

# Package ‘paleoMAS’

July 2, 2014

**Type** Package

**Title** Paleoecological Analysis

**Version** 2.0-1

**Date** 2011-01-30

**Author** Alexander Correa-Metrio, Dunia H. Urrego, Kenneth R. Cabrera, Mark B. Bush

**Maintainer** Alexander Correa-Metrio <acorrea@geologia.unam.mx>

**Depends** MASS, vegan, lattice

**Description** Transfer functions and statistical operations for paleoecology

**License** GPL (>= 2)

**LazyLoad** yes

**Repository** CRAN

**Date/Publication** 2012-10-29 08:59:22

**NeedsCompilation** no

## R topics documented:

paleoMAS-package . . . . .	2
akaike.all . . . . .	3
akaike.l . . . . .	4
analog.mult . . . . .	6
analog.sing . . . . .	8
bcrossv.all . . . . .	9
bcrossv.l . . . . .	10
bcrossv.l1 . . . . .	12
bcrossv.sa . . . . .	13
change . . . . .	15
chron . . . . .	17

dispersion.all . . . . .	18
error.l . . . . .	20
filter.p . . . . .	22
fossil.dist . . . . .	23
fossil.values . . . . .	24
fossilq . . . . .	26
modernq . . . . .	27
percenta . . . . .	27
quexilchron . . . . .	29
quexildepths . . . . .	30
quexilper . . . . .	31
run.mean . . . . .	32
simulat . . . . .	32
simulat.t . . . . .	33
syas . . . . .	35
synthetic . . . . .	37
vcrossv.all . . . . .	39
vcrossv.da . . . . .	41

<b>Index</b>	<b>43</b>
--------------	-----------

---

paleoMAS-package	<i>Paleoecological Analysis</i>
------------------	---------------------------------

---

## Description

Transfer functions and statistical operations for paleoecology

## Details

Package:	paleoMAS
Type:	Package
Version:	2.0-1
Date:	2009-12-06
License:	GPL
LazyLoad:	yes

## Author(s)

Alexander Correa-Metrio, Dunia H. Urrego, Kenneth R. Cabrera, Mark B. Bush.

Maintainer: Alexander Correa-Metrio <acorrea@my.fit.edu>

## References

- Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene
- Correa-Metrio, A. 2010. Climate and vegetation of the Yucatan Peninsula during the late Pleistocene. PhD Dissertation, Florida Institute of Technology, Melbourne, FL. 194 p.
- Correa-Metrio, A., K.R. Cabrera, and M.B. Bush. 2010. Quantifying ecological change through discriminant analysis: a paleoecological example from the Peruvian Amazon. *Journal of Vegetation Science* 21: 695-704.
- Urrego, D., M.B. Bush, M. Silman, A. Correa-Metrio, M. Ledru, F. Mayle, and B. Valencia. 2009. Millennial-scale ecological changes in tropical South America since the Last Glacial Maximum. In: F. Vimieux, F. Sylvestre, and M. Khodri, editors. Past climate variability from the Last Glacial Maximum to the Holocene in South America and surrounding regions. Springer, Paris.

---

akaike.all	<i>LOESS regression parameters selection through AIC minimization for multiple taxa</i>
------------	---

---

## Description

This function applies Akaike Information Criterion (AIC) to varying  $\alpha$  and degree in LOESS regression for multiple taxa. The final result contains alpha and degree values that minimize AIC for each taxon LOESS regression as a function of the environmental parameter.

## Usage

```
akaike.all(x, y, interval = c(0.15, 1, 0.05))
```

## Arguments

- |          |  |
|----------|--|
| x        | A vector containing the environmental gradient value for each sample or location.  |
| y        | A matrix containing the observed taxa abundances along the environmental gradient. Taxa are in columns and samples in rows.  |
| interval | $\alpha$ values to be evaluated. Three numbers in the following order: lower and upper limits of the sequence of $\alpha$ values to be evaluated, and increment of the sequence. |

## Details

See [akaike.1](#) for details. This procedure is less computation intensive than cross-validation, but the resulting parameters do not provide the LOESS regression that best fit the data. Instead, it selects the parameters that best fit the data conditioned to simplicity of the model.

**Value**

A matrix containing the values of  $\alpha$  and degree that minimize AIC of the LOESS regression for each taxon, and the value of the AIC.

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera.

**References**

Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene

Hurvich, C.M., and J.S. Simonoff. 1998. Smoothing parameters selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Society, Series B* 60: 271-293.

Cleveland, W.S., and S.J. Devlin. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83: 596-610.

**See Also**

[loess](#) for details on LOESS regression, and [akaike.l](#) for details on AIC.

**Examples**

```
data(modernq)
# Calculate percentages
perq<-percenta(modernq,first=2,last=39)[,2:55]
akaike.all(modernq[,1],perq[,1:10])
```

---

akaike.l

*Loess regression parameters selection through AIC minimization for one taxon*

---

**Description**

This function calculates the Akaike Information Criterion (AIC) of LOESS regressions with varying  $\alpha$  and degree for a single taxon.

**Usage**

```
akaike.l(x, y, interval = c(0.15, 1, 0.05), plot = TRUE,
         parameters = FALSE)
```

**Arguments**

x	A vector containing the environmental gradient value for each sample or location.
y	A matrix containing the observed taxa abundances along the environmental gradient. Taxa are in columns and samples in rows.
interval	$\alpha$ values to be evaluated. Three numbers in the following order: lower and upper limits of the sequence of $\alpha$ values to be evaluated, and increment of the sequence.
plot	Logical indicating whether or not a figure displaying the behavior of AIC through $\alpha$ values conditioned to degree is plotted.
parameters	Logical indicating whether or not a matrix with the parameters of each LOESS regression is desired.

**Details**

AIC is calculated according to Hurvich & Simonoff (1998) for LOESS regressions (Cleveland and Devlin, 1988) generated through all possible combinations between degree (1 and 2) and the generated series of  $\alpha$  values.

**Value**

A list with two components containing:

aic.loess	A matrix with the AIC values for each combination of $\alpha$ and degree.
minimum	The values of $\alpha$ and degree that minimize AIC.
minimum	A matrix containing all the parameters of the LOESS regression for each combination of $\alpha$ and degree (see Cleveland and Devlin (1988) and Hurvich & Simonoff (1998) for details). Only relevant if parameters=TRUE.

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera.

**References**

- Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene
- Hurvich, C.M., and J.S. Simonoff. 1998. Smoothing parameters selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Society, Series B* 60: 271-293.
- Cleveland, W.S., and S.J. Devlin. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83: 596-610.

**See Also**

[loess](#) for details on LOESS regression.

## Examples

```
data(modernq)
# Calculate percentages
perq<-percenta(modernq,first=2,last=39)[,2:55]
akaike.l(modernq[,1],perq[,3])
akaike.l(modernq[,1],perq[,3],parameters=TRUE)
```

---

analog.mult

*Multiple analogs comparison*

---

## Description

This function calculates and plots distance or dissimilarity indices (degree of analogy) between all samples in a fossil record and a set of modern or fossil analog samples.

## Usage

```
analog.mult(base, fossil, age, res = 100, method = "euclidean",
            binary = FALSE, plot = TRUE, dimension = TRUE,
            wire = FALSE, aspect = c(0.5, 0.3), drape = TRUE)
```

## Arguments

base	A matrix containing analog samples with samples in rows and taxa in columns. Rows must be named after each analog sample that they contain (see <a href="#">rownames</a> ).
fossil	A matrix with the fossil dataset containing samples in rows and species in columns.
age	A vector with the age or depth of each sample.
res	Temporal resolution for the interpolation needed to plot the results.
method	Corresponds to squared chord distance ("schord") and the methods available in <code>vegdist</code> of <b>vegan</b> . Available methods are "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "morisita", "horn", "mountford", "raup", "binomial" or "chao". See <a href="#">vegdist</a> (package <b>vegan</b> ) for details.
binary	Logical indicating whether a transformation of the data into presence/absence is desired. See <a href="#">vegdist</a> (package <b>vegan</b> ) for details.
plot	Logical indicating whether or not a plot is desired.
dimension	An indicator of whether or not a three-dimensional plot is desired. If TRUE a 3D figure is plotted according to the argument specified in <code>wire</code> . If FALSE a set of bidimensional panels is produced.
wire	An indicator of the kind of 3D figure that is desired. If TRUE a figure type <a href="#">wireframe</a> (package <b>lattice</b> ) is produced. If FALSE a figure type <a href="#">levelplot</a> (package <b>MASS</b> ) is produced. Only relevant if <code>dimension=TRUE</code> .
aspect	Argument of the function <a href="#">wireframe</a> . Only relevant if <code>wire=TRUE</code> .
drape	Argument of the function <a href="#">wireframe</a> . Only relevant if <code>wire=TRUE</code> .

## Details

When 3D plots are desired, they are based on a grid with a uniform time resolution. Therefore, a linear interpolation is performed, which requires the definition of the desired time resolution through the argument `res`.

## Value

This function returns a list containing:

<code>plot.data</code>	A matrix with distances between fossil samples and analogs interpolated into regular time intervals. The size of the interval was defined in the argument <code>res</code> .
<code>distances</code>	A matrix with the age of fossil samples, and their distance or dissimilarity to each analog sample.

## Author(s)

Alexander Correa-Metrio.

## References

Correa-Metrio, A. 2010. Climate and vegetation of the Yucatan Peninsula during the late Pleistocene. PhD Dissertation, Florida Institute of Technology, Melbourne, FL. 194 p.

## See Also

[analog.sing](#). See [vegdist](#) for details on the dissimilarity indexes.

## Examples

```
data(quexilchron,quexildepths,quexilper)
#Define chronolgy
age<-quexilchron
ages<-chron(age,quexildepths,max.depth=1957)$chronology[,2]
#If samples are used as representative of time periods
quexilbase<-quexilper[c(1,27,68),]
rownames(quexilbase)<-c("Late Pleistocene","LGM","MIS3")
#comparison of all fossil samples with the three periods
analog.mult(quexilbase,quexilper,ages,res=500)
#three-dimensional figure
analog.mult(quexilbase,quexilper,ages,res=500,plot=TRUE,dimension=TRUE,wire=TRUE)
#multipanel figure
analog.mult(quexilbase,quexilper,ages,res=500,plot=TRUE,dimension=FALSE)
```

---

analog.sing

*Single analog comparison*


---

### Description

This function calculates and plots distance or dissimilarity indexes (analogy degree) between all samples in a fossil record and a single sample from the same record.

### Usage

```
analog.sing(fossil, base, age, dca = FALSE, method = "euclidean",
            binary = FALSE)
```

### Arguments

fossil	A matrix of fossil data containing samples in rows and species in columns.
base	Row number where the analog sample is.
age	A vector with samples ages.
dca	Logical indicating whether or not distances should be calculated based on detrended correspondence analysis scores (DCA performed according to decorana of <b>vegan</b> ). If dca=TRUE, the dissimilarity between samples is calculated using scores of the first 4 axes of a DCA. Some DCA scores are negative, therefore caution is needed to select an appropriate dissimilarity method when dca=TRUE. If method="schord" DCA cannot be used.
method	Corresponds to squared chord distance ("schord"), and methods available in <code>vegdist</code> of <b>vegan</b> . Available methods are "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "morisita", "horn", "mountford", "raup", "binomial" or "chao". See <a href="#">vegdist</a> (package <b>vegan</b> ) for details.
binary	Argument of the function <code>vegdist</code> (package <b>vegan</b> ) that standardizes the data into presence/absence before calculating dissimilarities or distances.

### Value

Returns a matrix with sample ages and resulting distance or dissimilarity index between each sample and the selected analog.

### Author(s)

Dunia H. Urrego, Alexander Correa-Metrio, Mark B. Bush.

### References

Correa-Metrio, A. 2010. Climate and vegetation of the Yucatan Peninsula during the late Pleistocene. PhD Dissertation, Florida Institute of Technology, Melbourne, FL. 194 p.

Overpeck, J.T. T.I. Webb, and I.C. Prentice. 1985, Quantitative interpretation of fossil pollen spectra: Dissimilarity coefficients and the method of modern analogs. *Quaternary Research* 23: 87-708.



**See Also**

[analog.mult](#). See [vegdist](#) for details on the dissimilarity indices.

**Examples**

```
data(quexilper, quexilchron, quexildepths)
age<-quexilchron
d<-quexildepths
ages<-chron(age,d,max.depth=1957,dates="bars")$chronology[,2]
#distance to the first sample without DCA, using canberra metric
analog.sing(quexilper,base=1,ages,method="canberra")
#distance to the first sample with DCA, using euclidean distance
analog.sing(quexilper,base=1,ages,dca=TRUE)
```

---

bcrossv.all	<i>LOESS regression parameters selection through bootstrap cross-validation for multiple taxa</i>
-------------	---

---

**Description**

This function applies  $v$ -fold crossvalidation (CV) of loess regressions with varying  $\alpha$  and degree for multiple taxa. The final result contains  $\alpha$  and degree values that minimize the error of the loess regression for each taxon.

**Usage**

```
bcrossv.all(x, y, interval = c(0.15, 1, 0.05),
           trials = c(10, 0.25), target = c("rse", "rmse"))
```

**Arguments**

x	A vector containing the environmental gradient value for each sample or location.
y	A matrix containing the observed taxa abundances along the environmental gradient. Taxa are in columns and samples in rows.
interval	$\alpha$ values to be evaluated. Three numbers in the following order: lower and upper limits of the sequence of $\alpha$ values to be evaluated, and increment of the sequence.
trials	Values for number of repetitions of the crossvalidation (100 by default), and proportion of observations to be left out each time the CV is repeated (25 percent of observations by default). The observations left out each time are randomly selected with replacement.
target	If target="rse" the values of alpha and degree that minimize the LOESS regression root of the predictive squared error are selected. If target="rmse" the values of $\alpha$ and degree that minimize the LOESS regression root mean squared error are selected.

**Value**

A matrix containing  $\alpha$  and degree values that minimize the LOESS regression rse or rmse (depending on the choice in target.)

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera.

**References**

Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene  
 Cleveland, W.S., and S.J. Devlin. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83: 596-610.

**See Also**

[loess](#) for details on loess regression. [bcrossv.l1](#) and [bcrossv.l](#) for details on the bootstrap cross-validation.

**Examples**

```
data(modernq)
# Calculate percentages
perq<-percenta(modernq,first=2,last=39)[,2:55]
bcrossv.all(modernq[,1],perq[,1:5],trials=c(10,0.1),target="rse")
bcrossv.all(modernq[,1],perq[,1:5],trials=c(10,0.1),target="rmse")
```

---

bcrossv.l

*LOESS regression parameters selection through bootstrap cross-validation for one taxon*

---

**Description**

This function applies v-fold crossvalidation (CV) of LOESS regressions with varying  $\alpha$  and degree for a single taxon.

**Usage**

```
bcrossv.l(x, y, interval = c(0.15, 1, 0.05),
          trials = c(10, 0.25), plot = TRUE)
```

**Arguments**

x	A vector containing the environmental gradient value for each sample or location.
y	A vector containing the observed taxa abundances along the environmental gradient.
interval	$\alpha$ values to be evaluated. Three numbers in the following order: lower and upper limits of the $\alpha$ -values sequenceto be evaluated, and the increment of the sequence.
trials	Values for number of repetitions of the crossvalidation (100 by default), and proportion of observations to be left out each time the CV is repeated (0.25 by default). The observations left out each time are randomly selected with replacement.
plot	Logical indicating whether or not an observed vs. error is desired

**Details**

A sequence of  $\alpha$  values is generated according to the range and increments declared in `interval` (from 0.15 to 1 in 0.05 increments). The sequence of  $\alpha$  values is combined with degree 1 and 2, and the LOESS regression produced using each combination is cross-validated as explained in [bcrossv.11](#). See Cleveland and Devlin (1988) for details on loess regression.

**Value**

A matrix of 4 columns. All possible combinations of  $\alpha$  values generated by the sequence `interval` and degree 1 and 2 are contained in the first and second columns. The third column contains the root of the predictive squared error (`rse`,  $(1/n * \sum(\text{est-obs})^2)^{0.5}$ ). The fourth column contains the root mean squared error (`rmse`,  $1/n * \sum(\text{abs}(\text{obs-est}))$ ). If `plot=TRUE`, a scatter plot of the `rse` and the `rmse` conditioned to the degree of the loess regression is produced.

This function involves very intensive computations. Therefore, the user must find a balance between values of `interval` and `trials`, and the computational capabilities. Large number of trials may require long computation times.

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera.

**References**

Cleveland, W.S., and S.J. Devlin. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83: 596-610.

**See Also**

[loess](#) for details on loess regression. [bcrossv.11](#) for details on the bootstrap cross-validation.

**Examples**

```

data(modernq)
# Calculate percentages
perq<-percenta(modernq,first=2,last=39)[,2:55]
# Cross-validation for Pinus
bcrossv.l(modernq[,1],perq[,3],trials=c(10,0.1))

```

bcrossv.11

*LOESS regression bootstrap cross-validation with fixed parameters***Description**

This function performs v-fold cross-validation (CV) of a LOESS regression of known  $\alpha$  and degree for a single taxon.

**Usage**

```

bcrossv.11(x, y, trials = c(100, 0.25), span = 0.75, degree = 2,
           plot = TRUE, estimated = FALSE)

```

**Arguments**

x	A vector containing the environmental gradient value for each sample or location.
y	A vector containing the observed taxa abundances along the environmental gradient.
trials	Values for number of repetitions of the crossvalidation (100 by default), and proportion of observations to be left out each time the CV is repeated (0.25 by default). The observations left out each time are randomly selected with replacement.
span	$\alpha$ value for the LOESS regression.
degree	Degree of the LOESS regression. Limited to 1 and 2.
plot	Logical indicating whether or not an observed vs. error plot is desired
estimated	Logical indicating whether or not the matrix with the estimated values is desired.

**Details**

The LOESS regression (Cleveland and Devlin, 1988) CV is repeated as many times as indicated in the first component of `trials`. Each repetition leaves out a fraction of the total number of observations (indicated in the second component of `trials`).  $\alpha$  and degree values for the LOESS regression are fixed and indicated in `span` and `degree` respectively. Returns the results of the crossvalidation and an error matrix containing the predictive squared error (se,  $1/n \cdot \sum(\text{est-obs})^2$ ), the root predictive squared error (rse,  $\text{pse}^{0.5}$ ), and root mean squared error (rmse,  $1/n \cdot \sum(\text{abs}(\text{obs-est}))$ ).

If the remaining number of observations after leaving out a certain number of elements is too low, the fitting of the LOESS regression is not possible. Therefore, it is recommended to keep the second component of `trials` low, unless the number of observations is large enough.

**Value**

A matrix error containing the predictive squared error (se), the root predictive squared error (rse), and the root mean squared error (rmse). If `estimated=TRUE`, returns an additional matrix of three columns containing the values of the environmental parameter ("x"), and observed ("observed") and predicted ("predicted") percentages.

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera.

**References**

Cleveland, W.S., and S.J. Devlin. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83: 596-610.

**See Also**

[loess](#) for details on LOESS regression.

**Examples**

```
data(modernq)
# Calculate percentages
perq<-percenta(modernq,first=2,last=39)[,2:55]
# Cross-validation for Pinus
bcrossv.l1(modernq[,1],perq[,3])
```

---

bcrossv.sa

*Bootstrap cross-validation of synthetic assemblages fossil estimates*

---

**Description**

This function applies repeated v-fold cross-validation to the synthetic assemblages fossil estimations, producing an estimation of the error associated with the reconstruction.

**Usage**

```
bcrossv.sa(x, y, z, delta, method = "canberra", quant = 0.05,
           trials = c(10, 0.05), detrend = FALSE)
```

**Arguments**

x                    A vector containing the environmental gradient value for each sample or location.

y                    A matrix containing the observed taxa abundances along the environmental gradient. Taxa are in columns and samples in rows. See [synthetic](#) for details

z	A matrix containing the parameters for the LOESS regression of each species ( $\alpha$ and degree in that order). Objects produced by <code>bcrossv.all</code> and <code>akaike.all</code> could be used as z.
delta	Increments for the LOESS regression prediction. See <code>predict.loess</code> for details.
method	Corresponds to methods available in <code>vegdist</code> of <b>vegan</b> . Available methods are "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "morisita", "horn", "mountford", "raup", "binomial" or "chao". See <code>vegdist</code> (package <b>vegan</b> ) for details.
quant	The quantile of the distances that will be used to obtain the mean environmental parameter. See <code>fossil.values</code> for details.
trials	Values for number of repetitions of the crossvalidation (10 by default), and proportion of observations to be left out each time the CV is repeated (5 percent of observations by default). The observations left out each time are randomly selected with replacement.
detrend	Whether or not a detrending procedure is desired.

### Details

Detrending is used when the distribution of residuals along estimated values produces a heteroscedastic pattern. The evaluation of such a pattern is done by fitting a linear regression of residuals as a function of estimated values ( $RES = a + b * ESTIMATE$ ). The significance of the regression is given in terms of its intercept, slope, and p-value, which are shown in the `coef` component of the list produced when applying the function with `detrend=FALSE`. If the regression is significant, it is recommended to run the analysis again using `detrend=TRUE`. The detrending is done translating the estimated values using the found regression line ( $DETRENDED\ ESTIMATE = ESTIMATE + (a + b * ESTIMATE)$ ).

### Value

A list containing:

estimated	A matrix with three columns: observed, estimated and error. If <code>detrend=TRUE</code> , additional columns provide estimated error through regression line ( $a + b * ESTIMATE$ , <code>Est error</code> ), rotated estimate ( $ESTIMATE + (a + b * ESTIMATE)$ , <code>Rotated</code> ), translated error ( $RES + MEAN(RES)$ , <code>Translated</code> ), and detrended error ( $OBSERVED - TRANSLATED\ ESTIMATE$ , <code>Det error</code> ).
error	A matrix displaying squared sum of errors ( <code>sse</code> ), and root squared mean error ( <code>rmse</code> ). If <code>detrend=TRUE</code> , these two parameters are also provided for detrended estimation.
coef	A matrix containing intercept, slope and p-value of the residuals vs estimates regression.
transl	The average error that is added or subtracted to each estimate. Only relevant if <code>detrend=TRUE</code> .

If `detrend=FALSE`, a figure shows residuals distribution along estimated (upper panel) and observed values (lower panel). If `detrend=TRUE`, a figure shows residuals distribution along estimates (upper panel) and detrended estimates (lower panel).

**Author(s)**

Alexander Correa-Metrio.

**References**

Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene

**See Also**

[synthetic](#), [fossil.dist](#), [fossil.values](#), [syas](#).

**Examples**

```
data(modernq, fossilq)
# Calculate percentages
perq<-percenta(modernq,first=2,last=39)[,2:55]
# filter data set to include only samples with at least 0.5
# percent in 20 percent of the samples
perq1<-filter.p(perq,presen=0.5,persist=0.2)$filtered
# calculate alpha and degree for each taxon through AIC
a.d<-akaike.all(modernq[,1],perq1)
# select taxa that have acceptable dispersion and normally
# distributed residuals in percentages and AIC matrices
perq2<-perq1[,-c(3:5,8,17,19)]
#cross-validation without detrending
bcrossv.sa(modernq[,1],perq2,a.d,delta=25,method="canberra",quant=0.05,
  trials=c(5, 0.05),detrend=FALSE)
#cross-validation with detrending
bcrossv.sa(modernq[,1],perq2,a.d,delta=25,method="canberra",quant=0.05,
  trials=c(5, 0.05),detrend=TRUE)
```

---

change

*Dissimilarities, distances and rates of change*

---

**Description**

This function calculates dissimilarity or distances between contiguous samples (timeslices), as well as rates of ecological change when a robust age model is available.

**Usage**

```
change(x, age, dca = FALSE, meth = "euclidean", bin = FALSE,
  roc = FALSE, digits = 1)
```

**Arguments**

x	A matrix with samples in rows and species in columns.
age	Vector with sample ages.
dca	Logical indicating whether or not detrended correspondence analysis scores (DCA, performed according to <a href="#">decorana</a> package <b>vegan</b> ) will be used as a basis for dissimilarity computation. If <code>dca=TRUE</code> the dissimilarity between samples is calculated as euclidean distance between the first four DCA axis scores. If <code>FALSE</code> dissimilarity is calculated according to the methods specified in <code>meth</code> .
meth	Corresponds to methods available in <code>vegdist</code> of <b>vegan</b> . Available methods are "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "morisita", "horn", "mountford", "raup", "binomial" or "chao". See <a href="#">vegdist</a> (package <b>vegan</b> ) for details.
bin	Argument of the function <a href="#">vegdist</a> (package <b>vegan</b> ) that standardizes the data into presence/absence before calculating the dissimilarity.
roc	Logical argument of whether or not the calculation of rates of change is desired.
digits	Number that specifies the digits desired for the rounded-up ages.

**Details**

Rates of change are calculated as:  $RoC[jk] = \text{vegdist}[jk]/\text{res}$ ). For further discussion on assumptions involved, see Urrego et al (2009). DCA is calculated according to [decorana](#) (package **vegan**.)

**Value**

Returns a matrix with ages and their corresponding dissimilarity, distances or RoC. As calculations between subsequent samples return  $n-1$  observations, the distance, dissimilarity, or RoC are assigned to the youngest of the two contiguous samples. When `roc=TRUE`, it also returns a vector `res` with time steps between samples.

**Author(s)**

Dunia H. Urrego, Alexander Correa-Metrio.

**References**

Urrego DH, Bush M, Silman MR, Correa-Metrio A, Ledru M-P, Mayle FE, Valencia BG (2009). Millennial-scale ecological changes in tropical South America since the Last Glacial Maximum. Past climate variability from the Last Glacial Maximum to the Holocene in South America and surrounding regions. (eds. Vimeux F, Sylvestre F, Khodri M). Springer.

**See Also**

See [vegdist](#) and [decorana](#) for details on dissimilarity indexes.



**Examples**

```

data(quexilchron,quexildepths,quexilper)
ages<-chron(quexilchron,quexildepths,max.depth=1957)$chronology[,2]
#Absolute change
change(quexilper,ages,meth="bray")
change(quexilper,ages,dca=TRUE)
#Rate of change
change(quexilper,ages,meth="bray",roc=TRUE)
change(quexilper,ages,dca=TRUE,roc=TRUE)

```

---

chron *Age-depth computation*

---

**Description**

This function computes and returns interpolated ages for specified sample depths based on a matrix with depths and calibrated ages.

**Usage**

```

chron(age, depths, max.depth, type = "l", linear = TRUE,
      method = "fmm", ci = TRUE, dates = "points", length = 0.05,
      digits = 2)

```

**Arguments**

age	A numeric matrix with three columns: depths of dated intervals, calibrated ages and standard deviation. The date or suspected date of the uppermost sample MUST be provided (usually 0 cm are present).
depths	A vector with a list of depths for which ages will be interpolated. The depths MUST be in ascendant order. Repeated depths are not allowed (see details).
max.depth	Maximum depth of the stratigraphic sequence (including analyzed and dated depths). It is used to extrapolate ages when the bottom of the sequence goes beyond the deepest dated interval.
type	Graphic parameter for the output plot. See <a href="#">plot</a> for details.
linear	Interpolation method desired. If TRUE, ages are calculated using linear interpolation. If FALSE, a spline is fitted to the dataset.
method	Argument of the function <a href="#">spline</a> . This must be one of "fmm", "natural", "periodic" or "monoH.FC". Only relevant if linear=FALSE. See <a href="#">spline</a> for details.
ci	A logical value indicating whether confidence intervals are desired in the output plot.
dates	Graphical argument. "points" to display dates as red points, or "bars" to display dates as error bars spanning the provided standard deviation of calibrated ages.
length	Graphical parameter. Length in inches for the horizontal bar heads.
digits	Number of digits desired for interpolated ages.

**Details**

chron calculates interpolated ages using linear interpolation or spline as method. It is based on a three-column matrix with depth of dated interval, calibrated age, and calibrated standard deviation. The interpolated ages are calculated for a set of depths (defined by the user in depths). Whenever the stratigraphic sequence goes beyond the last dated point, the age for the bottom samples is extrapolated. max.depth is the maximum depth of the stratigraphic sequence, including dates and sampled points.

Confidence intervals correspond to the provided chronology deviation. Therefore, it does not necessarily correspond to 0.95 interval.

If there is need to repeat a depth (e.g the floor and roof of a hiatus), the two numbers must be differentiated by at least decimal positions.

**Value**

A list containing:

base	A matrix with the initial chronology.
chronology	A matrix with interpolated ages up to max.depth and with deviation if ci=TRUE. The deviation is calculated according to the error provided by the user, therefore it does not necessarily correspond to 0.95 intervals.

**Author(s)**

Alexander Correa-Metrio, Dunia H. Urrego.

**See Also**

See [plot](#) for details on graphic parameters, [approx](#) for details on linear interpolation, and [spline](#) for details on non-linear interpolation.

**Examples**

```
data(quexilchron,quexildepths)
#linear interpolation
chron(quexilchron,quexildepths,max.depth=1957,dates="points")
# Spline estimation
chron(quexilchron,quexildepths,max.depth=1957,dates="points")
```

---

dispersion.all

*Taxon response to environmental variables through LOESS regression*

---

**Description**

This function generates a family of responses of 1 or several taxa to an environmental gradient. A bootstrap procedure is performed to calculate the dispersion of the species response.

**Usage**

```
dispersion.all(x, y, z, delta, trials = c(100, 0.25),
              conf = c(0.025, 0.975), outfile = "Dispersion.pdf")
```

**Arguments**

x	A vector containing the environmental gradient value for each sample or location.
y	A matrix containing the observed taxa abundances along the environmental gradient. Taxa are in columns and samples in rows.
z	A matrix containing the parameters for the LOESS regression of each species ( $\alpha$ and degree in that order). Objects produced by <a href="#">bcrossv.all</a> and <a href="#">akaike.all</a> could be used as z.
delta	Increments for the loess regression prediction. See <a href="#">predict.loess</a> for details.
trials	Values for number of repetitions of the cross-validation (100 by default), and proportion of observations to be left out each time the CV is repeated (0.25 by default). The observations left out each time are randomly selected with replacement.
conf	Quantiles for the confidence intervals of the LOESS curves.
outfile	The name of a file (format pdf) where the resulting plots will be saved. The file is saved in the working directory.

**Details**

The response of each taxa to the environmental gradient is calculated through a LOESS regression (Cleveland and Devlin, 1988). To capture the variability the LOESS is recalculated as many times as indicated in the first component of `trials` leaving out a proportion equal to the second component of `trials`. The LOESS regression is used to predict the response of each taxon in a sequence that goes through the range of the observed environmental gradient in increments equal to `delta`. For each `delta` the minimum, maximum, average and `conf` quantiles are calculated. The coefficient of determination ( $R^2$ ) is calculated as in linear regression, where  $R^2 = SS_{\text{regression}} / SS_{\text{total}}$ .  $SS_{\text{regression}} = \sum(\text{ESTIMATE} - \text{mean}(\text{OBSERVED}))^2$ , and  $SS_{\text{total}} = \sum(\text{OBSERVED} - \text{mean}(\text{OBSERVED}))^2$ .

**Value**

The function returns plots of taxa responses to the environmental gradient. The plots are saved in file named after `outfile` in pdf format. The total variation detected through the CV is signaled by the shaded area. Dashed lines indicate the confidence intervals, and a continuous line indicates the mean response of the taxon to the environmental gradient. Filled diamonds represent data points, whereas red crosses represent data points with percentage values higher than those displayed. Additionally, Q-Q plots of the LOESS regression residuals of each taxon are saved in the same file. Solid line represents the ideal residual distribution (theoretical quantiles equal to observed quantiles), and dashed line is built as a line passing through the first and third observed quartiles.

Additionally, a list is returned containing:

limits	A list containing the mean response of each taxon to the environmental gradient, and the limits of its dispersion.
--------	--

coef.det      A matrix containing the sum of squares of the LOESS regression (SS regression), the total sum of squares (SS total), and the coefficient of determination (R2) for each taxon.

### Author(s)

Alexander Correa-Metrio, Kenneth R. Cabrera.

### References

Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene  
 Cleveland, W.S., and S.J. Devlin. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83: 596-610.

### See Also

[loess](#) for details on LOESS regression. [bcrossv.all](#) and [akaike.all](#) for details related to z.

### Examples

```
data(modernnq)
# Calculate percentages
perq<-percenta(modernnq,first=2,last=39)[,2:55]
# filter data set to include only samples with at least 0.5
# percent in 20 percent of the samples
perq1<-filter.p(perq,presen=0.5,persist=0.2)$filtered
# calculate alpha and degree for each taxon through AIC
a.d<-akaike.all(modernnq[,1],perq1)
# evaluate dispersion and residuals for each loess regression
dispersion.all(modernnq[,1],perq1,a.d,trials=c(50,0.1),delta=25)
```

---

error.l

*LOESS regression error estimation for multiple taxa*

---

### Description

This function applies bootstrap crossvalidation (CV) of LOESS regressions of multiple taxa. The regression parameters for each taxon are previously selected.

### Usage

```
error.l(x, y, z, trials = c(100, 0.25))
```

**Arguments**

x	A vector containing the environmental gradient value for each sample or location.
y	A matrix containing the observed taxa abundances along the environmental gradient. Taxa are in columns and samples in rows.
z	A matrix containing the parameters for the LOESS regression of each taxon ( $\alpha$ and degree in that order). Objects produced by <a href="#">bcrossv.all</a> and <a href="#">akaike.all</a> could be used as z.
trials	Values for number of repetitions of the crossvalidation (100 by default), and proportion of observations to be left out each time the CV is repeated. The observations left out each time are randomly selected with replacement.

**Details**

See [bcrossv.11](#) for details.

**Value**

A matrix of 6 columns showing error diagnostics for each taxon LOESS regression (root of the predictive squared error (rse), and root of the mean squared error (rmse) see [bcrossv.11](#) for details). The proportion of the error relative to the data range and mean is presented in the last two columns.

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera.

**References**

Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene  
 Cleveland, W.S., and S.J. Devlin. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83: 596-610.

**See Also**

[loess](#) and [bcrossv.11](#)

**Examples**

```
data(modernq)
# Calculate percentages
perq<-percenta(modernq,first=2,last=39)[,2:55]
#if alpha and degree are selected using AIC
a.d<-akaike.all(modernq[,1],perq[,1:10])
error.l(modernq[,1],perq[,1:10],a.d)
```

---

 filter.p

---

*Filter for Presence and Persistence*


---

### Description

This function reduces a given dataset based on filters for minimum presence (abundance) AND minimum persistence (number of samples), or maximum persistence.

### Usage

```
filter.p(x, rare=TRUE, presen = 1, persist = 0.05)
```

### Arguments

x	Matrix or dataframe with taxa in columns and samples in rows.
rare	Argument indicating if filter aims rare or common taxa. If rare=TRUE, the filter eliminates rare taxa with presence and persistence indicated in the respective arguments. If rare=FALSE, the filter eliminates taxa with a persistence higher than that indicated in the argument persist.
presen	Criterion for minimum percentual presence, 1 percent by default. Only relevant if rare=TRUE.
persist	Criterion for minimum persistence as a fraction of the number of samples where the taxon is expected to occur, 0.05 of the total number of samples by default.

### Details

This function applies both the presence and persistence filters when rare=TRUE. If the user desires to apply only one of the filters at a given time, a criterion that is met by all elements (taxa) should be chosen, e.g. persist=0. If rare=FALSE, only the criterion of being under the given persistence threshold is applied.

### Value

Returns a list with three components

filtered	Reduced dataset after both filters are applied.
filter	This component is returned only when rare=TRUE. Matrix with three columns: n, number of samples where taxon is present; n over minimum presence, number of samples where percentage is greater than the defined minimum or filter; quality, binary that lets the user know whether or not a taxon meets the filter criteria.
result	List with two or three components: percentage, minimum presence, only relevant when rare=TRUE; minimum or maximum, minimum or maximum persistence; and number of taxa, total number of taxa that meet the filter criteria.

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera, Dunia H. Urrego.

**References**

Correa-Metrio, A., K.R. Cabrera, and M.B. Bush. 2010. Quantifying ecological change through discriminant analysis: a paleoecological example from the Peruvian Amazon. *Journal of Vegetation Science* 21: 695-704.

**Examples**

```
#For a minimum presence of 2 percent in 20 percent of the samples
data(quexilper)
filter.p(quexilper)
quexilfil<-filter.p(quexilper,presen=2,persist=0.2)
#Filtered database
quexilfil$filtered
```

---

fossil.dist	<i>Distance calculation between fossil samples and modern synthetic assemblages</i>
-------------	---

---

**Description**

This function calculates the distance between each fossil sample and the synthetic assemblages, according to the selected distance method.

**Usage**

```
fossil.dist(modern, fossil, method = "canberra")
```

**Arguments**

modern	An object containing the synthetic assemblages produced by the function <code>synthetic</code> .
fossil	A matrix containing the fossil pollen percentages with samples in rows and taxa in columns.
method	Corresponds to methods available in <code>vegdist</code> of <b>vegan</b> . Available methods are "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "morisita", "horn", "mountford", "raup", "binomial" or "chao". See <a href="#">vegdist</a> (package <b>vegan</b> ) for details.

**Value**

A matrix containing the distance or dissimilarity between fossil samples and each one of the modern synthetic assemblages. Fossil samples are in rows and modern synthetic assemblages are in columns. The first row contains the values of the environmental gradient associated with each synthetic assemblage.

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera.

**References**

Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene

**See Also**

See [synthetic](#) for details on the construction of modern synthetic assemblages. See [vegdist](#) for details on dissimilarity indices.

**Examples**

```
data(modernq, fossilq)
# Calculate percentages
perq<-percenta(modernq,first=2,last=39)[,2:55]
# filter data set to include only samples with at least 0.5
# percent in 20 percent of the samples
perq1<-filter.p(perq,presen=0.5,persist=0.2)$filtered
# calculate alpha and degree for each taxon through AIC
a.d<-akaike.all(modernq[,1],perq1)
# select taxa that have acceptable dispersion and normally
# distributed residuals in percentages and AIC matrices
perq2<-perq1[,-c(3:5,8,17,19)]
a.d1<-a.d[-c(3:5,8,17,19),]
#build synthetic assemblages
syas.q<-synthetic(modernq[,1],perq2,a.d1,delta=25,plot=TRUE,
percen=rep(100,53))
#calculate distance from synthetic assemblages to each fossil sample
fossil.dist(syas.q,fossilq)
```

---

fossil.values

*Fossil environmental parameters calculation*

---

**Description**

This function recovers the values of the fossil environmental parameters by evaluating the distance of each fossil sample to each modern synthetic assemblage.

**Usage**

```
fossil.values(distan, age, quant = 0.05, detrend = FALSE, crossv)
```



**Arguments**

distan	A matrix containing the distance of each fossil sample to each modern synthetic assemblage, produced by <code>fossil.dist</code> .
age	A vector containing the age or depth of each fossil sample.
quant	The quantile of the distances that will be used to obtain the mean environmental parameter. See details below.
detrend	Whether or not a detrending procedure is desired. See <code>bcrossv.sa</code> for details.
crossv	An object produced by <code>bcrossv.sa</code> . Only relevant when <code>detrend=TRUE</code> .

**Details**

The fossil environmental value is assumed to be the value associated with the modern synthetic assemblage that displays the least distance to the fossil sample. To counteract possible bias produced by local resemblances, the value is calculated as the average among the values associated with the synthetic assemblages that display distances within the quantile defined in `quant`.

**Value**

A matrix with ages or depths in the first column and values of the environmental parameter in the second column.

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera.

**References**

Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene

**See Also**

[bcrossv.sa](#)

**Examples**

```
data(modernnq, fossilq)
# Calculate percentages
perq<-percenta(modernnq,first=2,last=39)[,2:55]
# filter data set to include only samples with at least 0.5
# percent in 20 percent of the samples
perq1<-filter.p(perq,presen=0.5,persist=0.2)$filtered
# calculate alpha and degree for each taxon through AIC
a.d<-akaike.all(modernnq[,1],perq1)
# select taxa that have acceptable dispersion and normally
# distributed residuals in percentages and AIC matrices
perq2<-perq1[,-c(3:5,8,17,19)]
a.d1<-a.d[-c(3:5,8,17,19),]
#build synthetic assemblages
```

```
syas.q<-synthetic(modernq[,1],perq2,a.d1,delta=25,plot=TRUE,
percen=rep(100,53))
#calculate distance from synthetic assemblages to each fossil sample
distances<-fossil.dist(syas.q,fossilq)
#calculate the fossil values of the environmental variable
# without detrending
fossil.values(distances,fossilq[,1],quant=0.05,detrend=FALSE)
#if detrending is desired, the detrending parameters have to be obtained
crossv.q<-bcrossv.sa(modernq[,1],perq2,a.d1,delta=25,method="canberra",
quant=0.05,trials=c(5, 0.05),detrend=TRUE)
#calculate the fossil values of the environmental variable
# with detrending
fossil.values(distances,fossilq[,1],quant=0.05,detrend=TRUE,
crossv=crossv.q)
```

---

fossilq

*Pollen percentages Aluik Pond*

---

### Description

This dataset provides pollen percentages of a core from Aluik Pond, Labrador, Canada (Jordan, 1975).

### Usage

```
data(fossilq)
```

### Format

A data frame with 9 rows (samples) and 25 columns (age and taxa).

### Source

<http://www.lpc.uottawa.ca/data/cpd>

### References

Jordan, R.H. 1975. Pollen diagrams from Hamilton Inlet, central Labrador, and their environmental implications for the northern maritime Archaic. *Arctic Anthropology* 12: 92-116.

### Examples

```
data(fossilq)
dim(fossilq)
```

---

`modernq`*Pollen counts Northern Quebec Treeline*

---

**Description**

This data set contains pollen counts for surface samples of 53 lakes in Northern Quebec, Canada (Gajewski, 1991).

**Usage**

```
data(modernq)
```

**Format**

A data frame with 53 observations (lakes) and 55 variables. The first column contains the elevation of each location, and the other 54 the pollen counts for the taxa found in the samples.

**Source**

<http://www.lpc.uottawa.ca/data/cpd>

**References**

Gajewski, K. 1991. Representation pollinique a la limite arbres au Nouveau-Quebec. Canadian Journal of Earth Sciences 25: 643-648.

**Examples**

```
data(modernq)
dim(modernq)
```

---

`percenta`*Computation of percentage matrix*

---

**Description**

This function calculates percentage matrices based on a dataframe with species in columns and samples in rows.

**Usage**

```
percenta(x, first, last)
```

### Arguments

<code>x</code>	Dataframe or matrix containing raw data with species in columns and samples in rows.
<code>first</code>	Number of first column in <code>x</code> (taxon) to be included in the sum that serves as basis for the percentage calculation (e.g. pollen sum, see Birks & Birks (1980)).
<code>last</code>	Number of last column in <code>x</code> (taxon) to be included in the sum that serves as basis for the percentage calculation (e.g. pollen sum, see Birks & Birks (1980)).

### Details

Percentages can be calculated based on a selected group of columns or taxa (e.g. terrestrial taxa included in the pollen sum for palynological analyses, see Birks & Birks (1980) for illustration on pollen sum). For the rest of the columns (taxa) percentages are calculated based on totals from included columns (taxa), e.g. spores and aquatics.

The dataframe should be organized in such a way that taxa or columns to be included in the percentage calculation are contiguous. Percentages of other columns outside that range (first to last) are calculated based on the total sum of elements included (e.g. pollen sum).

### Value

The function returns a dataframe containing calculated percentages, with species in columns and samples in rows.

### Author(s)

Alexander Correa-Metrio, Dunia H. Urrego.

### References

Birks, H. J. B., and Birks, H. H. (1980). Quaternary Palaeoecology. University Park Press, Baltimore.

### See Also

[scale](#)

### Examples

```
pollen<-matrix(nrow=10,sample(c(1:100),50))
colnames(pollen)<-c("Sp1","Sp2","Sp3","Sp4","Sp5")
percenta(pollen,first=1,last=5)
```

---

quexilchron

*Chronology Lake Quexil*

---

### Description

This data set provides the age basis for the chronology of Core 80-1 from Lake Quexil, Guatemala (Leyden 1984; Leyden et al 1993, 1994).

### Usage

```
data(quexilchron)
```

### Format

A matrix with 3 rows (dates) and 3 columns (depth, age and uncertainty). Depth 0 cm is assumed to be modern.

### Source

<http://www.ncdc.noaa.gov/paleo>

### References

Leyden, B.W. 1984. Guatemalan forest synthesis after Pleistocene aridity. Proceedings of the National Academy of Sciences USA 81:4856-4859.

Leyden, B.W., M. Brenner, D.A. Hodell, and J.H. Curtis. 1993. Late Pleistocene climate in the Central American lowlands. Pages 165-178 in P.K. Swart, K.C. Lohmann, J. McKenzie, and S. Savin, editors. Climate change in continental isotopic records. American Geophysical Union Geophysical Monograph 78, Washington, DC, USA.

Leyden, B.W., M. Brenner, D.A. Hodell, and J.H. Curtis. 1994. Orbital and internal forcing of climate on the Yucatan Peninsula for the past ca. 36 ka. Palaeogeography, Palaeoclimatology, Palaeoecology 109:193-210.

### Examples

```
data(quexilchron)
plot(quexilchron[,1:2], type="l")
```

---

quexildepths

*Depths Lake Quexil*

---

### **Description**

This data set provides the depths analyzed for pollen of Core 80-1 from Lake Quexil, Guatemala (Leyden 1984; Leyden et al 1993, 1994).

### **Usage**

```
data(quexildepths)
```

### **Format**

A numeric vector of length 68.

### **Source**

<http://www.ncdc.noaa.gov/paleo>

### **References**

Leyden, B.W. 1984. Guatemalan forest synthesis after Pleistocene aridity. Proceedings of the National Academy of Sciences USA 81:4856-4859.

Leyden, B.W., M. Brenner, D.A. Hodell, and J.H. Curtis. 1993. Late Pleistocene climate in the Central American lowlands. Pages 165-178 in P.K. Swart, K.C. Lohmann, J. McKenzie, and S. Savin, editors. Climate change in continental isotopic records. American Geophysical Union Geophysical Monograph 78, Washington, DC, USA.

Leyden, B.W., M. Brenner, D.A. Hodell, and J.H. Curtis. 1994. Orbital and internal forcing of climate on the Yucatan Peninsula for the past ca. 36 ka. Palaeogeography, Palaeoclimatology, Palaeoecology 109:193-210.

### **Examples**

```
data(quexildepths)  
range(quexildepths)
```

---

quexilper

*Pollen percentages Lake Quexil*

---

### **Description**

This data set provides pollen percentages of Core 80-1 from Lake Quexil (Leyden 1984; Leyden et al 1993, 1994).

### **Usage**

```
data(quexilper)
```

### **Format**

A matrix with 68 rows (samples) and 76 columns (taxa).

### **Source**

<http://www.ncdc.noaa.gov/paleo>

### **References**

Leyden, B.W. 1984. Guatemalan forest synthesis after Pleistocene aridity. *Proceedings of the National Academy of Sciences USA* 81:4856-4859.

Leyden, B.W., M. Brenner, D.A. Hodell, and J.H. Curtis. 1993. Late Pleistocene climate in the Central American lowlands. Pages 165-178 in P.K. Swart, K.C. Lohmann, J. McKenzie, and S. Savin, editors. *Climate change in continental isotopic records*. American Geophysical Union Geophysical Monograph 78, Washington, DC, USA.

Leyden, B.W., M. Brenner, D.A. Hodell, and J.H. Curtis. 1994. Orbital and internal forcing of climate on the Yucatan Peninsula for the past ca. 36 ka. *Palaeogeography, Palaeoclimatology, Palaeoecology* 109:193-210.

### **Examples**

```
data(quexilper)  
colnames(quexilper)
```

---

run.mean *Moving averages*

---

### Description

This function calculates moving averages for data series.

### Usage

```
run.mean(x, window)
```

### Arguments

x	A vector with the data series.
window	Number of observations to be taken into account in the estimation of the local mean.

### Value

Returns a vector with the estimated local averages.

### Author(s)

Alexander Correa-Metrio.

### Examples

```
run.mean(c(1:10),window=3)
```

---

simulat *Sample simulation*

---

### Description

This function simulates new samples based on known percentages

### Usage

```
simulat(x, pop = 10000, nsamples = 10, ssample = 300)
```

### Arguments

x	Relative frequencies of taxa in the sample of interest. It should correspond to a row in a matrix where species are in columns and samples in rows.
pop	Size of the population to be simulated.
nsamples	Number of samples to be drawn.
ssample	Sample size.



**Details**

A population of size `pop` and relative frequencies `x` is simulated, and randomly sampled to produce `nsamples` new samples. This procedure assumes that the percentages contained in `x` represent the real population.

**Value**

Returns a matrix `samples` that contains simulated samples in rows and species in columns.

**Author(s)**

Alexander Correa-Metrio, Kenneth R. Cabrera.

**References**

Correa-Metrio, A., K.R. Cabrera, and M.B. Bush. 2010. Quantifying ecological change through discriminant analysis: a paleoecological example from the Peruvian Amazon. *Journal of Vegetation Science* 21: 695-704.

**Examples**

```
# data from lake Quexil (Leyden et al 1994)
data(quexilper)
simulat(quexilper[1,])
```

---

simulat.t

*Model desaturation by samples simulation*

---

**Description**

This function simulates new samples for a sequence of fossil data to desaturate models.

**Usage**

```
simulat.t(x, pop = 1000, nsamples = 10, ssample = 300,
  percenta = TRUE, last)
```

**Arguments**

<code>x</code>	A percentage matrix containing samples in rows and taxa in columns.
<code>pop</code>	Size of the population to be simulated.
<code>nsamples</code>	Number of samples to be drawn.
<code>ssample</code>	Sample size.
<code>percenta</code>	Logical indicating whether or not the outcome matrix should be one of percentages.
<code>last</code>	If <code>percenta=TRUE</code> , <code>last</code> refers to the column where the last species to be taken into account is located. Only relevant if <code>percenta=TRUE</code> . See <a href="#">percenta</a> for details.

## Details

When all species are taken into the pollen count, last corresponds to the number of columns. If only some species are to be taken into the pollen count, they must be contiguous and start in the first column.

## Value

When percenta=FALSE, samples.t is returned. When percenta=TRUE an additional element percentages is returned.

samples.t	An absolute abundance matrix with samples in rows and species in columns. The first column contains sample number.
percentages	A relative abundance matrix with samples in rows and species in columns. Each species is repeated nsamples times.

## Author(s)

Alexander Correa-Metrio, Kenneth R. Cabrera.

## References

Correa-Metrio, A., K.R. Cabrera, and M.B. Bush. 2010. Quantifying ecological change through discriminant analysis: a paleoecological example from the Peruvian Amazon. *Journal of Vegetation Science* 21: 695-704.

## See Also

This function is an iteration of [simulat](#)

## Examples

```
data(quexilper)
colnames(quexilper)
# Cyperaceae and spores are left out of the pollen sum.
# For absolute abundance data
simulat.t(quexilper,percenta=FALSE,last=73)
# or
simulat.t(quexilper,last=73)$samples.t
# For percentages data
simulat.t(quexilper,last=73)$percentages
```

syas

*Synthetic assemblages***Description**

This function performs all the procedures for synthetic assemblages construction, and estimation of fossil environmental parameters.

**Usage**

```
syas(x, y, z, fossil, age, delta, percen, method = "canberra",
     quant = 0.05, plot = TRUE, window = 3, detrend = FALSE,
     crossv, maximum = 0.99, classes = 200, colors = c("red", "blue"),
     line = TRUE, ylabel = "Age", xlabel = "Parameter")
```

**Arguments**

x	A vector containing the environmental gradient value for each sample or location.
y	A matrix containing the observed taxa abundances along the environmental gradient. Taxa are in columns and samples in rows. Taxa included in this analysis should be selected by evaluating dispersion and residuals distribution through the function <a href="#">dispersion.all</a> .
z	A matrix containing the parameters for the LOESS regression of each species ( $\alpha$ and degree in that order). Objects produced by <a href="#">bcrossv.all</a> and <a href="#">akaike.all</a> could be used as z.
fossil	A matrix containing the fossil pollen percentages with samples in rows and taxa in columns.
age	A vector containing the age or depth of each fossil sample.
delta	Increments for the LOESS regression prediction. See <a href="#">predict.loess</a> for details.
percen	A vector containing total percentage of pollen in each modern sample. See <a href="#">synthetic</a> for details.
method	Corresponds to methods available in <code>vegdist</code> of <b>vegan</b> . Available methods are "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "morisita", "horn", "mountford", "raup", "binomial" or "chao". See <a href="#">vegdist</a> (package <b>vegan</b> ) for details.
quant	The quantile of the distances that will be used to obtain the mean environmental parameter.
plot	Logical indicating whether or not a plot of the fossil reconstruction is desired.
window	The plot is done using a moving average along the estimated values to avoid the excess of noise. This parameter refers to the size of the moving average. Only relevant if <code>plot=TRUE</code> .

detrend	Logical indicating whether or not a detrending procedure is desired. See <code>bcrossv.sa</code> for details.
crossv	An object produced by <code>bcrossv.sa</code> , only relevant when <code>detrend=TRUE</code> .
maximum	Percentage of data used to build the color tables, as extreme high distance values can make all values fall into low classes, showing only two colors. Only relevant if <code>plot=TRUE</code> .
classes	Number of classes in the color ramp. See <code>colorRamp</code> for details. Only relevant if <code>plot=TRUE</code> .
colors	Color to interpolate for the color ramp. See <code>colorRamp</code> for details. Only relevant if <code>plot=TRUE</code> .
line	Logical indicating whether or not a line showing the estimated values is desired. Only relevant if <code>plot=TRUE</code> .
ylabel	Label for y axis, which is depth or age. Only relevant if <code>plot=TRUE</code> .
xlabel	Label for x axis, which corresponds to the environmental variable. Only relevant if <code>plot=TRUE</code> .

### Value

A list containing:

syn.as	A list containing the elements <code>loess.f</code> and <code>predicted</code> produced by the function <code>synthetic</code> .
values	Fossil environmental reconstruction values. If <code>detrend=TRUE</code> , both estimates and detrended estimates are returned in this component.

Additionally, if `plot=TRUE` a plot of the environmental estimation through time or depth is produced.

### Author(s)

Alexander Correa-Metrio.

### References

Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene

### See Also

See `synthetic`, `fossil.dist`, and `fossil.values`, for details on the fossil reconstruction through synthetic assemblages. See `bcrossv.sa` for details on detrending procedure and error estimation.

### Examples

```
data(modernnq, fossilq)
# Calculate percentages
perq<-percenta(modernnq, first=2, last=39)[,2:55]
# filter data set to include only samples with at least 0.5
```

```

# percent in 20 percent of the samples
perq1<-filter.p(perq,presen=0.5,persist=0.2)$filtered
# calculate alpha and degree for each taxon through AIC
a.d<-akaike.all(modernq[,1],perq1)
# select taxa that have acceptable dispersion and normally
# distributed residuals in percentages matrix
perq2<-perq1[,-c(3:5,8,17,19)]
#cross-validation of the method
bcrossv.sa(modernq[,1],perq2,a.d,delta=25,method="canberra",quant=0.05,trials=c(5, 0.05),detrend=FALSE)
#synthetic assemblages fossil estimation without detrending
syas(modernq[,1],perq2,a.d,fossilq,fossilq[,1],delta=25,percen=rep(100,53),
plot=TRUE>window=3,detrend=FALSE,xlabel="Elevation (m asl)")
#synthetic assemblages fossil estimation with detrending
#first, estimation of the detrending parameters
crossv.q<-bcrossv.sa(modernq[,1],perq2,a.d,delta=25,method="canberra",
quant=0.05,trials=c(5, 0.05),detrend=TRUE)
#second, application of the synthetic assemblages using the
# results of the cross-validation for the detrending.
syas(modernq[,1],perq2,a.d,fossilq,fossilq[,1],delta=25,percen=rep(100,53),
plot=TRUE,detrend=TRUE,crossv=crossv.q,xlabel="Elevation (m asl)")

```

---

synthetic

*Calculation of synthetic assemblages*


---

### Description

This function calculates the ideal pollen assemblages associated with a series of values of a given environmental factor.

### Usage

```

synthetic(x, y, z, delta, plot = FALSE, percen, upmargin = 0.1,
widths = c(1, 4), steps = 0.05)

```

### Arguments

x	A vector containing the environmental gradient value for each sample or location.
y	A matrix containing the observed taxa abundances along the environmental gradient. Taxa are in columns and samples in rows. Taxa included in this analysis should be selected by evaluating dispersion and residuals distribution through the function <a href="#">dispersion.all</a> .
z	A matrix containing the parameters for the LOESS regression of each species ( $\alpha$ and degree in that order). Objects produced by <a href="#">bcrossv.all</a> and <a href="#">akaike.all</a> could be used as z.
delta	Increments for the LOESS regression prediction. See <a href="#">predict.loess</a> for details.
plot	Logical indicating whether or not a plot of the synthetic assemblages is desired.

percen	A vector containing total percentage of pollen in each modern sample. As some taxa might have been excluded from the pollen sum, this value could be larger than 100. See details below.
upmargin	A number indicating the margin size for the upper part of each panel. <code>par</code> function, <code>mai</code> parameter.
widths	A vector of values for the widths of columns in the plotting device. <code>layout</code> parameter.
steps	The proportion of the range to draw lines separating each synthetic assemblage.

### Details

When applying the synthetic assemblages technique, it is desirable to know the proportion of data from each sample that was used in the the construction of the modern ideal assemblages. The `avg.info` returns the proportion of pollen data that was used from each sample. If no taxa are excluded from the pollen sum, such proportion will coincide with the sum of the percentages of the taxa included. In this case, the argument `percent` can be filled by repeating 100 as many times as number of of modern samples (`percent = rep(100, nrow(y))`). When one or more taxa are excluded from the pollen sum, samples might have total percentages that are larger than 100. Therefore, the `percent` has to be calculated by adding all the percentages in each sample (`percent = apply (Y, 1, sum)`, where `Y` is the original percentage matrix including all taxa).

### Value

This function returns a list containing:

<code>loess.f</code>	A list with the LOESS regression for each taxon.
<code>predicted</code>	A matrix containing the predicted percentages of each taxon along the evaluated environmental gradient. The first column contains the values of the environmental variable, and taxa are contained in the subsequent columns.
<code>avg.info</code>	if <code>plot=TRUE</code> , this value is the average pollen percentage used from each sample in the construction of the synthetic assemblages.

Additionally, if `plot=TRUE`, a figure containing the synthetic assemblages is produced. Given differences in pollen representativity among species, individual taxa are standardized to facilitate illustration.

### Author(s)

Alexander Correa-Metrio, Kenneth R. Cabrera.

### References

Correa-Metrio, A., M.B. Bush, L.Perez, A. Schwalb, and K. Cabrera. Accepted. Pollen distribution along climatic and biogeographic gradients in northern Central America. The Holocene

### See Also

[par](#) and [layout](#) for details on graphic parameters.

**Examples**

```

data(modernq)
# Calculate percentages
perq<-percenta(modernq,first=2,last=39)[,2:55]
# filter data set to include only samples with at least 0.5
# percent in 20 percent of the samples
perq1<-filter.p(perq,presen=0.5,persist=0.2)$filtered
# calculate alpha and degree for each taxon through AIC
a.d<-akaike.all(modernq[,1],perq1)
# select taxa that have acceptable dispersion and normally
# distributed residuals in percentages and AIC matrices
perq2<-perq1[,-c(3:5,8,17,19)]
a.d1<-a.d[-c(3:5,8,17,19),]
#build synthetic assemblages
syas.q<-synthetic(modernq[,1],perq2,a.d1,delta=25,plot=TRUE,
percen=rep(100,53))
# predic percentages of Cyperacea for an elevation sequence from
# 100 to 550 in 25-m increments
predict(syas.q[[1]]$Cyperaceae,seq(100,550,25))

```

vcrossv.all

*V-fold iterative cross-validation for discriminant analysis***Description**

This function v-fold cross-validates a discriminant analysis through the leave-v-out procedure, with v varying from 1 to v. It also does repetitions of the cross-validation at each value of v to make estimates of the confidence limits for the accuracy of the function. This function involves very intensive computations. Therefore, if only specific values of v need to be evaluated, it is recommended to use vcrossv.da instead.

**Usage**

```
vcrossv.all(x, f, to, nsimulat, funct, ntrials, plot = TRUE)
```

**Arguments**

x	A matrix with samples in columns and taxa in rows. The rows must be named after taxa names (see <a href="#">rownames</a> ).
f	An object of class factor containing the discriminant factor (See Venables & Ripley (2002) for details on discriminant analysis).
to	The upper value of v. The v-fold crossvalidation is performed for each value from 1 to v.
nsimulat	Number of samples simulated to desaturate the model (see Correa-Metrio et al (2010) for details). If no samples were simulated nsimulat=1.
funct	lda for linear discriminant analysis, and qda for quadratic discriminant analysis.
ntrials	Number of desired repetitions for the cross-validation at each value of v.

**plot** Whether or not a plot of the behavior of the accuracy estimated for the discriminant function at each value of  $v$  is desired.

### Details

The function was designed for discrimination of pollen taxa into dichotomous ecological groups (only admits two factors). The prior information corresponds to the affinity of certain taxa to known environmental conditions. Therefore, while the taxa correspond to the objects to classify, the percentages through the fossil dataset correspond to the attributes. Each time the discriminant function is adjusted,  $v$  elements are left out with no replacements. Therefore, it is recommended that  $v$  be smaller than half of the total taxa, unless there is a considerable number of species. Take also into consideration that each time a taxon is left out for the crossvalidation, all the samples that were simulated for such taxon are left out too.

### Value

`vcrossv.all` returns a matrix with four columns. `fold` contains the values of  $v$ . `mean` accuracy contains the average discriminant function accuracy obtained from repeating the cross-validation `ntrials` times at the given value of  $v$ . `lower` (0.025) and `upper` (0.975) contain the 0.025 and 0.975 quantiles of the discriminant function accuracy obtained from the same procedure. Note that for  $v=1$  the results are the same for all repetitions given that leaving only one element out has no random component associated.

### Author(s)

Alexander Correa-Metrio, Kenneth R. Cabrera.

### References

Correa-Metrio, A., K.R. Cabrera, and M.B. Bush. 2010. Quantifying ecological change through discriminant analysis: a paleoecological example from the Peruvian Amazon. *Journal of Vegetation Science* 21: 695-704.

Venables, W.N., and B.D. Ripley. 2002. "Modern applied statistics with S". Springer, New York.

### See Also

[vcrossv.all.lda](#) and [qda](#) (package **MASS**) for details on the discriminant functions. [simulat](#) and [simulat.t](#) for details on samples simulations.

### Examples

```
data(quexilper)
# Taking only a fraction of the data base so the model is not saturated
a<-quexilper[1:10,1:20]
a<-t(a)
#build a dummy factor assuming that the first 10 species belong to group1 and the send ten belong to group 2
b<-as.factor(rep(c("group1", "group2"),each=10))
#apply the function
vcrossv.all(a,b,to=5,nsimulat=1,funct=lda,ntrials=20,plot=TRUE)
```



---

vcrossv.da

*V-fold cross-validation for discriminant analysis*


---

### Description

This function v-fold cross-validates a discriminant analysis through the leave-v-out procedure.

### Usage

```
vcrossv.da(x, f, fold, nsimulat, funct)
```

### Arguments

x	A matrix with samples in columns and taxa in rows. The rows must be named after taxa names (see <a href="#">rownames</a> ).
f	An object of class factor containing the discriminant factor (See Venables & Ripley (2002) for details on discriminant analysis).
fold	Value of v, i.e. number of elements to be left out in each validation.
nsimulat	Number of samples simulated to desaturate the model (see Correa-Metrio et al (in review) for details). If no samples were simulated nsimulat=1.
funct	lda for linear discriminant analysis, and qda for quadratic discriminant analysis.

### Details

The function was designed for discrimination of pollen taxa into dichotomous ecological groups (only admits two factors). The prior information corresponds to the affinity of certain taxa to known environmental conditions. Therefore, while the taxa correspond to the objects to classify, the percentages through the fossil dataset correspond to the attributes. Each time the discriminant function is adjusted, v elements are left out with no replacements. Therefore, it is recommended that v be smaller than half of the total taxa, unless there is a considerable number of species. Take also into consideration that each time a taxon is left out for the crossvalidation, all the samples that were simulated for such taxon are left out too.

### Value

A list containing:

posterior	The <i>a posteriori</i> probability of each taxa belonging to each one of the defined groups.
comp2	Binary classification of the taxa.
accuracy	The percentage of cases well classified in the cross-validation.

### Author(s)

Alexander Correa-Metrio, Kenneth R. Cabrera.

## References

Correa-Metrio, A., K.R. Cabrera, and M.B. Bush. 2010. Quantifying ecological change through discriminant analysis: a paleoecological example from the Peruvian Amazon. *Journal of Vegetation Science* 21: 695-704.

Venables, W.N., and B.D. Ripley. 2002. "Modern applied statistics with S". Springer, New York.

## See Also

[vcrossv.all.lda](#) and [qda](#) (package **MASS**) for details on the discriminant functions. [simulat](#) and [simulat.t](#) for details on samples simulations.

## Examples

```
data(quexilper)
# Taking only a fraction of the data base so the model is not saturated
a<-quexilper[1:10,1:20]
a<-t(a)
# build a dummy factor assuming that the first 10 species belong
# to group1 and the send ten belong to group 2
b<-as.factor(rep(c("group1", "group2"),each=10))
#to apply ordinary crossvalidation (leave-one-out)
vcrossv.da(a,b,fold=1,nsimulat=1,funct=lda)
#to apply 3-fold cross-validation
vcrossv.da(a,b,fold=3,nsimulat=1,funct=lda)
```

# Index

## \*Topic **arith**

- analog.mult, 6
- analog.sing, 8
- bcrossv.sa, 13
- change, 15
- chron, 17
- fossil.dist, 23
- fossil.values, 24
- percenta, 27
- run.mean, 32
- syas, 35

## \*Topic **chron**

- chron, 17

## \*Topic **datagen**

- simulat, 32
- simulat.t, 33

## \*Topic **datasets**

- fossilq, 26
- modernq, 27
- quexilchron, 29
- quexildepths, 30
- quexilper, 31

## \*Topic **graphs**

- synthetic, 37

## \*Topic **manip**

- analog.mult, 6
- analog.sing, 8
- bcrossv.sa, 13
- change, 15
- chron, 17
- filter.p, 22
- fossil.dist, 23
- fossil.values, 24
- percenta, 27
- run.mean, 32
- simulat, 32
- simulat.t, 33
- syas, 35

## \*Topic **math**

- analog.mult, 6
- analog.sing, 8
- bcrossv.sa, 13
- change, 15
- chron, 17
- fossil.dist, 23
- fossil.values, 24
- syas, 35

## \*Topic **models**

- akaike.all, 3
- akaike.l, 4
- bcrossv.all, 9
- bcrossv.l, 10
- bcrossv.l1, 12
- dispersion.all, 18
- error.l, 20
- synthetic, 37
- vcrossv.all, 39
- vcrossv.da, 41

## \*Topic **multivariate**

- analog.mult, 6
- analog.sing, 8
- bcrossv.sa, 13
- change, 15
- fossil.dist, 23
- fossil.values, 24
- syas, 35

## \*Topic **package**

- paleoMAS-package, 2

## \*Topic **robust**

- akaike.all, 3
- akaike.l, 4
- bcrossv.all, 9
- bcrossv.l, 10
- bcrossv.l1, 12
- bcrossv.sa, 13
- dispersion.all, 18
- error.l, 20
- vcrossv.all, 39

- vcrossv.da, 41
- \*Topic **smooth**
  - dispersion.all, 18
  - synthetic, 37
- akaike.all, 3, 14, 19–21, 35, 37
- akaike.l, 3, 4, 4
- analog.mult, 6, 9
- analog.sing, 7, 8
- approx, 18
  
- bcrossv.all, 9, 14, 19–21, 35, 37
- bcrossv.l, 10, 10
- bcrossv.l1, 10, 11, 12, 21
- bcrossv.sa, 13, 25, 36
  
- change, 15
- chron, 17
- colorRamp, 36
  
- decorana, 16
- dispersion.all, 18, 35, 37
  
- error.l, 20
  
- filter.p, 22
- fossil.dist, 15, 23, 36
- fossil.values, 14, 15, 24, 36
- fossilq, 26
  
- layout, 38
- lda, 40, 42
- levelplot, 6
- loess, 4, 5, 10, 11, 13, 20, 21
  
- modernq, 27
  
- paleoMAS (paleoMAS-package), 2
- paleoMAS-package, 2
- par, 38
- percenta, 27, 33
- plot, 17, 18
- predict.loess, 14, 19, 35, 37
  
- qda, 40, 42
- quexilchron, 29
- quexildepths, 30
- quexilper, 31
  
- rownames, 6, 39, 41
- run.mean, 32
  
- scale, 28
- simulat, 32, 34, 40, 42
- simulat.t, 33, 40, 42
- spline, 17, 18
- syas, 15, 35
- synthetic, 13, 15, 24, 35, 36, 37
  
- vcrossv.all, 39, 40, 42
- vcrossv.da, 41
- vegdist, 6–9, 14, 16, 23, 24, 35
  
- wireframe, 6