c(45.417,-75.7,53.533333,-113.5,48.422151,-123.3657,
    49.899444,-97.139167,45.95,-66.666667,47.5675,-52.707222,
    44.647778,-63.571389,43.7,-79.4,46.24,-63.1399,
    46.816667,-71.216667,50.454722,-104.606667,62.442222,-114.3975,
    63.748611,-68.519722,60.716667,-135.05),14L,2L,byrow=TRUE,
    dimnames=list(c("Ottawa","Edmonton","Victoria","Winnipeg",
      "Fredericton","St-John's","Halifax","Toronto",
      "Charlottetown","Quebec City","Regina",
      "Yellowknife","Iqaluit","Whitehorse"),c("Lon","Lat"))

#
sphericalcosdists <- gcd.slc(CapitalCitiesOfCanada)
vincentydists <- gcd.vife(CapitalCitiesOfCanada)
#
cor(as.numeric(sphericalcosdists),as.numeric(vincentydists))
percentdev <- 100*(vincentydists-sphericalcosdists)/vincentydists
mean(percentdev)
# Spherical Law of Cosines underestimated these distances by ~0.26
# percent.
#
```

MCA

Multiple-descriptors, Multiscale Codependence Analysis

Description

Functions to perform Multiscale Codependence Analysis (MCA)

Usage

```
MCA(Y, X, emobj)
test.mca(object, alpha = 0.05, max.step, response.tests = TRUE)
permute.mca(object, permute, alpha = 0.05, max.step, response.tests = TRUE)
```

Arguments

Y	a numeric matrix or vector containing the response variable(s).
X	a numeric matrix or vector containing the explanatory variable(s).
emobj	a eigenmap-class object.
object	a mca-class object.
alpha	type I (α) error threshold used by the testing procedure.
max.step	maximum number of steps to perform when testing for statistical significance.
response.tests	a boolean specifying whether to test individual response variables.
permute	The number of random permutations used for testing. When omitted, the number of permutations is calculated using function minpermute .

Details

Multiscale Codependence Analysis (MCA) allows to calculate correlation-like (i.e. codependence) coefficients between two variables with respect to structuring variables (Moran's eigenvector maps). The purpose of this function is limited to parameter fitting. Test procedures are handled through `test.mmca` (parametric testing) or `permute.mmca` (permutation testing). Additionally, methods are provided for printing, displaying the testing summary, plotting results, calculating fitted and residuals values, and making predictions. It is noteworthy that the test procedure used by MCA deviates from the standard R workflow since intermediate testing functions (`test.mca` and `permute.mca`) need first to be called before any testing be performed. For MCA, testing functionalities had been moved away from `summary.mca` because testing is computationally intensive.

Value

A `mca-class` object.

Author(s)

Guillaume Guénard, Département des sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

- Guénard, G., Legendre, P., Boisclair, D., and Bilodeau, M. 2010. Multiscale codependence analysis: an integrated approach to analyse relationships across scales. *Ecology* 91: 2952-2964
- Guénard, G. Legendre, P., and crews. 2016. Assessing multivariate relationships across spatial scales: introducing the Multivariate Multiscale Codependence Analysis. *Ecology* paper to be.

See Also

[eigenmap](#)

Examples

```
#
##### Begin {Salmon exemple}
#
data(Salmon)
#
## Converting the data from data frames to matrices:
Abundance <- log1p(as.matrix(Salmon[, "Abundance", drop=FALSE]))
Environ <- as.matrix(Salmon[, 3L:5])
#
## Creating a spatial eigenvector map:
map1 <- eigenmap(x=Salmon[, "Position"], weighting=Wf.binary, boundaries=c(0,20))
#
## Case of a single descriptor:
mca1 <- MCA(Y=Abundance, X=Environ[, "Substrate", drop=FALSE], emobj=map1)
mca1
mca1_partest <- test.mca(mca1)
mca1_partest
```

```

summary(mca1_partest)
par(mar = c(6,4,2,4))
plot(mca1_partest, las = 3)
mca1_pertest <- permute.mca(mca1)
mca1_pertest
summary(mca1_pertest)
plot(mca1_pertest, las = 3)
mca1_pertest$UpYXcb$C # Array containing the codependence coefficients
#
## With all descriptors at once:
mca2 <- MCA(Y=log1p(as.matrix(Salmon[, "Abundance", drop=FALSE])),
            X=as.matrix(Salmon[, 3L:5]), emobj=map1)

mca2
mca2_partest <- test.mca(mca2)
mca2_partest
summary(mca2_partest)
par(mar = c(6,4,2,4))
plot(mca2_partest, las = 3)
## Not run:
mca2_pertest <- permute.mca(mca2)
mca2_pertest
summary(mca2_pertest)
plot(mca2_pertest, las = 3)
mca2_pertest$UpYXcb$C # Array containing the codependence coefficients
mca2_pertest$UpYXcb$C[, 1L, ] # now turned into a matrix.
#
##### End {Salmon exemple}
#
##### Begin {Doubs exemple}
#
data(Doubs)
#
## Creating a spatial eigenvector map:
map2 <- eigenmap(x=Doubs.geo[, "DFS"])
#
mca3 <- MCA(Y=log1p(Doubs.fish), X=Doubs.env, emobj=map2)
mca3
mca3_pertest <- permute.mca(mca3)
mca3_pertest
summary(mca3_pertest)
par(mar = c(6,4,2,4))
plot(mca3_pertest, las = 2)
mca3_pertest$UpYXcb$C # Array containing the codependence coefficients
#
## Display the results along the transect
spmeans <- colMeans(log1p(Doubs.fish))
pca1 <- svd(log1p(Doubs.fish) - rep(spmeans, each=nrow(Doubs.fish)))
par(mar = c(5,5,2,5)+0.1)
plot(y = pca1$u[, 1L], x = Doubs.geo[, "DFS"], pch = 21L, bg = "red",
     ylab = "PCA1 loadings", xlab = "Distance from river source (km)")
#
x <- seq(0, 450, 1)
newdists <- matrix(NA, length(x), nrow(Doubs.geo))

```

```

for(i in 1L:nrow(newdists))
  newdists[i,] <- abs(Doubs.geo[, "DFS"] - x[i])
#
## Calculating predictions for arbitrary sites under the same set of
## environmental conditions that the codependence model was built with.
prd1 <- predict(mca3_pertest,
  newdata=list(target = eigenmap.score(map2, newdists)))
#
## Projection of the predicted species abundance on pca1:
Uprd1 <- (prd1 - rep(spmeans, each = nrow(prd1))) %*% pca1$v %*% diag(pca1$d^-1)
lines(y = Uprd1[,1L], x = x, col=2, lty = 1)
#
## Projection of the predicted species abundance on pca2:
plot(y = pca1$u[,2L], x = Doubs.geo[, "DFS"], pch = 21L, bg = "red",
  ylab = "PCA2 loadings", xlab = "Distance from river source (km)")
lines(y = Uprd1[,2L], x = x, col=2, lty = 1)

#
## Displaying only the observed and predicted abundance for Brown Trout.
par(new=TRUE)
plot(y = log1p(Doubs.fish[, "TRU"]), Doubs.geo[, "DFS"], pch=21L, bg="green",
  ylab="", xlab="", new=FALSE, axes=FALSE)
axis(4)
lines(y = prd1[, "TRU"], x = x, col=3)
mtext(side=4, "log(Abundance+1)", line = 2.5)
#
##### End {Doubs exemple}
#

## End(Not run)
##### Begin {Oribatid exemple}
#
data(Mite)
#
map3 <- eigenmap(x = mite.geo)
# Organize the environmental variables
mca4 <- MCA(Y = log1p(mite.species), X = mite.env, emobj = map3)
mca4_partest <- test.mca(mca4, response.tests = FALSE)
summary(mca4_partest)
plot(mca4_partest, las = 2, lwd = 2)
plot(mca4_partest, col = rainbow(1200)[1L:1000], las = 3, lwd = 4,
  main = "Codependence diagram", col.signif = "white")
#
rng <- list(x = seq(min(mite.geo[, "x"]) - 0.1, max(mite.geo[, "x"]) + 0.1, 0.05),
  y = seq(min(mite.geo[, "y"]) - 0.1, max(mite.geo[, "y"]) + 0.1, 0.05))
grid <- cbind(x = rep(rng[["x"]], length(rng[["y"]])),
  y = rep(rng[["y"]], each = length(rng[["x"]]))))
newdists <- matrix(NA, nrow(grid), nrow(mite.geo))
for(i in 1L:nrow(grid)) {
  newdists[i,] <- ((mite.geo[, "x"] - grid[i, "x"])^2 +
    (mite.geo[, "y"] - grid[i, "y"])^2)^0.5
}
#

```

```

spmeans <- colMeans(mite.species)
pca2 <- svd(log1p(mite.species) - rep(spmeans, each = nrow(mite.species)))
#
prd2 <- predict(mca4_partest,
  newdata = list(target = eigenmap.score(map3, newdists)))
Uprd2 <- (prd2 - rep(spmeans, each = nrow(prd2))) %*% pca2$v %*% diag(pca2$d^-1)
#
### Printing the response variable
prmat <- Uprd2[,1L]
dim(prmat) <- c(length(rng$x),length(rng$y))
zlim <- c(min(min(prmat),min(pca2$u[,1L])),max(max(prmat),max(pca2$u[,1L])))
image(z = prmat, x = rng$x, y = rng$y, asp = 1, zlim = zlim,
  col = rainbow(1200L)[1L:1000], ylab = "y", xlab = "x")
points(x = mite.geo[, "x"], y = mite.geo[, "y"], pch = 21,
  bg = rainbow(1200L)[round(1+(999*(pca2$u[,1L]-zlim[1L])/(zlim[2L]-zlim[1L])),0)])
#
##### End {Oribatid exemple}
#

```

mca-class

Class and methods for Multiscale Codependence Analysis involving multiple descriptors

Description

Class and methods to handle Multiscale Codependence Analysis (MMCA)

Usage

```

## S3 method for class 'mca'
print(x, ...)
## S3 method for class 'mca'
plot(x, col, col.signif=2, main="", ...)
## S3 method for class 'mca'
summary(object, ...)
## S3 method for class 'mca'
fitted(object, selection, components=FALSE, ...)
## S3 method for class 'mca'
residuals(object, selection, ...)
## S3 method for class 'mca'
predict(object, selection, newdata, components=FALSE, ...)

```

Arguments

x, object	A mca-class object.
col	A vector of color values to be used for plotting the multivariate codependence coefficients.

<code>col.signif</code>	Color of the frame used to mark the statistically significant codependence coefficients .
<code>main</code>	Text for the main title of the plot.
<code>selection</code>	A numeric vector of indices or character vector variable names to test or force-use. Mandatory if object is untested.
<code>components</code>	A boolean specifying whether the components of fitted or predicted values associated with single eigenfunctions in the map should be returned.
<code>newdata</code>	A list with elements <code>\$X</code> , <code>\$meanY</code> , and <code>\$target</code> that contain the information needed to make predictions (see details).
<code>...</code>	Further parameters to be passed to other functions or methods.

Details

The fitted, residuals, and predict methods return a matrix of fitted, residuals, or predicted values, respectively. The fitted and predict methods return a list a list when the parameter component is TRUE. The list contains the fitted or predicted values as a first element and an array components as a second. That 3-dimensional array has one matrix for each statistically significant codependence coefficient.

For making predictions, parameter newdata may contain three elements: `$X`, a matrix of new values of the explanatory variables, `$meanY`, a vector of the predicted mean values of the responses, and `$target`, a matrix of target scores for arbitrary locations within the study area. When no `$X` is supplied, the descriptor given to `MCA` is recycled, while when no `$meanY` is supplied, the mean values of the response variables given to `MCA` are used. Finally, when element `$target` is omitted from newdata, predictions are made at the sites were observations were done. When none of the above is provided, or if newdata is omitted when calling the prediction method, the behaviour of the predict method is identical to that of the fitted method.

Value

`mca-class` objects contain:

<code>data</code>	A list with two elements: the first being a copy of the response (Y) and the second being a copy of the explanatory variables (X). This is the variables that were given to <code>MCA</code> .
<code>emobj</code>	The <code>eigenmap-class</code> object that was given to <code>MCA</code> .
<code>UpYXcb</code>	A list with five elements: the first (<code>UpY</code>) is a matrix of the cross-products of structuring variable (U) and the response variable Y, the second (<code>UpX</code>) is a matrix of the cross-product of the structuring variable and the explanatory variables (X), the third (C) is a 3-dimensional array of the codependence coefficients, the fourth (B) is a 3-dimensional array of the coregression coefficients, and the fifth (CM) is a matrix of the multivariate codependence coefficients.
<code>test</code>	Results of statistical testing as performed by <code>test.mca</code> or <code>permute.mca</code> . NULL if no testing was performed, such as when only <code>MCA</code> had been called. The results of statistical testing is a list containing the following members:
<code>\$permute</code>	The number of randomized permutations used by <code>permute.mmca</code> for permutation testing. 0 or FALSE for parametric testing obtained using <code>test.mca</code> .

\$significant	The indices of codependence coefficient describing statistically significant codependence between Y and X, in decreasing order of magnitude.
\$global	The testing table (a 5-column matrix) with ϕ statistics, degrees-of-freedom, and testwise and familywise probabilities of type I (α) error. It contains one line for each statistically significant global coefficient (if any) in addition to test results for the first, non-significant coefficient, on which the testing procedure stopped.
\$response	Tests of every single response variable (a 3-dimensional array), had such tests been requested while calling the testing function, NULL otherwise.
\$permutations	Details about permutation testing not shown in test\$global or test\$response. NULL for parametric testing.

Author(s)

Guillaume Guénard, Département des sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

Guénard, G., Legendre, P., Boisclair, D., and Bilodeau, M. 2010. Multiscale codependence analysis: an integrated approach to analyse relationships across scales. *Ecology* 91: 2952-2964

Guénard, G. Legendre, P., and crews. 2016. Assessing multivariate relationships across spatial scales: introducing the Multivariate Multiscale Codependence Analysis. *Ecology* paper to be.

minpermute	<i>Number of permutations for MCA</i>
------------	---------------------------------------

Description

Calculate the number of permutations suitable for testing Multiscale Codependence Analysis.

Usage

```
minpermute(alpha,nbtest,margin=1,ru=3)
```

Arguments

alpha	The familywise type I error threshold allowable for the complete analysis.
nbtest	The number of test performed (the number of eigenvectors in the 'mem' object in the case of MCA).
margin	A margin allowed for the number of permutation. Default value: 1.
ru	The magnitude of the round-up to apply to the number of permutations.

Details

This function calculate the number of permutations for use with `permute.mca`. Parameter `margin` allows to apply a safe margin to the number of permutations. The minimal suitable value for this parameter is 1. Parameter `ru` allows one to round-up the number of permutations. A value of 0 implies no round-up, a value of 1 a round-up to the next ten, 2 a round-up to the next hundred, and so on. Function `minpermute` is called internally by `permute.mca` in case `permute = NA`. In that case, the margin is set to 10 (`margin = 10`) and the outcome is rounded-up to the next thousand (`ru = 3`). This function is meant for users that wish to apply their own margins and round-up factors to calculate the number of permutations for use with `permute.mca`.

Value

The minimum number of permutation to be used for `permute.mca`.

Author(s)

Guillaume Guénard, Département des sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

Guénard, G., Legendre, P., Boisclair, D., and Bilodeau, M. 2010. Multiscale codependence analysis: an integrated approach to analyse relationships across scales. *Ecology* 91: 2952-2964

See Also

[permute.mca](#)

Examples

```
# For a 5% threshold under 50 tests.
minpermute(alpha = 0.05, nbtest=50)
# Allowing more margin (implies more computation time).
minpermute(alpha = 0.05, nbtest=50, margin=10, ru=3)
```

Mite

Lac Geai oribatid mites community data

Description

A data set containing information on the microfauna of 70 sites located within the the peat on shore of Lac Geai, a bog lake located at the Station de Biologie de l'Université de Montréal, St-Hippolyte, QC, Canada (+45.9954; -73.9936).

Usage

Mite

Format

Contains three matrices: `mite.species` the abundance of 35 morpho-species of oribatid mites (Acari).

`mite.env` 14 environmental variables (quantitative and binary).

`mite.geo` the relative coordinates of the samples.

Details

Values in `mite.species` are counts of individuals of each of the morpho-species obtained from 5 cm diameter cores going from the surface of the peat down to a depth of 7 cm. See Bordard & Legendre (1994) and reference therein for details about sample treatment and species identification.

`mite.env` contains two quantitative variables, namely the substratum density (g/L) and water content (percent wet mass over dry mass), in addition to 12 dummy variables. The first seven represent the composition of the substratum: *Sphagnum magellacinum* (with a majority of *S. rubellum*), *S. rubellum*, *S. nemorum*, (with a majority of *S. angustifolium*), *S. rubellum* + *S. magellanicum* (in equal proportions), lignous litter, bare peat, and interface between *Sphagnum* species. The next three dummy variables represent the presence and abundance of shrubs (*Kalmia polifolia*, *K. angustifolia*, and *Rhododendron groenlandicum*): none, few, and many. The last two dummy variables represent the microtopography of the peat: blanket (flat) or hummock (raised).

`mite.geo` contains the location of the samples, in meters, with respect to the sampling grid. Point (0,0) is the lower left end of the plot for an observer looking from the shore towards the water. The x coordinate is the offset along the shore (from left to right) while the y coordinate is the offset from the shore while moving towards the water (See Borcard & Legendre, 1994, Fig. 1 for details on the sampling area).

Source

Daniel Borcard, Département de sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

Borcard, D. & Legendre, P. 1994. Environmental control and spatial structure in ecological communities: an example using Oribatid mites (Acari, Oribatei). *Environ. Ecol. Stat.* 1: 37-61

See Also

Borcard, D.; P. Legendre & P. Drapeau. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73: 1045-1055

Legendre, P. 2005. Species associations: the Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological and Environmental Statistics* 10: 226-245

Borcard, D.; Gillet, F. & Legendre, P. 2011. *Numerical Ecology* with R. Springer, New-York, NY, USA.

Examples

```
data(Mite)
summary(mite.species)
summary(mite.env)
summary(mite.geo)
```

Product-distribution *Frequency distributions for MCA parametric testing*

Description

Density and distribution functions of the phi statistic, which is the product of two Fisher-Snedecor distributions with particular degrees of freedom.

Usage

```
dphi(x, nu1, nu2, tol = .Machine$double.eps ^ 0.5)
pphi(q, nu1, nu2, lower.tail = TRUE, tol = .Machine$double.eps ^ 0.5)
dtau(x, nu, tol = .Machine$double.eps ^ 0.5)
ptau(q, nu, lower.tail = TRUE, tol = .Machine$double.eps ^ 0.5)
```

Arguments

<code>x, q</code>	vector of quantile.
<code>nu1, nu2, nu</code>	degrees of freedom (>0, may be non-integer). Inf is allowed.
<code>lower.tail</code>	logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.
<code>tol</code>	tolerance used for numerical estimation.

Details

The density distribution of a variable z that is the product of two random variables x and y with density distributions $f(x)$ and $g(y)$, respectively, is the integral over the intersection of the domains of x and y of $f(x) * g(z/x) / \text{abs}(x)$ dx.

`dphi` estimates density values using numerical integration ([integrate](#)) the Fisher-Snedecor [df](#) density distribution function. Following the algebra of Multiscale Codependence Analysis, $f(x)$ has $\text{df1} = \text{nu1}$ and $\text{df2} = \text{nu1} * \text{nu2}$ degrees of freedom and $g(x)$ has ' $\text{df1} = 1$ ' and ' $\text{df2} = \text{nu2}$ ' degrees of freedom. Hence, that product distribution has two parameters.

`pphi` integrates `dphi` in the interval $[0, q]$ when '`lower.tail = TRUE`' (the default) and on the interval $[q, \text{Inf}]$ when '`lower.tail = FALSE`'.

`dtau` and `ptau` are similar to `dphi` integrates `pphi`, but with $f(x)$ and $f(y)$ being two Student's t distribution with nu degrees of freedom. It is called by functions [test.mca](#) and [permute.mca](#) to perform hypothesis tests for single response variables, in which case unilateral tests can be performed.

Value

dphi and dtau return the density functions whereas pphi and ptau return the distribution functions.

Author(s)

Guillaume Guénard, Département des sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

Springer, M. D. 1979. The algebra of random variables. John Wiley and Sons Inc., Hoboken, NJ, USA.

Guénard, G., Legendre, P., Boisclair, D., and Bilodeau, M. 2010. Multiscale codependence analysis: an integrated approach to analyse relationships across scales. *Ecology* 91: 2952-2964

Guénard, G. Legendre, P., and crews. 2016. Assessing multivariate relationships across spatial scales: introducing the Multivariate Multiscale Codependence Analysis. *Ecology paper to be.*

See Also

[test.mca](#)

Examples

```
#
### Displays the phi probability distribution for five different numbers
### of degrees of freedom:
#
x <- 10^seq(-4, 0.5, 0.05)
plot(y = dphi(x, 1, 10), x = x, type = "l", col = "black", las = 1, ylab = "pdf",
     ylim = c(0, 0.5))
lines(y = dphi(x, 3, 10), x = x, col = "purple")
lines(y = dphi(x, 5, 70), x = x, col = "blue")
lines(y = dphi(x, 12, 23), x = x, col = "green")
lines(y = dphi(x, 35, 140), x = x, col = "red")
#
### Displays the density distribution function for 10 degrees of freedom
### and the cumulative probability above x = 1.
#
x <- 10^seq(-4, 0.5, 0.05)
y <- dphi(x, 5, 70)
plot(y = y, x = x, type = "l", col = "black", las = 1, ylab = "Density",
     ylim = c(0, 0.5))
polygon(x = c(x[81L:91], x[length(x)], 1), y = c(y[81L:91], 0, 0),
       col = "grey")
text(round(pphi(1, 5, 70, lower.tail=FALSE), 3), x = 1.75, y = 0.05)
#
### Idem for the tau distribution:
#
x <- c(-(10^seq(0.5, -4, -0.05)), 10^seq(-4, 0.5, 0.05))
plot(y = dtau(x, 1), x = x, type = "l", col = "black", las = 1,
```

```

      ylab = "pdf", ylim = c(0, 0.5))
lines(y = dtau(x, 2), x = x, col = "purple")
lines(y = dtau(x, 5), x = x, col="blue")
lines(y = dtau(x, 10), x = x, col="green")
lines(y = dtau(x, 100), x = x, col="red")
#
y <- dtau(x, 10)
plot(y = y, x = x, type = "l", col = "black", las = 1, ylab = "Density",
     ylim = c(0, 0.5))
polygon(x = c(x[which(x==1):length(x)], x[length(x)],1),
       y = c(y[which(x==1):length(x)], 0, 0), col = "grey")
text(round(ptau(1, 10, lower.tail = FALSE), 3), x = 1.5, y = 0.03)
polygon(x = c(-1, x[1L], x[1L:which(x==1)]),
       y = c(0, 0, y[1L:which(x==1)]), col="grey")
text(round(ptau(-1, 10), 3), x = -1.5, y = 0.03)
#

```

Salmon

*Juvenile Atlantic salmon (parr) density in St-Marguerite river, NA***Description**

A 1520m transect of the St-Marguerite River.

Usage

Salmon

Format

A 76 rows by 5 columns [data.frame](#).

Details

Contains (1) the 76 sampling site positions along a 1520 m river segment beginning at a location called 'Bardville' (Lat: 48°23'01.59" N ; Long: 70°12'10.05" W), (2) the number of parr (young salmon, ages I+ and II+) observed at the sampling sites, (3) the mean water depths (m), (4) the mean current velocity (m/s), and (5) the mean substrate size (mm). Sampling took place on July 7, 2002, in the 76 sites, each 20 m long. The 'Bardville' river segment is located in the upper portion of Sainte-Marguerite River, Quebec, Canada.

Source

Daniel Boisclair, Département de sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

References

Guénard, G., Legendre, P., Boisclair, D., and Bilodeau, M. 2010. Multiscale codependence analysis: an integrated approach to analyse relationships across scales. *Ecology* 91: 2952-2964

See Also

Bouchard, J. and Boisclair, D. 2008. The relative importance of local, lateral, and longitudinal variables on the development of habitat quality models for a river. *Can. J. Fish. Aquat. Sci.* 65: 61-73

Examples

```
data(Salmon)
summary(Salmon)
```


Index

*Topic **Doubs**

Doubs, [4](#)

*Topic **Mite**

Mite, [19](#)

*Topic **Salmon**

Salmon, [23](#)

codep (codep-package), [2](#)

codep-package, [2](#)

cthreshold, [3](#), [3](#)

data.frame, [23](#)

df, [21](#)

dist, [6](#), [11](#)

Doubs, [4](#)

dphi (Product-distribution), [21](#)

dtau (Product-distribution), [21](#)

eigenmap, [2](#), [5](#), [7](#), [10](#), [13](#)

eigenmap-class, [9](#)

eigenmap.score, [2](#), [7](#)

fitted.mca, [2](#)

fitted.mca (mca-class), [16](#)

gcd, [10](#)

hclust, [6](#)

integrate, [21](#)

MCA, [2](#), [8](#), [10](#), [12](#), [17](#), [18](#)

mca (mca-class), [16](#)

mca-class, [16](#)

minpermute, [3](#), [12](#), [18](#)

Mite, [19](#)

mite.env (Mite), [19](#)

mite.geo (Mite), [19](#)

mite.species (Mite), [19](#)

permute.mca, [2](#), [17](#), [19](#), [21](#)

permute.mca (MCA), [12](#)

plot.eigenmap, [2](#)

plot.eigenmap (eigenmap-class), [9](#)

plot.mca, [2](#)

plot.mca (mca-class), [16](#)

pphi (Product-distribution), [21](#)

predict.mca, [2](#)

predict.mca (mca-class), [16](#)

print.eigenmap, [2](#)

print.eigenmap (eigenmap-class), [9](#)

print.mca, [2](#)

print.mca (mca-class), [16](#)

Product-distribution, [21](#)

ptau (Product-distribution), [21](#)

residuals.mca, [2](#)

residuals.mca (mca-class), [16](#)

Salmon, [3](#), [23](#)

summary.mca, [2](#)

summary.mca (mca-class), [16](#)

test.mca, [2](#), [17](#), [21](#), [22](#)

test.mca (MCA), [12](#)

Wf.binary, [7](#)

Wf.binary (eigenmap), [5](#)

Wf.Drayf1, [7](#)

Wf.Drayf1 (eigenmap), [5](#)

Wf.Drayf2, [7](#)

Wf.Drayf2 (eigenmap), [5](#)

Wf.Drayf3, [7](#)

Wf.Drayf3 (eigenmap), [5](#)

Wf.PCNM, [7](#)

Wf.PCNM (eigenmap), [5](#)

Wf.sqrd, [7](#)

Wf.sqrd (eigenmap), [5](#)