Package ‘LOST’

July 4, 2024

Version 2.1.1
Date 2024-7-3
Title Missing Morphometric Data Simulation and Estimation
Author J. Arbour,
       C. Brown
Maintainer J. Arbour <jessica.arbour@mtsu.edu>
Imports gdata, shapes, e1071, pcaMethods, MASS, miscTools, stats, rgl,
       geomorph
Description Functions for simulating missing morphometric
       data randomly, with taxonomic bias and with anatomical bias. LOST also
       includes functions for estimating linear and geometric morphometric data.
License GPL (>= 2)
Depends R (>= 3.5.0)
NeedsCompilation no
Repository CRAN
Date/Publication 2024-07-04 02:20:02 UTC

Contents

LOST-package .......................................................... 2
align.missing .......................................................... 2
best.reg ............................................................ 3
bilat.align ........................................................ 4
byclade ............................................................... 5
complete.specimens .................................................. 6
crocs ................................................................. 7
crocs.landmarks ..................................................... 8
dacrya ................................................................. 8
darters ............................................................... 9
est.reg ............................................................... 10
flipped ............................................................... 11
how.many.missing .................................................... 12
Description

LOST includes functions for simulating missing morphometric data randomly, with taxonomic bias and with anatomical bias as described by Brown et al. 2012. This package also includes functions for estimating missing morphometric data based on regression analysis and a function for checking the percentage of missing data in a matrix.

Author(s)

J. Arbour and C. Brown

Maintainer: jessica.arbour@mtsu.edu

References


align.missing

Procrustes superimposition of landmark datasets with some missing values

Description

This function carries out a generalized procrustes superimposition on all fully complete specimens and produces a consensus configuration (using "Shapes" procGPA). Each incomplete specimen is then individually rotated and aligned with the consensus configuration based on any landmarks are available (using "Shapes" procOPA). Data is returned superimposed.

Usage

align.missing(X)
**Arguments**

- **X**: An l X 2 (or 3) X n array of coordinate data, where n is the number of specimens and l is the number of landmarks.

**Value**

Returns an l X 2 (or 3) X n array of coordinate data.

**Author(s)**

J. Arbour

**References**


**See Also**

*MissingGeoMorph*

**Examples**

```r
data(dacrya)

## make some specimens incomplete
dac.miss<-missing.data(dacrya, remsp=0.2, land.vec=c(1,2,3,4,5,6))

## align all specimens
dac.aliigned<-align.missing(dac.miss)
```

---

**best.reg**

*Estimate missing morphometric data with a highly correlated variable*

**Description**

Estimates missing morphometric using regression on the most highly correlated morphological variable available.

**Usage**

```r
best.reg(x)
```

**Arguments**

- **x**: A n X m matrix of morphometric data with n specimens and m variables, containing some percentage of missing values input as NA.
Value

Returns a n X m matrix containing both the original morphometric values as well as estimates for all previously missing values.

Author(s)

J. Arbour and C. Brown

References


See Also

- est.reg

---

**bilat.align**

*Align a bilaterally symmetric landmark configuration with a plane*

**Description**

Aligns a bilaterally symmetric landmark dataset to a specific plane by minimized the sum of squared distances of one coordinate (x, y or z). Useful for averaging bilateral landmarks or in preparation for correcting for artifacts like bending.

**Usage**

bilat.align(coords, land.pairs, average = TRUE, restricted = NULL)

**Arguments**

- **coords**
  
  Either a matrix or array of landmark data with columns representing the x, y, z coordinates and rows representing landmarks. See details for how this is applied for a single vs. multiple specimens.

- **land.pairs**
  
  A 2 column matrix indicating bilaterally paired landmarks. All "left" landmarks should be in the same column (and likewise for "right landmarks")

- **average**
  
  An optional term indicating that bilaterally paired landmarks should be mirrored and averaged, leaving only one "side" and the midline landmarks.

- **restricted**
  
  A set of row numbers indicating which landmarks should be considered by "optim" when selecting the optimal rotation. Typically landmarks representing a rigid structure if some landmarks represent articulated/moveable features.
**byclade**

_Simulate missing morphometric data with taxonomic bias_

**Details**

If a matrix for a single specimen’s landmarks is provided this is aligned to a plane, if an array of multiple specimens is provided, these should be previously aligned with Procrustes superimposition, and the entire configuration is optimized with a single rotation applied to all specimens. SS are minimized across the third axis (coords[,3] or coords[,3]).

**Value**

A matrix or array giving the rotated landmark configuration

**Author(s)**

J.H. Arbour

**References**

Arbour, J.H. In Prep. Get Unbent! R Tools for the removal of arching and bending of fish specimens in geometric morphometric shape analysis

**See Also**

`unbend.spine, unbend.tps.poly`

**Examples**

```r
library(rgl)
data(darters)
## align darter configuration by head landmarks (restricted)
aligned<-bilat.align(darters$coords[,,1],
darters$land.pairs,average=FALSE,darters$restricted)
plot3d(aligned, aspect=FALSE)
```

**Description**

This function simulates higher frequency of missing data points in groups that are less numerically well represented in the whole sample, relative to other group. These groups may represent taxa (as used in Brown et al., 2012), but may also represent any other group of interest (e.g. populations, trials, subsamples, etc.). From a morphometric dataset, this function first selects a number of specimens to have data points removed from at random. A vector containing the number of measurements to remove from each specimen is sorted into descending order. Specimens are then sampled without replacement with a probability relative to the sum of the entire sample sizes divided by the number of specimens its respective group. The order the specimens are sampled determines the number of data points to be removed (i.e. the first to be sampled has the most removed). A complete mathematical description may be found in Brown et al. (2012).
Usage

byclade(x, remperc , groups)

Arguments

x     A n X m matrix of morphometric data with n specimens and m variables. Or an
      1 X 2 or 3 X n array of geometric morphometric coordinates (2D or 3D), where
      l is the number of landmarks.
remperc The percentage of data to be removed from the matrix, expressed as a decimal
      (ex: 30 percent would be entered as 0.3)
groups  A vector of length n specifying taxonomic group membership as integers (ex:
      c(1,1,2,2,3,3,...) )

Value

returns a matrix or array (depending on input) of morphometric data with missing variables input
as 'NA'

Author(s)

J. Arbour and C. Brown

References

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and

See Also

missing.data, obliterator

---

**complete.specimens**

Remove incomplete specimens from a landmark dataset

Description

This function takes a dataset containing both complete and incomplete specimens and removes all
incomplete specimens.

Usage

complete.specimens(dataset, nlandmarks)
Arguments

dataset      A n* l X 2 matrix of coordinate data, where n is the number of specimens and l is the number of landmarks. All landmarks from one specimen should be grouped together.

nlandmarks   The number of landmarks per specimen

Value

Returns an c * l X 2 matrix of landmark data, where c is the number of complete specimens and l is the number of landmarks.

Author(s)

J. Arbour

References


See Also

align.missing, MissingGeoMorph

Description

A linear morphometric dataset featuring 23 cranial measurements from 223 specimens representing 21 crocodilian species.

Usage

data(crocs)

Format

A n X m dataframe, where n is the number of specimens and m is the number of variables.

Source

http://datadryad.org/resource/doi:10.5061/dryad.m01st7p0

References

## crocs.landmarks

*Coordinate data for a crocodilian reference skull*

### Description

Landmark data for the measurements points on a reference crocodilian skull, for use with the `obliterator` function.

### Usage

```r
data(crocs.landmarks)
```

### Format

A 6 X m dataframe in which each column gives the start and end points for each cranial measurement in the crocs dataset, from a single reference specimen. 3D Coordinates are listed as x1, x2, y1, y2, z1, z2 in each column.

### Source


### See Also

`obliterator`, `byclade`, `missing.data`, `crocs`
**Source**


**See Also**

`align.missing`, `MissingGeoMorph`

---

### darters

<table>
<thead>
<tr>
<th><code>darters</code></th>
<th>Darter landmarks</th>
</tr>
</thead>
</table>

**Description**

A 3D landmark dataset from 30 species of darter fishes (Etheostomatinae; Percidae)

**Usage**

```r
data("darters")
```

**Format**

The format is: List of 6 $ coords : num [1:220, 1:3] -1.458 -0.489 -0.037 1.705 0.959 ... ...
 attr(*, "dimnames")=List of 3...

- $ land.pairs: data.frame: 101 obs. of 2 variables: ...
  - left : int [1:101] 1 3 5 7 9 11 13 15 17 19 ...
  - right: int [1:101] 2 4 6 8 10 12 14 16 18 20 ...

- $ sliders: data.frame: 32 obs. of 3 variables: ...
  - start: int [1:32] 22 23 24 25 26 27 28 29 30 32 ...
  - slide: int [1:32] 23 24 25 26 27 28 29 30 32 33 ...
  - end : int [1:32] 24 25 26 27 28 29 30 31 33 34 ...

- $ surface: data.frame: 144 obs. of 1 variable: ...

- $ restricted: int [1:58] 1 2 3 4 5 6 7 8 9 10 ...


**Details**

Includes landmark coordinates (coords), a matrix indicating bilaterally paired landmarks (land.pairs), curve sliders (sliders), surface sliders (surface), rows of head landmarks (restricted) and landmarks approximating the spine/long axis (reference).

**Source**

Arbour, J.H. In Prep. Get Unbent! R Tools for the removal of arching and bending of fish specimens in geometric morphometric shape analysis

**References**

Arbour, J.H. In Prep. Get Unbent! R Tools for the removal of arching and bending of fish specimens in geometric morphometric shape analysis
See Also

unbend.spine, bilat.align, unbend.tps.poly

Examples

```r
data(darters)
library(rgl)
plot3d(darters$coords[,,1], aspect=FALSE)
```

est.reg

*A-priori size regression for missing data estimation*

Description

Estimates missing data using regression on a designated size variable. Any values of the size variable missing are estimated with the variable best correlated with size.

Usage

```r
est.reg(x, col_indep)
```

Arguments

- `x`: A n X m matrix of morphometric data with n specimens and m variables, containing some percentage of missing values input as NA
- `col_indep`: The number of the column in which the independant size variable is stored. This column will be used to estimate missing values in the other columns.

Value

Returns a n X m matrix containing both the original morphometric values as well as estimates for all previously missing values.

Author(s)

J. Arbour and C. Brown

References


See Also

best.reg
flipped

Reflected Relabelling

Description
This function carries out reflected relabelling to estimate missing geometric morphometric landmarks using bilateral symmetry following Gunz et al 2009.
A set of 3D landmarks are mirrored and aligned with the original data (using procOPA from package "shapes"). Missing landmarks are interpolated from the mirrored specimen.

Usage
flipped(specimen, land.pairs, show.plot = FALSE, axis = 1)

Arguments
- specimen: An l X 3 matrix of coordinate data, where l is the number of landmarks. Some data should be missing and designated with NA.
- land.pairs: A 2 column matrix, each row should contain row numbers (from matrix specimen) indicating bilateral pairs of landmarks. Unpaired landmarks do not need to be included. See also bilateral symmetry analyses in package "geomorph".
- show.plot: Optionally plot the specimen using plot3d from rgl. Estimated landmarks are given in red. Defaults to FALSE.
- axis: Which axis should be mirrored across. Default is x (1).

Value
Returns a l X 3 matrix of landmarks.

Author(s)
J. Arbour

References

See Also
MissingGeoMorph
how.many.missing  Calculate the percentage of missing morphometric data

Description

Calculates the percentage of morphometric data points that have been replaced with `NA` by functions such as `missing.data`, `byclade` or `obliterator` from LOST. Used to verify the amount of missing data inputted into complete morphometric matrices.

Usage

how.many.missing(x)

Arguments

x  A n X m matrix of morphometric data with n specimens and m variables, or a or l X 2(or 3) array of geometric morphometric data containing some percentage of missing data

Value

Returns the percentage (as a decimal) of missing data points present in x

Author(s)

J. Arbour and C. Brown

References


See Also

missing.data

---

missing.data  Randomly input missing data points

Description

Randomly replaces a set percentage of data points in a matrix of morphometric measurements with NA to simulate missing data. This is function RMD from Brown et al (2012). The amount of missing data can be chosen as an overall percentage of data (simple morphometric data) or specimens and can be constrained to a set of landmarks (for landmarks).
Usage

```r
missing.data(x, remperc, remsp = NULL, land.vec = NULL, land.identity = NULL)
```

Arguments

- `x`: A `n` X `m` matrix of morphometric data with `n` specimens and `m` variables. Or an array of geometric morphometrics landmarks (∁ X `m` X `n`)
- `remperc`: The percentage of data to be removed from the matrix or array, expressed as a decimal (ex: 30 percent would be entered as 0.3)
- `remsp`: The percentage of specimens to be removed from the array, expressed as a decimal (ex: 30 percent would be entered as 0.3)
- `land.vec`: The number of landmarks to remove per specimen in an array. This can be a single value or vector with unique or repeating values.
- `land.identity`: A vector to constrain the landmarks to choose from when assigning missing data. The values correspond to row numbers in an array.

Value

Returns a `n` X `m` matrix or ∁ X `m` X `n` array of morphometric data with missing variables input as NA

Author(s)

J. Arbour and C. Brown

References


See Also

`byclade.obliterator`

Examples

```r
data(dacrya)

### remove 1 to 6 landmarks from 20% of specimens
dac.miss<-missing.data(dacrya, remsp=0.2, land.vec=c(1,2,3,4,5,6))
dac.miss
```
missing.specimens  Simulate incomplete specimens

Description

Randomly selects a pre-determined number of specimens from a landmark dataset (2D or 3D) and removes some of their landmarks.

Usage

missing.specimens(dataset, nspremove, nldremove, nlandmarks)

Arguments

dataset  A n*l X 2 (or 3) matrix of coordinate data, where n is the number of specimens and l is the number of landmarks. All landmarks from one specimen should be grouped together.

nspremove  The number of specimens which should have landmarks removed.

nldremove  The number of landmarks to remove per specimen. This may be a single value or a vector of values, none of which can be >nlandmarks. If a vector is given, for each specimen selected, the function will randomly select a value from the vector and remove that many landmarks.

nlandmarks  The number of landmarks per specimen

Value

Returns an n * l X 2 (or 3) matrix with some complete and some incomplete specimens.

Author(s)

J. Arbour

References


See Also

align.missing, MissingGeoMorph
Description

This function provides several options for estimating landmark data (details of which can be found in the references below). The function first aligns the landmarks using Procrustes superimposition (align.missing). Both 2D and 3D coordinates can be accommodated.

Usage

MissingGeoMorph(x, method = "BPCA", original.scale = FALSE)

Arguments

x A n* l X 2 matrix (2D data only) or an l X m X n array (2D or 3D data) of coordinate data, where n is the number of specimens and l is the number of landmarks, and m is the number of dimensions. All landmarks from one specimen should be grouped together. Missing values should be given as NA

method Four methods are provided for estimating missing landmark data: 1) "BPCA" - Bayesian principal component analysis, 2) "mean" - mean substitution, 3) "reg" - values are estimated based on the most strongly correlated variable available, and 4) "TPS" - thin plate spline interpolation (only available for 2D). See Arbour and Brown (2014) for a comparison of the performance of each of these methods.

original.scale Rescale and translate the data back to its original size (TRUE) or leave it in the rescaled, superimposed configuration (FALSE)

Value

Returns an n * l X 2 (or 3) matrix of coordinate data, with missing values imputed. Landmarks have been aligned and are given in the original shape space.

Author(s)

J. Arbour

References


See Also

align.missing, missing.specimens
obliterator

Simulate missing morphometric data with anatomical bias

Description

This function simulates the effect of proximity between measurements in morphometric data on the distribution of missing values. This attempts to replicate specimens showing regional incompleteness. From a morphometric dataset, this function selects a number of specimens to have data points removed from and a number of measurements to remove from each of these specimens based on a random distribution of missing data. For each specimen, this function randomly selects one starting data point for removal. All subsequent data points have a probability of removal that is proportional to the inverse of the distance to all previously removed data points, based on a reference set of landmarks (matrix 'distances'). For a complete mathematical description see Brown et al. (2012). See function obliteratorGM for the geometric morphometric implementation.

Usage

obliterator(x, remperc, landmarks, expo=1)

Arguments

x
A n X m matrix of morphometric data with n specimens and m variables

remperc
The percentage of data to be removed from the matrix, expressed as a decimal (ex: 30 percent would be entered as 0.3)

landmarks
A 6 X m matrix that includes the start and end points (landmarks) for each morphometric measurement from a reference specimen (3D). The data in each column is ordered as x1,x2,y1,y2,z1,z2. See example crocs.landmarks

expo
An optional term for raising the denominator to an exponent, to increase or decrease the severity of the anatomical bias

Value

Returns a n X m matrix of morphometric data with missing variables input as NA

Author(s)

J. Arbour and C. Brown

References


See Also

missing.data, byclade, obliteratorGM
obliteratorGM

Simulation of geometric morphometric landmarks with anatomical bias

Description

This is the geometric morphometric implementation of the LOST function obliterator. This attempts to replicate specimens showing regional incompleteness. For each specimen, this function randomly selects one starting data point for removal. All subsequent data points have a probability of removal that is proportional to the inverse of the distance to all previously removed data points, based on the shape of that particular specimen (this differs from the linear morphometric implementation which requires a reference set of coordinates). For a complete mathematical description see Brown et al. (2012).

Usage

obliteratorGM(x, remperc, expo=1)

Arguments

x
A n X m matrix of morphometric data with n specimens and m variables. Or a 1 X 2 or 3 X n array of geometric morphometric coordinates, with 1 being the number of landmarks.

remperc
The percentage of data to be removed from the matrix, expressed as a decimal (ex: 30 percent would be entered as 0.3)

expo
An optional term for raising the denominator to an exponent, to increase or decrease the severity of the anatomical bias

Value

Returns a n X m matrix of morphometric data with missing variables input as NA

Author(s)

J. Arbour and C. Brown

References


See Also

missing.data, byclade, obliterator
unbend.spine

Correct for lateral bending in fish geometric morphometric landmarks

Description

Correct for the impact of lateral bending along the spine of a fish in geometric morphometric landmarks. Fits a polynomial function along the length and width of the specimen, determines the perpendicular residuals and arc length along the polynomial and these are used as the new length and width landmarks. Landmarks are first centered and bilaterally aligned using bilat.align.

Usage

unbend.spine(coords, land.pairs, deg = 3, restricted = NULL)

Arguments

coords A matrix of landmark coordinate data. Columns should be coordinates, and rows landmarks.
land.pairs A 2-column matrix giving the bilaterally paired landmarks. One column should be all "left" landmarks and one all "right" landmarks.
deg The degrees for the polynomial function, passed to the function "poly". Typically 2 or 3.
restricted A limited set of landmarks (row numbers for the coords matrix) to use for bilateral alignment. Typically those representing a rigid/fixed structure (e.g., head). Passed to bilat.align.

Details

Resulting landmark data is in the same scale as the original landmark configuration. Can be applied over multiple specimens using for-loops or apply functions.

Value

bilat.aligned Provides the bilaterally aligned landmark data as a matrix
unbent Provides the bilaterally aligned and unbent landmark data as a matrix

Author(s)

J.H. Arbour

References

Arbour,J.H. In Prep. Get Unbent! R Tools for the removal of arching and bending of fish specimens in geometric morphometric shape analysis
unbend.tps.poly

See Also

bilat.align, unbend.tps.poly

Examples

data(darters)
library(rgl)
## bilaterally aligned using only head landmarks
lands.unbent<-unbend.spine(darters$coords[,2],
darters$land.pairs,deg=3, restricted=darters$restricted)$unbent

plot3d(lands.unbent, aspect=FALSE)

unbend.tps.poly

TPS-style unbend specimens

Description

Remove dorsoventral arching effect from fish specimen landmark data. Function similar to "unbend specimens" utility in the TPS software suite. Fits a polynomial function along the length and height of the specimen, determines the perpendicular residuals and arc length along the polynomial, and these are used as the new length and width landmarks.

Usage

unbend.tps.poly(coords, reference, axes = NULL, deg = 3)

Arguments

coords  A matrix of landmark coordinate data. Columns should be coordinates, and rows landmarks.
reference The rows of the matrix over which the polynomial function will be fit. Should represent the spine or other proxy for the long axis of the body.
axes A vector with 2 values representing the "lateral" view of the fish. The first entry should be the "long" (anterior-posterior) axis and the second should be the vertical (dorso-central) axis.
deg The degrees for the polynomial function, passed to "poly". Typically 2 or 3 (default = 3).

Details

It is advisable to remove lateral bending with unbend.spine prior to using this function. Otherwise data should be at least bilaterally aligned to a plane (and see bilat.align) Resulting landmark data is in the same scale as the original landmark configuration. Can be applied over multiple specimens using for-loops or apply functions.
Value

Returns a matrix of landmark data with the effect of dorso-ventral arching removed.

Author(s)

J.H. Arbour

References

Arbour,J.H. In Prep. Get Unbent! R Tools for the removal of arching and bending of fish specimens in geometric morphometric shape analysis

See Also

bilat.align, unbend.spine

Examples

library(rgl)
data(darters)
## bilaterally aligned using only head landmarks
lands.unbent<-unbend.spine(darters$coords[,,3],
darters$land.pairs,deg=3, restricted=darters$restricted)$unbent

plot(lands.unbent[,c(1,3)],asp=1)

lands.unbent<-unbend.tps.poly(lands.unbent,darters$reference,axes=c(1,3))
plot(lands.unbent[,c(1,2)],asp=1)

plot3d(lands.unbent, aspect=FALSE)
Index

* datasets
  darters, 9
  align.missing, 2, 7, 9, 14, 15
  best.reg, 3, 10
  bilat.align, 4, 10, 18–20
  byclade, 5, 8, 12, 13, 16, 17
  complete.specimens, 6
  crocs, 7, 8
  crocs.landmarks, 8, 8, 16
  dacrya, 8
  darters, 9
  est.reg, 4, 10
  flipped, 11
  how.many.missing, 12
  LOST (LOST-package), 2
  LOST-package, 2
  missing.data, 6, 8, 12, 16, 17
  missing.specimens, 14, 15
  MissingGeoMorph, 3, 7, 9, 11, 14, 15
  obliterator, 6, 8, 12, 13, 16, 17
  obliteratortGM, 16, 17
  unbend.spine, 5, 10, 18, 19, 20
  unbend.tps.poly, 5, 10, 19, 19