Package ‘zetadiv’

June 11, 2022

**Title** Functions to Compute Compositional Turnover Using Zeta Diversity

**Version** 1.2.1

**Date** 2022-06-08

**Description** Functions to compute compositional turnover using zeta-diversity, the number of species shared by multiple assemblages. The package includes functions to compute zeta-diversity for a specific number of assemblages and to compute zeta-diversity for a range of numbers of assemblages. It also includes functions to explain how zeta-diversity varies with distance and with differences in environmental variables between assemblages, using generalised linear models, linear models with negative constraints, generalised additive models, shape constrained additive models, and I-splines.

**Depends** R (>= 4.0.0), scam

**License** GPL-3

**Imports** stats, graphics, grDevices, utils, car, mgcv, vegan, geodist, nnls, glm2

**LazyData** true

**Encoding** UTF-8

**RoxygenNote** 7.2.0

**NeedsCompilation** no

**Author** Guillaume Latombe [aut, cre], Melodie A. McGeoch [aut], David A. Nipperess [aut], Cang Hui [aut]

**Maintainer** Guillaume Latombe <latombe.guillaume@gmail.com>

**Repository** CRAN

**Date/Publication** 2022-06-10 22:30:02 UTC

R topics documented:

- bird.env.coarse .......................... 2
**bird.env.coarse**

**Description**

Projected coordinates and environmental variables in 123, 100 x 100 km sites.

**Usage**

```r
data(bird.env.coarse)
```

**Format**

A data frame with 123 rows (sites) and 9 columns (xy coordinates and environmental variables).
Details

The data set contains the following variables:

- x: x-position in meters in UTM 53 South projection
- y: y-position in meters in UTM 53 South projection
- Natural: Proportion of area of conservation and natural environments
- Irrigated: Proportion of area of production from irrigated agriculture and plantations
- Water: Proportion of area of water features
- Elevation: Elevation
- ApP: Area per person
- Temp: Temperature
- Precip: Precipitation

Location: Australia – 51° 27’ 2.27” S, 135° 21’ 35.19” E
Data owners: ABARES, Australian Bureau of Statistics, GEBCO, WorldClim

References

http://data.daff.gov.au/anrdl/metadata_files/pa_luav4g9abl07811a00.xml
http://www.gebco.net/
http://www.worldclim.org/

Details

The data set contains the following variables:

- \( x \): x-position in meters in UTM 53 South projection
- \( y \): y-position in meters in UTM 53 South projection
- Natural: Proportion of area of conservation and natural environments
- Irrigated: Proportion of area of production from irrigated agriculture and plantations
- Water: Proportion of area of water features
- Elevation: Elevation
- ApP: Area per person
- Temp: Temperature
- Precip: Precipitation

Location: Australia – 50° 33' 5.03" S, 135° 21’ 10.40” E

Data owners: ABARES, Australian Bureau of Statistics, GEBCO, WorldClim

References

http://data.daff.gov.au/anrdl/metadata_files/pa_luav4g9abl07811a00.xml
http://www.gebco.net/
http://www.worldclim.org/


Description

Inventory of bird species occurrence in 123, 100 x 100 km sites.

Usage

data(bird.spec.coarse)

Format

A data frame with 123 rows (sites) and 193 columns (xy coordinates and species).
Details

• x: x-position in meters in UTM 53 South projection
• y: y-position in meters in UTM 53 South projection
• columns 3-193: bird species occurrence

The original bird occurrence data were arranged into a continuous grid covering South-East Australia. Only cells whose richness was within 10 percents of real estimated richness are included here, so that the data corresponds to presence-absence data.

Location: Australia – 51° 27' 2.27" S, 135° 21' 35.19" E
Data owner: BirdLife Australia

References

glm.cons is an adaptation of function glm2 from package {glm2} in which the least squares estimation is replaced by a regression with signs constraint on the coefficients using function nnpcls from package {nnls}.

Usage

```r
glm.cons(
  formula,
  family = stats::gaussian(),
  data,
  weights,
  subset,
  na.action,
  start = NULL,
  etastart,
  mustart,
  offset,
  control = list(...),
  model = TRUE,
  method = "glm.fit.cons",
  cons = -1,
  cons.inter = 1,
  x = FALSE,
  y = TRUE,
  contrasts = NULL,
  ...
)
```

Arguments

- `formula` as for `glm`
- `family` as for `glm`
- `data` as for `glm`
- `weights` as for `glm`
- `subset` as for `glm`
- `na.action` as for `glm`
- `start` as for `glm`
- `etastart` as for `glm`
- `mustart` as for `glm`
- `model` as for `glm`
- `method` as for `glm`
- `cons` as for `glm`
- `cons.inter` as for `glm`
- `x` as for `glm`
- `y` as for `glm`
- `contrasts` as for `glm`
glm.cons

offset as for glm
control as for glm
model as for glm
method the method used in fitting the model. The default method "glm.fit.cons" uses function nnmpls from package nnls instead of lm.fit to impose the sign of the coefficients. As in glm, the alternative method "model.frame" returns the model frame and does no fitting.
cons type of constraint. Default is -1 for negative coefficients on the predictors. The other option is 1 for positive coefficients on the predictors.
cons.inter type of constraint for the intercept. Default is 1 for positive intercept, suitable for Gaussian family. The other option is -1 for negative intercept, suitable for binomial family.
x as for glm
y as for glm
contrasts as for glm
... as for glm

Value
The value returned by glm.cons has exactly the same structure as the value returned by glm and glm.2.

References

See Also
glm, glm2

Examples
## Dobson (1990) Page 93: Randomized Controlled Trial:
counts <- c(18,17,15,20,10,20,25,13,12)
outcome <- gl(3,1,9)
treatment <- gl(3,3)
print(d.AD <- data.frame(treatment, outcome, counts))
glm.D93 <- glm.cons(counts ~ outcome + treatment, family = poisson())
glm.D93.ngl <- glm.cons(counts ~ outcome + treatment, family = poisson(),
method="glm.fit.cons")
summary(glm.D93)
summary(glm.D93.ngl)
glm.fit.cons

*Generalized Linear Models fitting method with negative coefficients constraint*

Description

`glm.fit.cons` is an adaptation of function `glm.fit2` from package `{glm2}` in which the least squares estimation is replaced by a non-positive regression using function `nnnpls` from package `{nnls}`.

Usage

```r
glm.fit.cons(
  x,
  y,
  weights = rep(1, nobs),
  cons = -1,
  cons.inter = 1,
  start = NULL,
  etastart = NULL,
  mustart = NULL,
  offset = rep(0, nobs),
  family = stats::gaussian(),
  control = list(),
  intercept = TRUE
)
```

Arguments

- **x**: as for `glm.fit`
- **y**: as for `glm.fit`
- **weights**: as for `glm.fit`
- **cons**: type of constraint. Default is `-1` for negative coefficients on the predictors. The other option is `1` for positive coefficients on the predictors.
- **cons.inter**: type of constraint for the intercept. Default is `1` for positive intercept, suitable for Gaussian family. The other option is `-1` for negative intercept, suitable for binomial family.
- **start**: as for `glm.fit`
- **etastart**: as for `glm.fit`
- **mustart**: as for `glm.fit`
- **offset**: as for `glm.fit`
- **family**: as for `glm.fit`
- **control**: as for `glm.fit`
- **intercept**: as for `glm.fit`
Value

The value returned by glm.fit.cons has exactly the same structure as the value returned by glm.fit and glm.fit2.

References


See Also

glm.fit, glm.fit2

Examples

```r
## Dobson (1990) Page 93: Randomized Controlled Trial :
counts <- c(18,17,15,20,10,20,25,13,12)
outcome <- gl(3,1,9)
treatment <- gl(3,3)
print(d.AD <- data.frame(treatment, outcome, counts))
glm.D93 <- glm.cons(counts ~ outcome + treatment, family = poisson())
glm.D93.ngl <- glm.cons(counts ~ outcome + treatment, family = poisson(),
                       method="glm.fit.cons")
summary(glm.D93)
summary(glm.D93.ngl)
```

Ispline

Transform data using I-splines

Description

Evaluates the I-splines for all variables of a data frame, as performed in Zeta.msgdm.

Usage

Ispline(dat, order.ispline = 2, kn.ispline = 1, rescale = 0)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dat</td>
<td>A data frame whose columns are variables to be transformed using I-splines.</td>
</tr>
<tr>
<td>order.ispline</td>
<td>Order of the I-spline.</td>
</tr>
<tr>
<td>kn.ispline</td>
<td>Number of knots in the I-spline.</td>
</tr>
<tr>
<td>rescale</td>
<td>Indicates how to rescale the values between 0 and 1. Default is 0, which divides the data by the maximum value. Any other value corresponds to setting the minimum value to 0.</td>
</tr>
</tbody>
</table>
Marion.env

Value

Ispline returns a data frame with the same number of rows as dat and ncol(dat) * (order.ispline + kn.ispline) columns.

References


See Also

Zeta.msgdm

Examples

```r
utils::data(bird.env.coarse)
data.env <- bird.env.coarse[,3:9]
data.env.splines <- Ispline(data.env)
```

Description

Geographic coordinates, altitude and island side (East, West) at 12 plots (4 transects and 3 altitudes) on Marion Island.

Usage

data(Marion.env)

Format

A data frame with 12 rows (plots) and 4 columns (variables).

Details

The data set contains the following variables:

- x: x-position in meters in UTM 37 projection
- y: y-position in meters in UTM 37 projection
- Altitude: mean elevation
- Side: cardinal (East or West) side of the island

Location: Marion Island – 46° 53’ 34.2” S, 37 degrees 45’ 02.3” E
Data owner: Melodie A. McGeoch
References


---

**Marion.species**

**Marion Island Species Presence-Absence Dataset**

**Description**

Inventory of springtails and mite species presence-absence in 12 plots (4 transects and 3 altitudes) on Marion Island.

**Usage**

```r
data(Marion.species)
```

**Format**

A data frame with 12 rows (plots) and 33 columns (species).

**Details**

The data set contains the following variables:

- x: x-position in meters in UTM 37 South projection
- y: y-position in meters in UTM 37 South projection
- columns 3-24: mite species presence absence
- columns 25-33: springtail species presence absence

Location: Marion Island – 46° 53’ 34.2” S, 37 degrees 45’ 02.3” E

Data owner: Melodie A. McGeoch

**References**


Pie Chart, considering negative values as zeros

Description

Plots a pie chart, considering negative values as zeros, for the purpose of illustrating variation partitioning.

Usage

```r
description
Usage
Arguments
```

Arguments

- `x` A vector of non-negative numerical quantities. The values in `x` are displayed as the areas of pie slices.
- `labels` One or more expressions or character strings giving names for the slices. Other objects are coerced by `as.graphicsAnnot`. For empty or NA (after coercion to character) labels, no label nor pointing line is drawn.
- `edges` The circular outline of the pie is approximated by a polygon with this many edges.
- `radius` The pie is drawn centered in a square box whose sides range from -1 to 1. If the character strings labeling the slices are long it may be necessary to use a smaller radius.
- `clockwise` Logical indicating if slices are drawn clockwise or counter clockwise (i.e., mathematically positive direction, used by default).
- `init.angle` number specifying the starting angle (in degrees) for the slices. Defaults to 0 (i.e., '3 o’clock’) unless clockwise is true where init.angle defaults to 90 (degrees), (i.e., '12 o’clock’).
density

The density of shading lines, in lines per inch. The default value of NULL means that no shading lines are drawn. Non-positive values of density also inhibit the drawing of shading lines.

angle

The slope of shading lines, given as an angle in degrees (counter-clockwise).

col

A vector of colors to be used in filling or shading the slices. If missing a set of 6 pastel colours is used, unless density is specified when par("fg") is used.

border, lty

(possibly vectors) arguments passed to polygon which draws each slice.

main

An overall title for the plot.

warning

Boolean value. Set to FALSE to avoid displaying a warning if some values are negative and set to 0.

... 

Graphical parameters can be given as arguments to pie. They will affect the main title and labels only.

Details

This function is identical to the function pie in \{graphics\}, except that it considers all negative values as zeros, to allow for plotting variation partitioning outputs. The original pie function returns an error when negative values are present. However, variation partitioning can return negative values, which can then be treated as zeros (Legendre & Legendre, 2008). This function allows direct use of the results from Zeta.varpart without editing the data.

References


See Also

pie, Zeta.varpart

Examples

```r
pie.neg(rep(1, 24), col = rainbow(24), radius = 0.9)
```

**Plot.ispline**

Plots I-splines for Multi-Site Generalised Dissimilarity Modelling

Description

Plots I-splines computed by Return.ispline, or calls Return.ispline if the outputs from Zeta.msgdm are provided before plotting.
Usage

```r
Plot.ispline(
  isplines = NULL,
  msgdm,
  data.env,
  distance = FALSE,
  biotic = 0,
  pch = NULL,
  lty = NULL,
  legend = TRUE,
  lwd = 1,
  cex = 1,
  num.quantiles = 11
)
```

Arguments

- **isplines**: Output of function `Return.ispline`.
- **msgdm**: Output of function `Zeta.msgdm` computed with `reg.type = ispline`.
- **data.env**: Site-by-variable data frame used for the computation of `msgdm`, with sites as rows and environmental variables as columns.
- **distance**: Boolean, indicates is distance was used in the computation of `msgdm`.
- **biotic**: Boolean, indicates is zeta diversity from another community was used in the computation of `msgdm`.
- **pch**: Shapes of the points to be used in the plotting. If nothing is provided, `pch` is a sequence of integers from 1 to the number of variables used for the computation of `msgdm`.
- **lty**: Line types to be used in the plotting. If nothing is provided, `lty` is a sequence of integers from 1 to the number of variables used for the computation of `msgdm`.
- **legend**: Boolean, indicates if the legend must be drawn.
- **lwd**: Line width.
- **cex**: Point size.
- **num.quantiles**: Number of points to plot on the I-splines. Default is 11 to plot a point every 10 percents of the range of values.

Value

`Plot.ispline` returns a data frame with the same number of rows as `dat` and `ncol(dat) * (order.ispline + kn.ispline)` columns.

References


See Also

Zeta.msgdm

Examples

```r
utils::data(Marion.species)
xy.marion <- Marion.species[1:2]
data.spec.marion <- Marion.species[3:33]

utils::data(Marion.env)
data.env.marion <- Marion.env[3]

zeta.ispline <- Zeta.msgdm(data.spec.marion, data.env.marion, xy.marion, sam = 100,
                           order = 3, normalize = "Jaccard", reg.type = "ispline")
zeta.ispline
zeta.ispline.r <- Return.ispline(zeta.ispline, data.env.marion, distance = TRUE)
zeta.ispline.r

dev.new()
Plot.ispline(isplines = zeta.ispline.r, distance = TRUE)

dev.new()
Plot.ispline(msgdm = zeta.ispline, data.env = data.env.marion, distance = TRUE)
```

---

**Plot.zeta.ddecay**

**Zeta distance-decay plotting**

**Description**

Plots the output of the function Zeta.ddecay.

**Usage**

`Plot.zeta.ddecay(zeta.ddecay)`

**Arguments**

- `zeta.ddecay`: A list produced by the function Zeta.ddecay.

**Value**

A plot of the zeta distance-decay with distance on the x-axis and the value of zeta on the y-axis.

**References**

See Also

Zeta.decline.mc, Zeta.order.mc, Zeta.decline.ex, Zeta.order.ex, Zeta.ddecay.

Zeta.ddecays

Examples

```r
utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[,1:2]
data.spec.bird <- bird.spec.coarse[,3:193]

dev.new()
zeta.ddecay.bird <- Zeta.ddecay(xy.bird, data.spec.bird, sam = 100, order = 3,
confint.level = 0.95, plot=FALSE)
Plot.zeta.ddecay(zeta.ddecay.bird)

##########

utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]

zeta.ddecay.marion <- Zeta.ddecay(xy.marion, data.spec.marion, sam = 100, order = 3,
confint.level = 0.95, trsf = "log", normalize = "Jaccard", plot=FALSE)
dev.new()
Plot.zeta.ddecay(zeta.ddecay.marion)
```

Description

Plots the output of the function Zeta.ddecays.

Usage

```r
Plot.zeta.ddecays(zeta.ddecays)
```

Arguments

- `zeta.ddecays` A list produced by the function Zeta.ddecays.

Value

A plot of the zeta distance-decay with the orders on the x-axis and the slope of the linear distance-decays on the y-axis.
References

See Also
Zeta.decline.mc, Zeta.order.mc, Zeta.decline.ex, Zeta.order.ex, Zeta.ddecays, Zeta.ddecay, Plot.zeta.ddecay

Examples
```r
utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[,1:2]
data.spec.bird <- bird.spec.coarse[,3:193]

dev.new()
zeta.ddecays.bird <- Zeta.ddecays(xy.bird, data.spec.bird, sam = 100, orders = 2:5, plot = FALSE, confint.level = 0.95)
Plot.zeta.ddecays(zeta.ddecays.bird)

utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]

dev.new()
zeta.ddecays.marion <- Zeta.ddecays(xy.marion, data.spec.marion, sam = 100, orders = 2:5, plot = FALSE, confint.level = 0.95)
Plot.zeta.ddecays(zeta.ddecays.marion)
```

---

**Plot.zeta.decline**

*Zeta diversity decline plotting*

**Description**
Plots the output of the functions Zeta.decline.mc and Zeta.decline.ex.

**Usage**
```r
Plot.zeta.decline(zeta, sd.plot = TRUE, arrange.plots = TRUE)
```

**Arguments**
- *zeta*: A list produced by the function Zeta.decline.mc or Zeta.decline.ex.
- *sd.plot*: Boolean value (TRUE or FALSE) indicating if the standard deviation of each zeta diversity value must be plotted.
- *arrange.plots*: Boolean value (TRUE or FALSE) indicating if the graphics device must be divided into 4 subplots.
Value

A plot of the zeta decline with 4 subplots displaying (i) the raw decline, (ii) the ratios of the zeta values (computed as $\zeta_i/\zeta_{i-1}$), (iii) the fit in a log plot and (iv) the fit in a log-log plot.

References


See Also

Zeta.decline.mc, Zeta.order.mc, Zeta.decline.ex, Zeta.order.ex

Examples

```r
utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[1:2]
data.spec.bird <- bird.spec.coarse[3:193]

dev.new(width = 12, height = 4)
zeta.bird <- Zeta.decline.mc(data.spec.bird, orders = 1:5, sam=100, plot = FALSE)
Plot.zeta.decline(zeta.bird)

utils::data(Marion.species)
xy.marion <- Marion.species[1:2]
data.spec.marion <- Marion.species[3:33]

dev.new(width = 12, height = 4)
zeta.marion <- Zeta.decline.ex(data.spec.marion, orders = 1:5, plot = FALSE)
Plot.zeta.decline(zeta.marion)
```

Description

Plotting of zeta diversity scaling with sample grain dependency based on the minimum distance between sites.

Plots the output of the function Zeta.scale.min.dist.
Usage

\[
\text{Plot.zeta.scale.min.dist(}
\begin{array}{ll}
\text{zeta.scale.irreg,} \\
\text{size.init = 1,} \\
\text{add = FALSE,} \\
\text{ylim = NULL,} \\
\text{col = "black"}
\end{array}
\)
\]

Arguments

- **zeta.scale.irreg**: A list generated by the function `Zeta.scale.min.dist`.
- **size.init**: Initial size of the plots before aggregation.
- **add**: Boolean value indicating if the graph must be plotted in a new graphics device or added to the active one.
- **ylim**: Numeric vectors of length 2, giving the range of y values.
- **col**: String indicating the color of the graph.

Value

A plot of the zeta diversity scaling with the mapping grain \(m\) (the number of sites combined to generate data at a coarser grain) on the x-axis and the value of zeta on the y-axis.

References


See Also

- `Zeta.decline.mc`, `Zeta.order.mc`, `Zeta.decline.ex`, `Zeta.order.ex`
- `Zeta.scale.min.dist`, `rescale.regular`, `Zeta.scale.regular`, `rescale.regular`
- `Plot.zeta.scale.regular`

Examples

```r
utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]

zeta.scale.irreg.species <- Zeta.scale.min.dist(xy.marion, data.spec.marion, m = 1:3, order = 3, reorder = 3, sam = 50, normalize = "Jaccard", plot=FALSE)
dev.new()
Plot.zeta.scale.min.dist(zeta.scale.irreg.species)
```
**Plot.zeta.scale.regular**

*Plotting of zeta diversity scaling with sample grain using hierarchical increases in grain size*

**Description**

Plots the output of the function `Zeta.scale.regular`.

**Usage**

```r
Plot.zeta.scale.regular(
  zeta.scale.reg,
  size.init = 1,
  add = FALSE,
  ylim = NULL,
  col = "black"
)
```

**Arguments**

- `zeta.scale.reg`: A list generated by the function `Zeta.scale.regular`.
- `size.init`: initial Size of the plots before aggregation.
- `add`: Boolean value indicating if the graph must be plotted in a new graphics device or added to the active one.
- `ylim`: Numeric vectors of length 2, giving the range of y values.
- `col`: String indicating the color of the graph.

**Value**

A plot of the zeta diversity scaling with the mapping grain $n \times n$ (the number of sites combined to generate data at a coarser grain) on the x-axis and the value of zeta on the y-axis.

**References**


**See Also**

- `Zeta.decline.mc`
- `Zeta.order.mc`
- `Zeta.decline.ex`
- `Zeta.order.ex`
- `Zeta.scale.regular`
- `Zeta.scale.min.dist`
- `rescale.regular`
- `Plot.zeta.scale.min.dist`
Examples

```r
utils::data(bird.spec.fine)
xy.bird <- bird.spec.fine[1:400,1:2]
data.spec.bird <- bird.spec.fine[1:400,3:192]

## sam = 25 is used here for fast execution, but a higher value is advised
zeta.scale.reg <- Zeta.scale.regular(xy.bird, data.spec.bird, n = 1:3, order = 3,
   sam = 25, normalize = "Jaccard", plot=FALSE)
dev.new()
Plot.zeta.scale.regular(zeta.scale.reg)
```

Description

Predict the zeta values for new environmental and distance data from the models returned by `Zeta.msgdm`.

Usage

```r
Predict.msgdm(model = model.msgdm, reg.type, newdata, type = "response")
```

Arguments

- `model.msgdm`: A model returned by `Zeta.msgdm`. The class of the model depends on the type of regression used in `Zeta.msgdm`.
- `reg.type`: Type of regression used in `Zeta.msgdm`. Options are "glm" for generalised linear models, "ngls" for negative linear models, "gam" for generalised additive models, "scam" for shape constrained additive models (with monotonic decreasing by default), and "ispline" for I-spline models (forcing monotonic decreasing), as recommended in generalised dissimilarity modelling by Ferrier et al. (2007).
- `newdata`: A data frame with the new environmental and distance data. The names of the columns must be the same as the names used in the data frame used in `Zeta.msgdm`. For I-splines, the data frame must be generated beforehand from the original data by the function `Ispline`.
- `type`: The type of prediction required, as for `predict.glm`. The default is on the scale of the response variable; the alternative "link" is on the scale of the linear predictors.

Value

`Predict.msgdm` returns a vector of predicted zeta values.
References


See Also

Zeta.msgdm

Examples

```r
utils::data(bird.spec.fine)
xy.bird <- bird.spec.fine[1:500,1:2]
data.spec.bird <- bird.spec.fine[1:500,3:192]
utils::data(bird.env.fine)
data.env.bird <- bird.env.fine[1:500,3:9]

zeta.glm <- Zeta.msgdm(data.spec.bird, data.env.bird, sam = 100, order = 3)
newdata <- data.frame(matrix(NA,100,ncol(data.env.bird)))
names(newdata) <- names(data.env.bird)
for(z in 1:100){
  samp <- sample(1:104, 3, replace = FALSE)
  newdata[z,] <- apply(apply(bird.env.fine[501:604,3:9][samp,], 2,
    stats::dist), 2, mean)
}
# rescale the data like during MS-GDM
newdata <- newdata/matrix(rep(zeta.glm$rescale.factor,100),
  100,length(zeta.glm$rescale.factor),byrow=TRUE)
new.zeta.glm <- Predict.msgdm(model.msgdm = zeta.glm$model, reg.type = "glm",
  newdata = newdata)

zeta.ngls <- Zeta.msgdm(data.spec.bird, data.env.bird, sam = 100, order = 3,
  reg.type = "ngls", normalize = "Jaccard")
newdata <- data.frame(matrix(NA,100,ncol(data.env.bird)))
names(newdata) <- names(data.env.bird)
for(z in 1:100){
  samp <- sample(1:104, 3, replace = FALSE)
  newdata[z,] <- apply(apply(bird.env.fine[501:604,3:9][samp,], 2, stats::dist),
    2, mean)
}
# rescale the data like during MS-GDM
newdata <- newdata/matrix(rep(zeta.ngls$rescale.factor,100),
  100,length(zeta.ngls$rescale.factor),byrow=TRUE)
new.zeta.ngls <- Predict.msgdm(model.msgdm = zeta.ngls$model, reg.type = "ngls",
  newdata = newdata)
```
Perform an I-spline regression

Description
Evaluates the I-splines for all variables of a data frame of predictor variables, and perform a generalised linear regression with constraint on the parameters.

Usage
Reg.ispline(
  response,
  predictor,
  order.ispline = 2,
  kn.ispline = 1,
  family = stats::gaussian(),
  method.glm = "glm.fit.cons",
  cons = 1,
  cons.inter = 1,
  control = list(),
  Plot = TRUE,
  lty = NULL,
  lwd = 1
)

Arguments
response A vector of numeric values representing the response variable.
predictor A data frame of numeric variables representing the predictors.
order.ispline Order of the I-spline.
kn.ispline Number of knots in the I-spline.
family A description of the error distribution and link function to be used in the glm model (see family for details of family functions).
method.glm Method used in fitting the generalised linear model. The default method "glm.fit.cons" is an adaptation of method glm.fit2 from package glm2 using a constrained least squares regression in the reweighted least squares. Another option is "glm.fit2", which calls glm.fit2; see help documentation for glm.fit2 in package glm2.
cons type of constraint in the glm if method.glm = "glm.fit.cons". Default is 1 for positive coefficients on the predictors. The other option is -1 for negative coefficients on the predictors.
cons.inter type of constraint for the intercept. Default is 1 for positive intercept, suitable for Gaussian family. The other option is -1 for negative intercept, suitable for binomial family.
control As for glm.
Reg.ispline

Plot  Boolean value indicating if the I-splines must be plotted.

lty  Line types to be used in the plotting. If nothing is provided, lty is a sequence of integers from 1 to the number of variables used for the computation of msgdm.

lwd  Line width.

Details

Reg.ispline performs a non-linear regression using a combination of GLM and I-splines. It can, for example, be used to compare regression outputs when using MS-GDM with I-splines on environmental variables and biotic variables as in Zetya.msgdm to the same regression approach without environmental variables.

Value

Reg.ispline returns a list of the following elements:

splines  A data frame in which each columns contains the value resulting from the transformation of the predictors into individual I-splines. The number of columns of splines is the number of predictors times the number of splines (determined as the sum of order.ispline and kn.ispline).

spline  A data frame in which each columns contains the value resulting from the combinations of the individual I-splines. This combination is obtained by multiplying the coefficients of model and the values of the individual I-splines splines.

model  A glm model using response as the response variable, and splines as the predictors.

References


See Also

Zeta.msgdm,Ispline

Examples

```r
utils::data(Marion.species)
xy.marion <- Marion.species[1:2]
data.spec.marion <- Marion.species[3:33]

##random other communities
data.spec.marion2a <- data.spec.marion
data.spec.marion2a[which(data.spec.marion2a==1,arr.ind=TRUE)] <- 0
for(i in 1:ncol(data.spec.marion2a))
```
data.spec.marion2a[sample(nrow(data.spec.marion2a),8),i] <- 1
data.spec.marion2b <- data.spec.marion
data.spec.marion2b[which(data.spec.marion2b==1,arr.ind=TRUE)] <- 0
for(i in 1:ncol(data.spec.marion2b))
data.spec.marion2b[sample(nrow(data.spec.marion2b),8),i] <- 1

dat.spec.tot <- list(data.spec.marion,data.spec.marion2a,data.spec.marion2b)
zeta.tot <- Zeta.order.mc.mult(data.spec=dat.spec.tot,order=3,sam=200)
zeta.splines <- Ispline(zeta.tot$zeta.val[,2:3])
data.tot <- data.frame(zeta.val=zeta.tot$zeta.val[,1],zeta.splines$splines)

dev.new()
Reg.ispline(response = zeta.tot$zeta.val[,1], predictor = zeta.tot$zeta.val[,2:3], lwd=2, cons=1)

---

rescale.min.dist  
**Rescaling of data based on the minimum distance between sites**

Description

Combines sites based on the minimum distance between them.

Usage

```r
rescale.min.dist(
  xy, 
  data.spec, 
  data.env = NULL, 
  m, 
  distance.type = "Euclidean", 
  dist.custom = NULL, 
  method = "mean", 
  shuffle = FALSE
)
```

Arguments

- **xy**  
  Site-by-coordinate data frame, with sites as rows and coordinates as columns.
- **data.spec**  
  Site-by-species presence-absence data frame, with sites as rows and species as columns.
- **data.env**  
  Site-by-variable data frame, with sites as rows and environmental variables as columns.
- **m**  
  Mapping grain (the number of sites combined to generate data at a coarser grain). The m closest sites are grouped together.
- **distance.type**  
  Method to compute distance. Default is "Euclidean", for Euclidean distance. The other options are (i) "ortho" for orthodromic distance, if xy correspond to longitudes and latitudes (orthodromic distance is computed using the geodist function from package geodist); and (ii) "custom", in which case the user must provide a distance matrix for dist.custom.
**rescale.min.dist**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dist.custom</td>
<td>Distance matrix provided by the user when distance.type = &quot;custom&quot;.</td>
</tr>
<tr>
<td>method</td>
<td>Name of a function (as a string) indicating how to combine the coordinates and the environmental variables. It can be a basic R-function such as &quot;mean&quot; or &quot;max&quot;, but also a custom function.</td>
</tr>
<tr>
<td>shuffle</td>
<td>Boolean value (TRUE or FALSE) indicating if the order of the sites must be randomised, which can have an impact on the outputs if some distances are equal.</td>
</tr>
</tbody>
</table>

**Details**

The nearest neighbouring sites (plots, quadrates, or areas of varying shapes) are grouped as spatial clusters of 2, 3, 4, etc. sites, based on the minimum distance between them. Since the procedure is based on the relative distance between sites, the site order can have an impact on the output. This function is suitable for both regularly and irregularly spaced sites, contiguous or non contiguous. For regularly spaced sites, the use of `rescale.regular` is recommended.

**Value**

rescale.min.dist returns a data frame with the rescaled data.

**References**


**See Also**

`Zeta.decline.mc`, `Zeta.order.mc`, `Zeta.decline.ex`, `Zeta.order.ex`, `Zeta.scale.min.dist`, `Zeta.scale.regular`, `rescale.regular`

**Examples**

```r
utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]
data.rescale <- rescale.min.dist(xy.marion, data.spec.marion, m=2)
```
**Description**

Increases grain by hierarchically nesting regularly spaced sites.

**Usage**

```r
rescale.regular(xy, data.spec, data.env = NULL, method = "mean", n)
```

**Arguments**

- `xy`: Site-by-coordinate data frame, with sites as rows and coordinates as columns.
- `data.spec`: Site-by-species presence-absence data frame, with sites as rows and species as columns.
- `data.env`: Site-by-variable data frame, with sites as rows and environmental variables as columns.
- `method`: Name of a function (as a string) indicating how to combine the coordinates and the environmental variables. It can be a basic R-function such as "mean" or "max", but also a custom function.
- `n`: Mapping grain (the number of sites combined to generate data at a coarser grain). Regularly spaced sites are grouped as n x n sites.

**Details**

The sites (plots or quadrates) are aggregated as nearest neighbouring groups of n x n sites, using a nested approach, starting from the lowest x and y, to increase the grain. The sites can be spatially contiguous or discontiguous, as long as they are regularly spaced. This function is not suitable for irregularly spaced sites. If the total number of sites is not a multiple of n x n, the extra sites are discarded.

**Value**

`rescale.regular` returns a data frame with the rescaled data.

**References**


Return.ispline

Computing splines coordinates from I-spline-based multi-site generalised dissimilarity modelling

Description
Stores the coordinates of the I-splines resulting from Zeta.msgdm for plotting.

Usage
Return.ispline(msgdm, data.env, distance = FALSE, biotic = 0)

Arguments
- msgdm: Output of function Zeta.msgdm computed with reg.type = ispline.
- data.env: Site-by-variable data frame used for the computation of msgdm, with sites as rows and environmental variables as columns.
- distance: Boolean, indicates is distance was used in the computation of msgdm.
- biotic: Integer, indicates the number of other groups of taxa for which zeta diversity was computed and used in the computation of msgdm.

Details
Return.ispline allows to store the same number of coordinates for all I-splines, to average replicates and obtain confidence intervals.

Value
Return.ispline returns a list containing the following components used to plot the I-splines:
- env: A data frame containing the rescaled environmental (numeric and factor), distance and biotic x-values.
- Ispline: A data frame containing the I-spline values corresponding to the rescaled environmental (numeric and factor), distance and biotic x-values.
**Zeta.ddecay**

**See Also**

*Zeta.msgdm, Ispline*

**Examples**

```r
utils::data(Marion.species)
xy.marion <- Marion.species[1:2]
data.spec.marion <- Marion.species[3:33]

utils::data(Marion.env)
data.env.marion <- Marion.env[3]

zeta.ispline <- Zeta.msgdm(data.spec.marion, data.env.marion, xy.marion, sam = 100,
    order = 3, normalize = "Jaccard", reg.type = "ispline")
zeta.ispline
zeta.ispline.r <- Return.ispline(zeta.ispline, data.env.marion, distance = TRUE)
zeta.ispline.r

dev.new()
Plot.ispline(isplines = zeta.ispline.r, distance = TRUE)

dev.new()
Plot.ispline(msgdm = zeta.ispline, data.env = data.env.marion, distance = TRUE)
```

---

**Zeta.ddecay**

*Zeta distance decay for a specific number of assemblages or sites*

**Description**

Computes the distance decay of zeta diversity for a specific order (number of assemblages or sites), using either a generalised linear model with possible constraint on the coefficients, a generalised additive model, or a shape constrained additive model.

**Usage**

```r
Zeta.ddecay(xy,
data.spec,order = 2,sam = 1000,distance.type = "Euclidean",dist.custom = NULL,method = "mean",reg.type = "glm",family = stats::gaussian(),method.glm = "glm.fit.cons",
```
cons = -1,
cons.inter = 1,
confint.level = 0.95,
kn = -1,
bs = "mpd",
trsf = "NULL",
cutoff = NULL,
rescale = FALSE,
normalize = FALSE,
empty.row = "remove",
plot = TRUE
}

Arguments

xy
Site-by-coordinate data frame, with sites as rows and coordinates as columns.
data.spec
Site-by-species presence-absence data frame, with sites as rows and species as columns.
order
Specific number of assemblages or sites at which zeta diversity is computed.
sam
Number of samples for which the zeta diversity is computed.
distance.type
Method to compute distance. Default is "Euclidean", for Euclidean distance. The other options are (i) "ortho" for orthodromic distance, if xy correspond to longitudes and latitudes (orthodromic distance is computed using the geodist function from package geodist); and (ii) "custom", in which case the user must provide a distance matrix for dist.custom.
dist.custom
Distance matrix provided by the user when distance.type = "custom".
method
Name of a function (as a string) indicating how to combine the pairwise differences and distances for more than 3 sites. It can be a basic R-function such as "mean" or "max", but also a custom function.
reg.type
Type of regression. Options are "glm" for generalised linear models "gam" for generalised additive models and "scam" for shape constrained additive models (with monotonic decreasing by default).
family
A description of the error distribution and link function to be used in the glm, gam and scam models (see family for details of family functions).
method.glm
Method used in fitting the generalised linear model. The default method "glm.fit.cons" is an adaptation of method glm.fit2 from package glm2 using a negative least squares regression in the reweighted least squares. Another option is "glm.fit2", which calls glm.fit2.; see help documentation for glm.fit2 in package glm.
cons
type of constraint in the glm if method.glm = "glm.fit.cons". Default is -1 for negative coefficients on the predictors. The other option is 1 for positive coefficients on the predictors.
cons.inter
type of constraint for the intercept. Default is 1 for positive intercept, suitable for Gaussian family. The other option is -1 for negative intercept, suitable for binomial family.
confint.level Percentage for the confidence intervals of the coefficients from the generalised linear models.

kn Number of knots in the GAM and SCAM. Default is -1 for determining kn automatically using Generalized Cross-validation.

bs A two-letter character string indicating the (penalized) smoothing basis to use in the scam model. Default is "mpd" for monotonic decreasing splines. see smooth.terms for an overview of what is available.

trsf Name of a function (as a string) indicating how to transform distance.

cutoff If specified, maximum distance value for which the linear regression must be performed.

rescale Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by $\zeta_1$, to get a range of values between 0 and 1. Has no effect if normalize != FALSE.

normalize Indicates if the zeta values for each sample should be divided by the total number of species for this specific sample (normalize = "Jaccard"), by the average number of species per site for this specific sample (normalize = "Sorensen"), or by the minimum number of species in the sites of this specific sample (normalize = "Simpson"). Default value is FALSE, indicating that no normalization is performed.

empty.row Determines how to handle empty rows, i.e. sites with no species. Such sites can cause underestimations of zeta diversity, and computation errors for the normalized version of zeta due to divisions by 0. Options are "empty" to let the data untreated, "remove" to remove the empty rows, 0 to set the normalized zeta to 0 when zeta is divided by 0 during normalization (sites share no species, so are completely dissimilar), and 1 to set the normalized zeta to 1 when zeta is divided by 0 during normalization (i.e. sites are perfectly similar).

plot Boolean value (TRUE or FALSE) indicating if the outputs must be plotted.

Value

Zeta.ddecay returns a list containing the following components:

- order The order of zeta for which the distance decay was computed.
- reg.type A character string indicating the type of regression that was performed.
- reg An object whose class depends on the type of regression (glm, gam or scam), corresponding to the regression over distance for the number of assemblages or sites specified in 'order'.
- confint The confidence intervals for the coefficients from the generalised linear model. confint is not generated for generalised additive models and shape constrained additive models.
- zeta.val The values of zeta for the sampled sites used in the regression.
- distance The distances for the sampled sites used in the regression.

References

Zeta.ddecays

See Also

Zeta.decline.mc, Zeta.order.mc, Zeta.decline.ex, Zeta.order.ex, Zeta.ddecays,
Plot.zeta.ddecay

Examples

utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[,1:2]
data.spec.bird <- bird.spec.coarse[,3:193]

dev.new()
zeta.ddecay.bird <- Zeta.ddecay(xy.bird, data.spec.bird, sam = 100, order = 3,
   method.glm = "glm.fit2", confint.level = 0.95)

dev.new()
zeta.ddecay.bird <- Zeta.ddecay(data.spec=data.spec.bird, distance.type = "custom",
   dist.custom = as.matrix(dist(xy.bird)), cutoff = 800000, sam = 100, order = 3,
   reg.type = "gam", confint.level = 0.95)

############

utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]

dev.new()
zeta.ddecay.marion <- Zeta.ddecay(xy.marion, data.spec.marion, sam = 100, order = 3,
   method.glm = "glm.fit2", confint.level = 0.95, trsf = "log", normalize = "Jaccard")

Zeta.ddecays

Zeta distance decay for a range of numbers of assemblages or sites

Description

Computes the distance decay of zeta diversity for a range of orders (number of assemblages or sites), using generalised linear models.

Usage

Zeta.ddecays(
   xy,
   data.spec,
   orders = 2:10,
   sam = 1000,
   family = stats::gaussian(),
   distance.type = "Euclidean",
   dist.custom = NULL,
method = "mean",
confint.level = 0.95,
trsf = "NULL",
cutoff = NULL,
rescale = FALSE,
normalize = FALSE,
plot = TRUE
)

Arguments

xy
Site-by-coordinate data frame, with sites as rows and coordinates as columns.
data.spec
Site-by-species presence-absence data frame, with sites as rows and species as columns.
orders
Range of number of assemblages or sites at which zeta diversity is computed. All the orders must be strictly greater than 1.
sam
Number of samples for which the zeta diversity is computed.
family
A description of the error distribution and link function to be used in the generalised linear models (see family for details of family functions).
distance.type
Method to compute distance. Default is "Euclidean", for Euclidean distance. The other options are (i) "ortho" for orthodromic distance, if xy correspond to longitudes and latitudes (orthodromic distance is computed using the geodist function from package geodist); and (ii) "custom", in which case the user must provide a distance matrix for dist.custom.
dist.custom
Distance matrix provided by the user when distance.type = "custom".
method
Name of a function (as a string) indicating how to combine the pairwise differences and distances for more than 3 sites. It can be a basic R-function such as "mean" or "max", but also a custom function.
confint.level
Percentage for the confidence intervals of the coefficients from the linear regression.
trsf
Name of a function (as a string) indicating how to transform distance. Default is "NULL" for the identity transformation.
cutoff
If specified, maximum distance value for which the linear regression must be performed.
rescale
Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by $\zeta_1$, to get a range of values between 0 and 1. Has no effect if normalize = FALSE.
normalize
Indicates if the zeta values for each sample should be divided by the total number of species for this specific sample (normalize = "Jaccard"), by the average number of species per site for this specific sample (normalize = "Sorensen"), or by the minimum number of species in the sites of this specific sample (normalize = "Simpson"). Default value is FALSE, indicating that no normalization is performed.
plot
Boolean value (TRUE or FALSE) indicating if the outputs must be plotted.
Value

Zeta.ddecays returns a list containing the following components:

- **orders**: Range of number of assemblages or sites at which zeta diversity was computed.
- **coefs**: A vector of the coefficients from the generalised linear models for the numbers of sites specified by orders.
- **confint**: The confidence intervals for the coefficients from the generalised linear models.

References


See Also

- Zeta.decline.mc, Zeta.order.mc, Zeta.decline.ex, Zeta.order.ex, Zeta.ddecay

Examples

```r
utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[,1:2]
data.spec.bird <- bird.spec.coarse[,3:193]

dev.new()
zeta.ddecays.bird <- Zeta.ddecays(xy.bird, data.spec.bird, sam = 100, orders = 2:5, plot = TRUE, confint.level = 0.95)

################

utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]

dev.new()
zeta.ddecays.marion <- Zeta.ddecays(xy.marion, data.spec.marion, sam = 100, orders = 2:5, plot = TRUE, confint.level = 0.95)
```

### Zeta.decline.ex

**Expectation of zeta diversity decline**

Description

Computes the expectation of zeta diversity, the number of species shared by multiple assemblages for a range of orders (number of assemblages or sites), using a formula based on the occupancy of the species, and fits the decline to an exponential and a power law relationship.
Usage

Zeta.decline.ex(
  data.spec,
  orders = 1:10,
  sd.correct = TRUE,
  confint.level = 0.95,
  sd.plot = TRUE,
  rescale = FALSE,
  empty.row = "empty",
  plot = TRUE
)

Arguments

- **data.spec**: Site-by-species presence-absence data frame, with sites as rows and species as columns.
- **orders**: Range of number of assemblages or sites for which zeta diversity is computed.
- **sd.correct**: Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) or not (using the number of site combinations as the denominator).
- **confint.level**: Percentage for the confidence intervals of the coefficients from the regressions.
- **sd.plot**: Boolean value (TRUE or FALSE) indicating if the standard deviation of each zeta diversity value must be plotted.
- **rescale**: Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by \( \zeta_1 \), to get a range of values between 0 and 1.
- **empty.row**: Determines how to handle empty rows, i.e. sites with no species. Such sites can cause underestimations of zeta diversity. Options are "empty" to let the data untreated or "remove" to remove the empty rows.
- **plot**: Boolean value (TRUE or FALSE) indicating if the outputs must be plotted.

Details

Zeta.decline.ex is much faster than Zeta.decline.mc to compute the exact value of zeta diversity when the number of species is lower than \( C^N_i \), where \( N \) is the total number of sites and \( i \) is the order of zeta.

sd.correct should be set to TRUE if the assemblages represent a subsample of the whole system. It can be set to FALSE if the sampling is exhaustive, for example in case of a continuous regular grid covering the whole study area.

The exponential and the power law fit are performed using linear regressions on log-transformed data (only the zeta values are log-transformed for the exponential fit, and both the orders and the zeta values are log-transformed for the power law fit).
Value

Zeta.decline.ex returns a list containing the following components:

- **zeta.order**: The number of assemblages or sites for which the zeta diversity was computed.
- **combinations**: The number of possible combinations of sites for the chosen orders.
- **zeta.val**: The zeta diversity values.
- **zeta.val.sd**: The zeta diversity standard deviation values.
- **zeta.ratio**: The ratio of zeta diversity values by the zeta diversity values at the lower order \( \zeta_i / \zeta_{i-1} \).
- **zeta.exp**: Object of class "lm", containing the output of the exponential regression.
- **zeta.exp.confint**: The confidence intervals of the coefficients of the exponential regression.
- **zeta.pl**: Object of class "lm", containing the output of the power law regression.
- **zeta.pl.confint**: The confidence intervals of the coefficients of the power law regression.
- **aic**: AIC values for zeta.exp and zeta.pl.

References


See Also

Zeta.decline.mc, Zeta.order.mc, Zeta.order.ex, Plot.zeta.decline

Examples

```r
utils::data(bird.spec.coarse)
data.spec.bird <- bird.spec.coarse[,3:193]

dev.new(width = 12, height = 4)
zeta.bird <- Zeta.decline.ex(data.spec.bird, orders = 1:5)
zeta.bird

############

utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]

dev.new(width = 12, height = 4)
zeta.marion <- Zeta.decline.ex(data.spec.marion, orders = 1:5)
zeta.marion
```
Zeta.decline.mc

Zeta diversity decline using Monte Carlo sampling

Description

Computes zeta diversity, the number of species shared by multiple assemblages, for a range of orders (number of assemblages or sites), using combinations of sampled sites, and fits the decline to an exponential and a power law relationship.

Usage

Zeta.decline.mc(
  data.spec,
  xy = NULL,
  orders = 1:10,
  sam = 1000,
  sd.correct = TRUE,
  sd.correct.adapt = FALSE,
  confint.level = 0.95,
  sd.plot = TRUE,
  rescale = FALSE,
  normalize = FALSE,
  NON = FALSE,
  FPO = NULL,
  DIR = FALSE,
  empty.row = "empty",
  plot = TRUE,
  silent = TRUE
)

Arguments

data.spec  Site-by-species presence-absence data frame, with sites as rows and species as columns.
xy          Site coordinates. This is only used if NON = TRUE or DIR = TRUE.
orders      Range of number of assemblages or sites for which zeta diversity is computed.
sam         Number of samples for which the zeta diversity is computed for each number of assemblages or sites.
sd.correct  Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) or not (using the number of site combinations as the denominator).
sd.correct.adapt

Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) if sam is higher than the number of possible combinations,
38

or not (using the number of site combinations as the denominator) if sam is lower than the number of possible combinations. If sd.correct.adapt = TRUE, it takes precedence over sd.correct.

cfint.level Percentage for the confidence intervals of the coefficients from the regressions.

sd.plot Boolean value (TRUE or FALSE) indicating if the standard deviation of each zeta diversity value must be plotted.

rescale Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by \( \zeta_1 \), to get a range of values between 0 and 1. Has no effect if normalize \(!=\) FALSE.

normalize Indicates if the zeta values for each sample should be divided by the total number of species for this specific sample (normalize = "Jaccard"), by the average number of species per site for this specific sample (normalize = "Sorensen"), or by the minimum number of species in the sites of this specific sample (normalize = "Simpson"). Default value is FALSE, indicating that no normalization is performed.

NON Boolean value (TRUE or FALSE) indicating if the number of species in common should only be counted for the nearest neighbours.

FPO A vector with the coordinates of the fixed point origin from which the zeta diversity will be computed (overrides NON). In that case, \( \zeta_1 \) is the number of species in the closest site to the FPO, \( \zeta_2 \) is the number of species shared by the 2 closest sites, etc.

DIR Boolean value (TRUE or FALSE) indicating if zeta diversity must be computed using a directed nearest neighbour scheme in the direction away from the FPO, starting from any site.

empty.row Determines how to handle empty rows, i.e. sites with no species. Such sites can cause underestimations of zeta diversity, and computation errors for the normalized version of zeta due to divisions by 0. Options are "empty" to let the data untreated, "remove" to remove the empty rows, 0 to set the normalized zeta to 0 when zeta is divided by 0 during normalization (sites share no species, so are completely dissimilar), and 1 to set the normalized zeta to 1 when zeta is divided by 0 during normalization (i.e. sites are perfectly similar).

plot Boolean value (TRUE or FALSE) indicating if the outputs must be plotted.

silent Boolean value (TRUE or FALSE) indicating if messages must be printed.

Details

If the number of combinations of sites is lower than the value of the parameter sam, all the combinations are used and an exact solution is computed. In that case, using the number of site combinations as the denominator may be appropriate to compute the standard deviation, if all sites were sampled and the zeta values. This can be adjusted with parameters sd.correct and sd.correct.adapt.

Zeta.decline.mc is faster than Zeta.decline.ex to compute the exact value of zeta diversity when the number of species is higher than \( C_i^N \), where \( N \) is the total number of sites and \( i \) is the order of zeta.

The exponential and the power law fit are performed using linear regressions on log-transformed data (only the zeta values are log-transformed for the exponential fit, and both the orders and the zeta values are log-transformed for the power law fit).
Zeta.decline.mc enables accommodating richness heterogeneity by setting normalize = "Jaccard", normalize = "Sorensen" or normalize = "Simpson". This cannot be performed by Zeta.decline.ex.

**Value**

Zeta.decline.mc returns a list containing the following components:

- **zeta.order**: The number of assemblages or sites for which the zeta diversity was computed.
- **combinations**: The number of possible combinations of sites for the chosen orders.
- **zeta.val**: The zeta diversity values.
- **zeta.val.sd**: The zeta diversity standard deviation values.
- **zeta.ratio**: The ratio of zeta diversity values by the zeta diversity values at the lower order $\zeta_i/\zeta_i-1$.
- **zeta.exp**: Object of class "lm", containing the output of the exponential regression.
- **zeta.exp.confint**: The confidence intervals of the coefficients of the exponential regression.
- **zeta.pl**: Object of class "lm", containing the output of the power law regression.
- **zeta.pl.confint**: The confidence intervals of the coefficients of the power law regression.
- **aic**: AIC values for zeta.exp and zeta.pl.

**References**


**See Also**

Zeta.decline.ex, Zeta.order.ex, Zeta.order.mc, Plot.zeta.decline

**Examples**

```r
utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[,1:2]
data.spec.bird <- bird.spec.coarse[,3:193]

device.new(width = 12, height = 4)
zeta.bird <- Zeta.decline.mc(data.spec.bird, xy.bird, orders = 1:5, sam = 100, NON = TRUE)
zeta.bird

utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]
```
Zeta.msgdm

Multi-site generalised dissimilarity modelling for a set of environmental variables and distances

Description

Computes a regression model of zeta diversity for a given order (number of assemblages or sites) against a set of environmental variables and distances between sites. The different regression models available are generalised linear models, generalised linear models with negative constraints, generalised additive models, shape constrained additive models, and I-splines.

Usage

```
Zeta.msgdm(
  data.spec,  # data of species occurrences
  data.env,  # data of environmental variables
  xy = NULL,  # list of coordinates for species
  data.spec.pred = NULL,  # data of species occurrences to predict
  order = 1,  # order of sites or assemblages
  sam = 1000,  # number of random samples if resampling
  reg.type = "glm",  # type of regression model
  family = stats::gaussian(),  # family of the model
  method.glm = "glm.fit.cons",  # method for fitting the model
  cons = -1,  # constraint for the model
  cons.inter = 1,  # constraint for the intercept
  confint.level = 0.95,  # level of confidence interval
  bs = "mpd",  # basis function for I-splines
  kn = -1,  # knot for I-splines
  order.ispline = 2,  # order of I-splines
  kn.ispline = 1,  # knot of I-splines
  distance.type = "Euclidean",  # type of distance
  dist.custom = NULL,  # custom distance function
  rescale = FALSE,  # rescale data
  rescale.pred = TRUE,  # rescale predictions
  method = "mean",  # method for prediction
  normalize = FALSE,  # normalize data
  silent = FALSE,  # silent mode
  empty.row = 0,  # empty row
  control = list(),  # control parameters
  glm.init = FALSE  # initialize glm
)
```
### Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>data.spec</td>
<td>Site-by-species presence-absence data frame, with sites as rows and species as columns.</td>
</tr>
<tr>
<td>data.env</td>
<td>Site-by-variable data frame, with sites as rows and environmental variables as columns.</td>
</tr>
<tr>
<td>xy</td>
<td>Site coordinates, to account for distances between sites.</td>
</tr>
<tr>
<td>data.spec.pred</td>
<td>Site-by-species presence-absence data frame or list of data frames, with sites as rows and species as columns, for which zeta diversity will be computed and used as a predictor of the zeta diversity of data.spec.</td>
</tr>
<tr>
<td>order</td>
<td>Specific number of assemblages or sites at which zeta diversity is computed.</td>
</tr>
<tr>
<td>sam</td>
<td>Number of samples for which the zeta diversity is computed.</td>
</tr>
<tr>
<td>reg.type</td>
<td>Type of regression used in the multi-site generalised dissimilarity modelling. Options are &quot;glm&quot; for generalised linear models, &quot;ng1s&quot; for negative linear models, &quot;gan&quot; for generalised additive models, &quot;scan&quot; for shape constrained additive models (with monotonic decreasing by default), and &quot;ispline&quot; for I-spline models (forcing monotonic decline), as recommended in generalised dissimilarity modelling by Ferrier et al. (2007).</td>
</tr>
<tr>
<td>family</td>
<td>A description of the error distribution and link function to be used in the glm, gam and scam models (see family for details of family functions).</td>
</tr>
<tr>
<td>method.glm</td>
<td>Method used in fitting the generalised linear model. The default method &quot;glm.fit.cons&quot; is an adaptation of method glm.fit2 from package glm2 using a constrained least squares regression (default is negative coefficients) in the reweighted least squares. Another option is &quot;glm.fit2&quot;, which calls glm.fit2; see help documentation for glm.fit2 in package glm2.</td>
</tr>
<tr>
<td>cons</td>
<td>Type of constraint in the glm if method.glm = &quot;glm.fit.cons&quot;. Default is -1 for negative coefficients on the predictors. The other option is 1 for positive coefficients on the predictors.</td>
</tr>
<tr>
<td>cons.inter</td>
<td>Type of constraint for the intercept. Default is 1 for positive intercept, suitable for Gaussian family. The other option is -1 for negative intercept, suitable for binomial family.</td>
</tr>
<tr>
<td>confint.level</td>
<td>Percentage for the confidence intervals of the coefficients from the generalised linear models.</td>
</tr>
<tr>
<td>bs</td>
<td>A two-letter character string indicating the (penalized) smoothing basis to use in the scam model. Default is &quot;mpd&quot; for monotonic decreasing splines. see smooth.terms for an overview of what is available.</td>
</tr>
<tr>
<td>kn</td>
<td>Number of knots in the GAM and SCAM. Default is -1 for determining kn automatically using Generalized Cross-validation.</td>
</tr>
<tr>
<td>order.ispline</td>
<td>Order of the I-spline.</td>
</tr>
<tr>
<td>kn.ispline</td>
<td>Number of knots in the I-spline.</td>
</tr>
<tr>
<td>distance.type</td>
<td>Method to compute distance. Default is &quot;Euclidean&quot;, for Euclidean distance. The other options are (i) &quot;ortho&quot; for orthodromic distance, if xy correspond to longitudes and latitudes (orthodromic distance is computed using the geodist function from package geodist); and (ii) &quot;custom&quot;, in which case the user must provide a distance matrix for dist.custom.</td>
</tr>
</tbody>
</table>
dist.custom  Distance matrix provided by the user when distance.type = "custom".
rescale  Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by the total number of species in the dataset, to get a range of values between 0 and 1. Has no effect if normalize != FALSE.
rescale.pred  Boolean value (TRUE or FALSE) indicating if the spatial distances and differences in environmental variables should be rescaled between 0 and 1.
method  Name of a function (as a string) indicating how to combine the pairwise differences and distances for more than 3 sites. It can be a basic R-function such as "mean" or "max", but also a custom function.
normalize  Indicates if the zeta values for each sample should be divided by the total number of species for this specific sample (normalize = "Jaccard"), by the average number of species per site for this specific sample (normalize = "Sorensen"), or by the minimum number of species in the sites of this specific sample (normalize = "Simpson"). Default value is FALSE, indicating that no normalization is performed.
silent  Boolean value (TRUE or FALSE) indicating if warnings must be printed.
empty.row  Determines how to handle empty rows, i.e. sites with no species. Such sites can cause underestimations of zeta diversity, and computation errors for the normalized version of zeta due to divisions by 0. Options are "empty" to let the data untreated, "remove" to remove the empty rows, 0 to set the normalized zeta to 0 when zeta is divided by 0 during normalization (sites share no species, so are completely dissimilar), and 1 to set the normalized zeta to 1 when zeta is divided by 0 during normalization (i.e. sites are perfectly similar).
control  As for glm.
glm.init  Boolean value, indicating if the initial parameters for fitting the glm with constraint on the coefficients signs for reg.type == "ispline" should be initialised based on the correlation coefficients between the zeta values and the environmental difference or distance. glm.init = TRUE helps preventing the error message: error: cannot find valid starting values: please specify some.

Details

The environmental variables can be numeric or factorial.

If order = 1, the variables are used as such in the regression, and factorial variables must be dummy for the output of the regression to be interpretable.

For numeric variables, if order>1 the pairwise difference between sites is computed and combined according to method. For factorial variables, the distance corresponds to the number of unique values over the number of assemblages of sites specified by order.

If xy = NULL, Zeta.msgdm only uses environmental variables in the regression. Otherwise, it also computes and uses euclidian distance (average or maximum distance between multiple sites, depending on the parameters method) as an explanatory variable.

If rescale.pred = TRUE, zeta is regressed against the differences of values of the environmental variables divided by the maximum difference for each variable, to be rescaled between 0 and 1. If !is.null(xy), distances between sites are also divided by the maximum distance. If order = 1, the
variables are transformed by first subtracting their minimum value, and dividing by the difference of their maximum and minimum values.

If reg.type = "ispline", the variables are rescaled between 0 and 1 prior to computing the I-splines by subtracting their minimum value, and dividing by the difference of their maximum and minimum values.

Value

Zeta.msgdm returns a list whose component vary depending on the regression technique. The list can contain the following components:

- **val**: Vector of zeta values used in the MS-GDM.
- **predictors**: Data frame of the predictors used in the MS-GDM.
- **range.min**: Vector containing the minimum values of the numeric variables, used for rescaling the variables between 0 and 1 for I-splines (see Details).
- **range.max**: Vector containing the maximum values of the numeric variables, used for rescaling the variables between 0 and 1 for I-splines (see Details).
- **rescale.factor**: Factor by which the predictors were divided if rescale.pred = TRUE and order>1.
- **order.ispline**: The value of the original parameter, to be used in Plot.ispline.
- **kn.ispline**: The value of the original parameter, to be used in Plot.ispline.
- **model**: An object whose class depends on the type of regression (glm, nnnpls, gam or scam; I-splines return and object of class glm), corresponding to the regression over distance for the number of assemblages or sites specified in order.
- **confint**: The confidence intervals for the coefficients from generalised linear models with no constraint. confint is not generated for the other types of regression.
- **vif**: The variance inflation factors for all the variables for the generalised linear regression. vif is not generated for the other types of regression.

References


See Also

Zeta.decline.mc, Zeta.order.mc, Zeta.decline.ex, Zeta.order.ex, Predict.msgdm, Ispline
Examples

```r
utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[1:2]
data.spec.bird <- bird.spec.coarse[3:193]
utils::data(bird.env.coarse)
data.env.bird <- bird.env.coarse[,3:9]

zeta.glm <- Zeta.msgdm(data.spec.bird, data.env.bird, sam = 100, order = 3)
zeta.glm
dev.new()
graphics::plot(zeta.glm$model)

zeta.ngls <- Zeta.msgdm(data.spec.bird, data.env.bird, xy.bird, sam = 100, order = 3,
  reg.type = "ngls", rescale = TRUE)
zeta.ngls

##########

utils::data(Marion.species)
xy.marion <- Marion.species[1:2]
data.spec.marion <- Marion.species[3:33]
utils::data(Marion.env)
data.env.marion <- Marion.env[3]

zeta.gam <- Zeta.msgdm(data.spec.marion, data.env.marion, sam = 100, order = 3,
  reg.type = "gam")
zeta.gam
dev.new()
graphics::plot(zeta.gam$model)

zeta.ispline <- Zeta.msgdm(data.spec.marion, data.env.marion, xy.marion, sam = 100,
  order = 3, normalize = "Jaccard", reg.type = "ispline")
zeta.ispline

zeta.ispline.r <- Return.ispline(zeta.ispline, data.env.marion, distance = TRUE)
zeta.ispline.r
dev.new()
Plot.ispline(isplines = zeta.ispline.r, distance = TRUE)
dev.new()
Plot.ispline(msgdm = zeta.ispline, data.env = data.env.marion, distance = TRUE)
```

---

**Zeta.order.ex**

*Expectation of zeta diversity for a specific number of assemblages or sites*
Description

Computes the expectation of zeta diversity, the number of species shared by multiple assemblages, for a specific order (number of assemblages or sites) using a formula based on the occupancy of the species.

Usage

\texttt{Zeta.order.ex(}
\begin{verbatim}
data.spec,
order = 1,
sd.correct = TRUE,
rescale = FALSE,
empty.row = "empty"
\end{verbatim}
\texttt{)}

Arguments

data.spec Site-by-species presence-absence data frame, with sites as rows and species as columns.
order Specific number of assemblages or sites at which zeta diversity is computed.
sd.correct Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) or not (using the number of site combinations as the denominator).
rescale Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by $\zeta_1$, to get a range of values between 0 and 1.
empty.row Determines how to handle empty rows, i.e. sites with no species. Such sites can cause underestimations of zeta diversity. Options are "empty" to let the data untreated or "remove" to remove the empty rows.

Details

\texttt{Zeta.order.ex} is much faster than \texttt{Zeta.order.mc} to compute the exact value of zeta diversity when the number of species is lower than $C_N^i$, where $N$ is the total number of sites and $i$ is the order of zeta.

\texttt{sd.correct} should be set to TRUE if the assemblages represent a subsample of the whole system. It can be set to FALSE if the sampling is exhaustive, for example in case of a continuous regular grid covering the whole study area.

Value

\texttt{zeta.order.ex} returns a list containing the following components:

- \texttt{zeta.order} The number of assemblages or sites for which the zeta diversity was computed.
- \texttt{combinations} The number of possible combinations of sites for the chosen order.
- \texttt{zeta.val} The zeta diversity values.
- \texttt{zeta.val.sd} The standard deviation of zeta diversity.
References


See Also

`Zeta.order.mc`, `Zeta.decline.ex`, `Zeta.decline.mc`

Examples

```r
utils::data(bird.spec.coarse)
data.spec.bird <- bird.spec.coarse[,3:193]

zeta.bird <- Zeta.order.ex(data.spec.bird, order = 3)
zeta.bird

############
utils::data(Marion.species)
data.spec.marion <- Marion.species[,3:33]

zeta.marion <- Zeta.order.ex(data.spec.marion, order = 3)
zeta.marion
```

---

**Zeta.order.mc**

`Zeta diversity for a specific number of assemblages or sites using Monte Carlo sampling`

**Description**

Computes zeta diversity, the number of species shared by multiple assemblages, for a specific order (number of assemblages or sites).

**Usage**

```r
Zeta.order.mc(
  data.spec,
  xy = NULL,
  order = 1,
  sam = 1000,
  sd.correct = TRUE,
  sd.correct.adapt = FALSE,
)```
```r
rescale = FALSE,
normalize = FALSE,
NON = FALSE,
FPO = NULL,
DIR = FALSE,
empty.row = "empty",
silent = TRUE
)
```

**Arguments**

**data.spec**
Site-by-species presence-absence data frame, with sites as rows and species as columns.

**xy**
Site coordinates. This is only used if `NON` = TRUE or `DIR` = TRUE.

**order**
Specific number of assemblages or sites at which zeta diversity is computed.

**sam**
Number of samples for which the zeta diversity is computed.

**sd.correct**
Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) or not (using the number of site combinations as the denominator).

**sd.correct.adapt**
Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) if `sam` is higher than the number of possible combinations, or not (using the number of site combinations as the denominator) if `sam` is lower than the number of possible combinations. If `sd.correct.adapt` == TRUE, it takes precedence over `sd.correct`.

**rescale**
Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by $\zeta_1$, to get a range of values between 0 and 1.

**normalize**
Indicates if the zeta values for each sample should be divided by the total number of species for this specific sample (normalize = "Jaccard"), by the average number of species per site for this specific sample (normalize = "Sorensen"), or by the minimum number of species in the sites of this specific sample (normalize = "Simpson"). Default value is FALSE, indicating that no normalization is performed.

**NON**
Boolean value (TRUE or FALSE) indicating if the number of species in common should only be counted for the nearest neighbours.

**FPO**
A vector with the coordinates of the fixed point origin from which the zeta diversity will be computed (overrides NON). In that case, $\zeta_1$ is the number of species in the closest site to the FPO, $\zeta_2$ is the number of species shared by the 2 closest sites, etc.

**DIR**
Boolean value (TRUE or FALSE) indicating if zeta diversity must be computed using a directed nearest neighbour scheme in the direction away from the FPO, starting from any site.
empty.row  Determines how to handle empty rows, i.e. sites with no species. Such sites can cause underestimations of zeta diversity, and computation errors for the normalized version of zeta due to divisions by 0. Options are "empty" to let the data untreated, "remove" to remove the empty rows, 0 to set the normalized zeta to 0 when zeta is divided by 0 during normalization (sites share no species, so are completely dissimilar), and 1 to set the normalized zeta to 1 when zeta is divided by 0 during normalization (i.e. sites are perfectly similar).

silent  Boolean value (TRUE or FALSE) indicating if messages must be printed.

Details

If the number of combinations of sites is lower than the value of the parameter sam, all the combinations are used and an exact solution is computed. In that case, using the number of site combinations as the denominator may be appropriate to compute the standard deviation, if all sites were sampled and the zeta values. This can be adjusted with parameters sd.correct and sd.correct.adapt.

Zeta.order.mc is faster than Zeta.order.ex to compute the exact value of zeta diversity when the number of species is higher than $C^N_i$, where $N$ is the total number of sites and $i$ is the order of zeta.

Zeta.order.mc enables accommodating richness heterogeneity by setting normalize = "Jaccard", normalize = "Sorensen" or normalize = "Simpson". This cannot be performed by Zeta.order.ex.

Value

Zeta.order.mc returns a list containing the following components:

- zeta.order  The number of assemblages or sites for which the zeta diversity was computed.
- combinations  The number of possible combinations of sites for the chosen order.
- zeta.val  The zeta diversity values.
- zeta.val.sd  The standard deviation of zeta diversity.

References


See Also

Zeta.decline.mc

Examples

```r
utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[,1:2]
data.spec.bird <- bird.spec.coarse[,3:193]

zeta.bird <- Zeta.order.mc(data.spec.bird, order = 3, sam=100)
```
zeta.bird

 utils::data(Marion.species)
 xy.marion <- Marion.species[,1:2]
 data.spec.marion <- Marion.species[,3:33]

 zeta.marion <- Zeta.order.mc(data.spec.marion, xy.marion, order = 3, sam = 100,
 NON = TRUE)
 zeta.marion

---

Zeta.order.mc.mult

Number of species in common between a specific number of assemblages or sites using Monte Carlo sampling, for multiple combinations and several groups of taxa

---

Description

Computes the number of species shared by multiple assemblages, for a specific order (number of assemblages or sites), for multiple combinations and several groups of taxa.

Usage

Zeta.order.mc.mult(
  data.spec,  
  xy = NULL,  
  order = 1,  
  sam = 1000,  
  sd.correct = TRUE,  
  sd.correct.adapt = FALSE,  
  rescale = FALSE,  
  normalize = FALSE,  
  NON = FALSE,  
  FPO = NULL,  
  DIR = FALSE,  
  empty.row = "empty",  
  silent = TRUE
)

Arguments

data.spec A list of site-by-species presence-absence data frames, with sites as rows and species as columns.

xy Site coordinates. This is only used if NON = TRUE or DIR = TRUE.

order Specific number of assemblages or sites at which zeta diversity is computed.
sam  Number of samples for which the zeta diversity is computed.

sd.correct  Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) or not (using the number of site combinations as the denominator).

sd.correct.adapt  Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) if sam is higher than the number of possible combinations, or not (using the number of site combinations as the denominator) if sam is lower than the number of possible combinations. If sd.correct.adapt == TRUE, it takes precedence over sd.correct.

rescale  Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by \( \zeta_1 \), to get a range of values between 0 and 1.

normalize  Indicates if the zeta values for each sample should be divided by the total number of species for this specific sample (normalize = "Jaccard"), by the average number of species per site for this specific sample (normalize = "Sorensen"), or by the minimum number of species in the sites of this specific sample (normalize = "Simpson"). Default value is FALSE, indicating that no normalization is performed.

NON  Boolean value (TRUE or FALSE) indicating if the number of species in common should only be counted for the nearest neighbours.

FPO  A vector with the coordinates of the fixed point origin from which the zeta diversity will be computed (overrides NON). In that case, \( \zeta_1 \) is the number of species in the closest site to the FPO, \( \zeta_2 \) is the number of species shared by the 2 closest sites, etc.

DIR  Boolean value (TRUE or FALSE) indicating if zeta diversity must be computed using a directed nearest neighbour scheme in the direction away from the FPO, starting from any site.

empty.row  Determines how to handle empty rows, i.e. sites with no species. Such sites can cause underestimations of zeta diversity, and computation errors for the normalized version of zeta due to divisions by 0. Options are "empty" to let the data untreated, "remove" to remove the empty rows, 0 to set the normalized zeta to 0 when zeta is divided by 0 during normalization (sites share no species, so are completely dissimilar), and 1 to set the normalized zeta to 1 when zeta is divided by 0 during normalization (i.e. sites are perfectly similar).

silent  Boolean value (TRUE or FALSE) indicating if messages must be printed.

Details

Contrary to Zeta.order.mc, the number of species shared by the different combinations of assemblages are not averaged, but returned as is. This is useful to then compare local zeta diversity for different groups of taxa.

As for Zeta.order.mc, if the number of combinations of sites is lower than the value of the parameter sam, all the combinations are used and an exact solution is computed. In that case, using the
number of site combinations as the denominator may be appropriate to compute the standard deviation, if all sites were sampled and the zeta values. This can be adjusted with parameters `sd.correct` and `sd.correct.adapt`.

Value

`Zeta.order.mc.mult` returns a list containing the following components:

- `zeta.order`: The number of assemblages or sites for which the zeta diversity was computed.
- `sites`: A matrix in which each row contains the indices of a given combination, i.e. of the specific sam assemblages.
- `zeta.val`: A data frame in which each column is the number of species shared by the assemblages.

References


See Also

`Zeta.decline.mc`

Examples

```r
utils::data(Marion.species)
xy.marion <- Marion.species[1:2]
data.spec.marion <- Marion.species[3:33]

# random other communities
data.spec.marion2a <- data.spec.marion
data.spec.marion2a[which(data.spec.marion2a==1,arr.ind=TRUE)] <- 0
for(i in 1:ncol(data.spec.marion2a))
  data.spec.marion2a[sample(nrow(data.spec.marion2a),8),i] <- 1
data.spec.marion2b <- data.spec.marion
data.spec.marion2b[which(data.spec.marion2b==1,arr.ind=TRUE)] <- 0
for(i in 1:ncol(data.spec.marion2b))
  data.spec.marion2b[sample(nrow(data.spec.marion2b),8),i] <- 1
dat.spec.tot <- list(data.spec.marion,data.spec.marion2a,data.spec.marion2b)
zeta.tot <- Zeta.order.mc.mult(data.spec=dat.spec.tot,order=3,sam=200)
```
Zeta.sam.sensitivity  Sensitivity analysis for the sample size of zeta

Description

Computes zeta diversity for a given order (number of assemblages or sites) for a range of sample sizes, to assess the sensitivity to this parameter.

Usage

Zeta.sam.sensitivity(
  data.spec,  
  xy = NULL,  
  order = 1,  
  sam.seq,  
  reps = 20,  
  sd.correct = TRUE,  
  sd.correct.adapt = FALSE,  
  rescale = FALSE,  
  normalize = FALSE,  
  NON = FALSE,  
  FPO = NULL,  
  DIR = FALSE,  
  display = TRUE,  
  plot = TRUE,  
  notch = TRUE
)

Arguments

data.spec  Site-by-species presence-absence data frame, with sites as rows and species as columns.
xy  Site coordinates. This is only used if NON = TRUE or DIR = TRUE.
order  Specific number of assemblages or sites at which zeta diversity is computed.
sam.seq  Sequence of samples for which the zeta diversity is computed.
reps  Number of replicates of zeta diversity computations for each sample size.
sd.correct  Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) or not (using the number of site combinations as the denominator).

sd.correct.adapt  Boolean value (TRUE or FALSE) indicating if the standard deviation must be computed with an unbiased estimator (using the number of site combinations - 1 as the denominator) if sam is higher than the number of possible combinations, or not (using the number of site combinations as the denominator) if sam is lower
than the number of possible combinations. If sd.correct.adapt == TRUE, it takes precedence over sd.correct.

rescale

Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by $\zeta_1$, to get a range of values between 0 and 1.

normalize

Indicates if the zeta values for each sample should be divided by the total number of species for this specific sample (normalize = "Jaccard"), by the average number of species per site for this specific sample (normalize = "Sorensen"), or by the minimum number of species in the sites of this specific sample (normalize = "Simpson"). Default value is FALSE, indicating that no normalization is performed.

NON

Boolean value (TRUE or FALSE) indicating if the number of species in common should only be counted for the nearest neighbours.

FPO

A vector with the coordinates of the fixed point origin from which the zeta diversity will be computed (overrides NON). In that case, $\zeta_1$ is the number of species in the closest site to the FPO, $\zeta_2$ is the number of species shared by the 2 closest sites, etc.

DIR

Boolean value (TRUE or FALSE) indicating if zeta diversity must be computed using a directed nearest neighbour scheme in the direction away from the FPO, starting from any site.

display

Boolean value (TRUE or FALSE) indicating if the current value of the sample size must be displayed. Acts as a counter.

plot

Boolean value (TRUE or FALSE) indicating if the outputs must be plotted as a boxplot of the zeta diversity distributions for each sample size

notch

Boolean value (TRUE or FALSE) indicating if the notches must be plotted in the boxplot.

Details

Note that the execution of Zeta.sam.sensitivity can be quite lengthy, because of the number of replicates needed.

Value

Zeta.sam.sensitivity returns a matrix with (sam.max-sam.min)/sam.incr columns and reps rows.

References


See Also

Zeta.decline.mc, Zeta.order.mc, Zeta.decline.ex, Zeta.order.ex
Examples

#Note that the sensitivity analyses in the following two examples are quite long to run, #typically around 10 minutes for the first example and 1-2 minutes for the second.

utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[1:2]
data.spec.bird <- bird.spec.coarse[3:193]

dev.new()
zeta.sens.bird <- Zeta.sam.sensitivity(data.spec.bird, order = 3,
  sam.seq = seq(250,1000,250), reps = 20, display = TRUE, plot = TRUE, notch = TRUE)
zeta.sens.bird

############

utils::data(Marion.species)
xy.marion <- Marion.species[1:2]
data.spec.marion <- Marion.species[3:33]

dev.new()
zeta.sens.marion <- Zeta.sam.sensitivity(data.spec.marion, order = 3,
  sam.seq = seq(50,250,50), reps = 20, plot = TRUE, notch = TRUE)
zeta.sens.marion

---

**Zeta.scale.min.dist**

Zeta diversity scaling with sample grain dependency based on the minimum distance between sites

**Description**

Computes zeta diversity scaling with sample grain for a specific order (number of assemblages or sites), increasing grain by sequentially adding sites based on the minimum distance between them.

**Usage**

Zeta.scale.min.dist(
  xy,
  data.spec,
  m,
  order = 1,
  reorder = 100,
  shuffle = TRUE,
  sam = 1000,
  method = "mean",
  rescale = FALSE,
  normalize = FALSE,
plot = TRUE,
sd = TRUE,
distance.type = "Euclidean",
dist.custom = NULL,
zeta.type = "exact"
)

Arguments

xy Site-by-coordinate data frame, with sites as rows and coordinates as columns.
data.spec Site-by-species presence-absence data frame, with sites as rows and species as columns.
m Vector of mapping grains: m[i] sites are grouped together to generate data at a coarser grain.
order Specific number of assemblages or sites at which zeta diversity is computed.
reorder Number of times the sites are rearranged and grouped together for the computation of zeta (see Details).
shuffle Boolean value (TRUE or FALSE) indicating if the order of the sites must be randomised, which can have an impact on the outputs if some distances are equal.
sam Number of samples for which the zeta diversity is computed.
method Name of a function (as a string) indicating how to combine the coordinates. It can be a basic R-function such as "mean" or "max", but also a custom function.
rescale Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by $\zeta_1$, to get a range of values between 0 and 1. Has no effect if normalize != FALSE.
normalize Indicates if the zeta values for each sample should be divided by the total number of species for this specific sample (normalize = "Jaccard"), by the average number of species per site for this specific sample (normalize = "Sorensen"), or by the minimum number of species in the sites of this specific sample (normalize = "Simpson"). Default value is FALSE, indicating that no normalization is performed.
plot Boolean value (TRUE or FALSE) indicating if the outputs must be plotted.
sd Boolean value (TRUE or FALSE) indicating if the standard deviation must be plotted for each grain.
distance.type Method to compute distance. Default is "Euclidean", for Euclidean distance. The other options are (i) "ortho" for orthodromic distance, if xy correspond to longitudes and latitudes (orthodromic distance is computed using the geodist function from package geodist); and (ii) "custom", in which case the user must provide a distance matrix for dist.custom.
dist.custom Distance matrix provided by the user when distance.type = "custom".
zeta.type The function that must be used for the computation of zeta diversity. Default is "exact" for calling Zeta.order.ex. Use "monte carlo" for calling Zeta.order.mc.
Details

The nearest neighbouring sites (plots, quadrates, or areas of varying shapes) are grouped as spatial clusters of 2, 3, 4, etc. sites, based on the minimum distance between them. Since the procedure is based on the relative distance between sites, the site order can have an impact on the output. The procedure is therefore performed 'reorder' times, for which sites are randomly reordered each time, and the mean zeta is computed. This function is suitable for both regularly and irregularly spaced sites, contiguous or non contiguous (*sensu* Scheiner et al., 2011). For regularly spaced sites, the use of `Zeta.scale.regular` is recommended.

Value

`zeta.scale.min.dist` returns a list containing the following components:

- `order` The order of zeta.
- `m` The vector of mapping grains: m[i] sites are grouped together to generate data at a coarser grain.
- `values` A matrix containing the zeta diversity values over the 'reorder' computations, for each grain.
- `sd` A matrix containing the standard deviation of zeta diversity over the 'reorder' computations, for each grain.

References


See Also

`Zeta.decline.mc`, `Zeta.order.mc`, `Zeta.decline.ex`, `Zeta.order.ex`,

`Zeta.scale.regular`, `rescale.regular`

Examples

```r
utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]

dev.new()
zeta.scale.irreg.species <- Zeta.scale.min.dist(xy.marion, data.spec.marion, m = 1:3, order = 3, reorder = 3, sam = 50, normalize = "Jaccard")
```
Zeta.scale.regular

Zeta diversity scaling with sample grain using hierarchical increases in grain size

Description

Computes zeta diversity scaling with sample grain for a specific order (number of assemblages or sites), increasing grain by hierarchically nesting of regularly spaced sites.

Usage

Zeta.scale.regular(
  xy,
  data.spec,
  n,
  order = 1,
  sam = 1000,
  method = "mean",
  rescale = FALSE,
  normalize = FALSE,
  plot = TRUE,
  zeta.type = "exact"
)

Arguments

xy Site-by-coordinate data frame, with sites as rows and coordinates as columns.
data.spec Site-by-species presence-absence data frame, with sites as rows and species as columns.
n Vector of mapping grains: regularly spaced sites are grouped as n[i] x n[i] sites to generate data at a coarser grain.
order Specific number of assemblages or sites at which zeta diversity is computed.
sam Number of samples for which the zeta diversity is computed.
method Name of a function (as a string) indicating how to combine the coordinates. It can be a basic R-function such as "mean" or "max", but also a custom function.
rescale Boolean value (TRUE or FALSE) indicating if the zeta values should be divided by ζ1, to get a range of values between 0 and 1. Has no effect if normalize != FALSE.
normalize Indicates if the zeta values for each sample should be divided by the total number of species for this specific sample (normalize = "Jaccard"), by the average number of species per site for this specific sample (normalize = "Sorensen"), or by the minimum number of species in the sites of this specific sample (normalize = "Simpson"). Default value is FALSE, indicating that no normalization is performed.
plot Boolean value (TRUE or FALSE) indicating if the outputs must be plotted.
zeta.type  The function that must be used for the computation of zeta diversity. Default is "exact" for calling Zeta.order.ex. Use "monte carlo" for calling Zeta.order.mc.

Details

The sites (plots or quadrates) are incrementally aggregated as nearest neighbouring groups of 4, 9, etc. sites, using a nested approach, starting from the lowest x and y, to increase the grain. The sites can be spatially contiguous or discontiguous, as long as they are regularly spaced (see Scheiner et al., 2011). If the total number of sites is not a multiple of \( n[i] \times n[i] \), the extra sites are discarded.

Value

Zeta.scale.regular returns a list containing the following components:

- order  The order of zeta.
- n  The vector of mapping grains: regularly spaced sites are grouped as \( n[i] \times n[i] \) sites to generate data at a coarser grain.
- values  The zeta diversity values for each grain.
- sd  The standard deviation of zeta diversity for each grain.

References


See Also

Zeta.decline.mc, Zeta.order.mc, Zeta.decline.ex, Zeta.order.ex

Zeta.scale.min.dist, rescale.regular, rescale.min.dist

Examples

```r
utils::data(bird.spec.fine)
xy.bird <- bird.spec.fine[1:400,1:2]
data.spec.bird <- bird.spec.fine[1:400,3:192]

dev.new()
##sam = 25 is used here for fast execution, but a higher value is advised
zeta.scale.reg <- Zeta.scale.regular(xy.bird, data.spec.bird, n = 1:3, order = 3, sam = 25, normalize = "Jaccard", zeta.type="monte carlo")
```
Zeta.varpart

Variation partitioning for zeta diversity

Description

Variation partitioning of zeta diversity for a specific order (number of assemblages or sites) over distance and environmental variables.

Usage

Zeta.varpart(
  msgdm.mod,
  num.part = 2,
  reg.type = "glm",
  family = stats::gaussian(),
  method.glm = "glm.fit.cons",
  cons = -1,
  cons.inter = 1,
  kn = -1,
  bs = "mpd"
)

Arguments

msgdm.mod An object return by function Zeta.msgdm.
num.part Number of partitions of zeta diversity. Can be 2 or 3.
reg.type Type of regression for the multi-site generalised dissimilarity modelling. Options are "glm" for generalised linear models, "nqls" for negative linear models, "gam" for generalised additive models, "scam" for shape constrained additive models, and "ispline" for I-spline models, as recommended in generalised dissimilarity modelling by Ferrier et al. (2007).
family A description of the error distribution and link function to be used in the glm, gam and scam models (see family for details of family functions).
method.glm Method used in fitting the generalised linear model. The default method "glm.fit.cons" is an adaptation of method glm.fit2 from package glm2 using a negative least squares regression in the reweighted least squares. Another option is "glm.fit2", which corresponds to method glm.fit2; see help documentation for glm.fit2 in package glm.
cons type of constraint in the glm if method.glm = "glm.fit.cons". Default is -1 for negative coefficients on the predictors. The other option is 1 for positive coefficients on the predictors.
cons.inter type of constraint for the intercept. Default is 1 for positive intercept, suitable for Gaussian family. The other option is -1 for negative intercept, suitable for binomial family.
kn  Number of knots in the GAM and SCAM. Default is -1 for determining kn automatically using Generalized Cross-validation.

bs  A two-letter character string indicating the (penalized) smoothing basis to use in the scam model. Default is "mpd" for monotonic decreasing splines. see smooth.terms for an overview of what is available.

Details

Note that, for a given regression, the variation explained is computed as 1-(RSS/TSS)*(v-1)/(v-p-1), where RSS is the residual sum of squares and TSS is the total sum of squares, v is the number of variables used in the regression (which is greater than the original number of variables for I-splines) and p is the number of samples. 1-(RSS/TSS) corresponds to the classical R-squared for linear regression only, and results for non-linear regressions should be interpreted with caution.

The environmental variables can be numeric or factorial, and order must be greater than 1.

For numeric variables, the pairwise difference between sites is computed and combined according to method. For factorial variables, the distance corresponds to the number of unique values over the number of assemblages of sites specified by order.

Zeta is regressed against the differences of values of the environmental variables divided by the maximum difference for each variable, to be rescaled between 0 and 1. If !is.null(xy), distances between sites are also divided by the maximum distance.

Value

Zeta.varpart returns a data frame with one column containing the variation explained by each component a (the variation explained by distance alone), b (the variation explained by either distance or the environment), c (the variation explained by the environment alone) and d (the unexplained variation).

References


See Also

Zeta.decline.mc, Zeta.order.mc, Zeta.decline.ex, Zeta.order.ex, Zeta.msgdm, pie.neg

Examples

```r
utils::data(bird.spec.coarse)
xy.bird <- bird.spec.coarse[,1:2]
data.spec.bird <- bird.spec.coarse[,3:193]
utils::data(bird.env.coarse)
data.env.bird <- bird.env.coarse[,3:9]
```
```r
zeta.bird <- Zeta.msgdm(data.spec.bird, data.env.bird, xy.bird, sam = 100, order = 3)
zeta.varpart.bird <- Zeta.varpart(zeta.bird, method.glm = "glm.fit2")
zeta.varpart.bird
dev.new()
pie.neg(zeta.varpart.bird[4:7,1], density = c(4, 0, 8, -1),
angle = c(90, 0, 0, 0),
labels = c("distance", "undistinguishable", "environment", "unexplained"),
radius = 0.9)

 utils::data(Marion.species)
xy.marion <- Marion.species[,1:2]
data.spec.marion <- Marion.species[,3:33]
utils::data(Marion.env)
data.env.marion <- Marion.env[3:4]
zeta.marion <- Zeta.msgdm(data.spec.marion, data.env.marion, xy.marion, sam = 100,
order = 3, normalize = "Jaccard")
zeta.varpart.marion <- Zeta.varpart(zeta.marion, method.glm = "glm.fit2")
zeta.varpart.marion
dev.new()
pie.neg(zeta.varpart.marion[4:7,1], density = c(4, 0, 8, -1),
angle = c(90, 0, 0, 0),
labels = c("distance", "undistinguishable", "environment", "unexplained"),
radius = 0.9)
```
Index

* data
  bird.env.coarse, 2
  bird.env.fine, 3
  bird.spec.coarse, 4
  bird.spec.fine, 5
  Marion.env, 10
  Marion.species, 11

  as.graphicsAnnot, 12
  bird.env.coarse, 2
  bird.env.fine, 3
  bird.spec.coarse, 4
  bird.spec.fine, 5
  family, 23, 30, 33, 41, 59

  glm, 6, 7, 23, 24, 42
  glm.cons, 6
  glm.fit, 8, 9
  glm.fit.cons, 8
  glm.fit2, 9
  glm2, 7

  Ispline, 9, 21, 24, 29, 43

  Marion.env, 10
  Marion.species, 11

  pie, 13
  pie.neg, 12, 60
  Plot.ispline, 13
  Plot.zeta.ddecay, 15, 17, 32
  Plot.zeta.ddecays, 16
  Plot.zeta.decline, 17, 36, 39
  Plot.zeta.scale.min.dist, 18, 20
  Plot.zeta.scale.regular, 19, 20
  Predict.msgdm, 21, 43

  Reg.ispline, 23
  rescale.min.dist, 25, 28, 58
  rescale.regular, 19, 20, 26, 27, 56, 58
  Return.ispline, 28

  smooth.terms, 31, 41, 60

  Zeta.ddecay, 16, 17, 29, 34
  Zeta.ddecays, 16, 17, 32, 32
  Zeta.decline.ex, 16–20, 26, 28, 32, 34, 34,
  38, 39, 43, 46, 53, 56, 58, 60
  Zeta.decline.mc, 16–20, 26, 28, 32, 34–36,
  38, 43, 46, 48, 51, 53, 56, 58, 60
  Zeta.msgdm, 10, 15, 22, 24, 29, 39, 40, 59, 60
  Zeta.order.ex, 16–20, 26, 28, 32, 34, 36, 39,
  43, 44, 48, 53, 56, 58, 60
  Zeta.order.mc, 16–20, 26, 28, 32, 34, 36, 39,
  43, 45, 46, 48, 53, 56, 58, 60
  Zeta.order.mc.mult, 49
  Zeta.sam.sensitivity, 52
  Zeta.scale.min.dist, 19, 20, 26, 28, 54, 58
  Zeta.scale.regular, 19, 20, 26, 28, 56, 57
  Zeta.varpart, 13, 59