

# Package ‘LOST’

February 19, 2015

**Type** Package

**Title** Missing morphometric data simulation and estimation

**Version** 1.1

**Date** 2013-09-26

**Author** J. Arbour and C. Brown

**Maintainer** Jessica Arbour <jessica.arbour@utoronto.ca>

**Imports** gdata, shapes, e1071, pcaMethods, MASS, miscTools

**Description** LOST includes 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)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2013-10-11 01:36:31

## R topics documented:

LOST-package . . . . .	2
align.missing . . . . .	2
best.reg . . . . .	4
byclade . . . . .	5
complete.specimens . . . . .	6
crocs . . . . .	7
crocs.landmarks . . . . .	8
dacrya . . . . .	8
est.reg . . . . .	9
how.many.missing . . . . .	10
missing.data . . . . .	11
missing.specimens . . . . .	12
MissingGeoMorph . . . . .	13
obliterator . . . . .	14

<b>Index</b>	<b>16</b>
--------------	-----------

---

 LOST-package

*Missing morphometric data simulation and estimation*


---

### Description

LOST includes functions for simulating missing morphometric data randomly, with taxonomic bias and with anatomical bias as described by Brown et al. *In Press*. 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.

### Details

Package: LOST  
 Type: Package  
 Version: 1.1  
 Date: 2013-09-26  
 License: GPL (>= 2)

### Author(s)

J. Arbour and C. Brown

Maintainer: jessica.arbour@utoronto.ca

### References

Arbour, J. and Brown, C. *In Press*. Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution*

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

---

 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 whatever landmarks are available (using "Shapes" procOPA). Landmarks are returned in the same shape space as the original dataset.

**Usage**

```
align.missing(X, nlandmarks)
```

**Arguments**

X	A $n \times l \times 2$ 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  $n \times l \times 2$  matrix of aligned landmarks

**Author(s)**

J. Arbour

**References**

Arbour, J. and Brown, C. *In Press*. Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution*

**See Also**

[MissingGeoMorph](#)

**Examples**

```
data(dacrya)

## make some specimens incomplete
dac.miss<-missing.specimens(dacrya,10,c(1,2,3,4,5,6),16)
dac.miss

## align all specimens
dac.aligned<-align.missing(dac.miss,16)

## compare original and aligned
plot(dacrya,col="blue")
points(dac.aligned,col="red")
```

---

`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**

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

**Arguments**

<code>x</code>	A $n \times m$ matrix of morphometric data with $n$ specimens and $m$ variables, containing some percentage of missing values input as NA
----------------	---

**Value**

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

**Author(s)**

J. Arbour and C. Brown

**References**

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

**See Also**

[est.reg](#)

**Examples**

```
data(crocs)

### remove 30% of data points
croc.miss<-missing.data(crocs,0.3)
croc.miss

### estimating missing points
croc.new<-best.reg(croc.miss)
croc.new
```

byclade

*Simulate missing morphometric data with taxonomic bias***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., In Press), but may also represent any other group of interest (e.g. populations, trials, subsamples, etc.). 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 the distribution of missing data produced by [missing.data](#). A vector containing the number of measurements to remove from each specimen is produced and 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. (In Press).

**Usage**

```
byclade(x, remperc, ngroups, groups)
```

**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)
ngroups	The number of taxonomic groups present in the data matrix
groups	A vector of length n specifying taxonomic group membership as integers (ex: c(1,1,2,2,3,3,...) )

**Value**

returns a n X m matrix 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 Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

**See Also**

[missing.data](#), [obliterator](#)

## Examples

```
data(crocs)

## make some fake groups
rows<-nrow(crocs)
groups<-rep(1:4,length.out=rows)
groups

## remove 20% of data points
croc.miss<-byclade(crocs,0.2,4,groups)
croc.miss
```

---

complete.specimens	<i>Remove incomplete specimens from a landmark dataset</i>
--------------------	--

---

## 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 $n \times l \times 2$ 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  $c \times l \times 2$  matrix of landmark data, where  $c$  is the number of complete specimens and  $l$  is the number of landmarks.

## Author(s)

J. Arbour

## References

Arbour, J. and Brown, C. *In Press*. Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution*

## See Also

[align.missing](#), [codeMissingGeoMorph](#)

## Examples

```
data(dacrya)

#### make 10 specimens incomplete
dac.miss<-missing.specimens(dacrya,10,c(1,2,3,4,5,6),16)
dac.miss

#### remove 10 incomplete specimens
dac.com<-complete.specimens(dac.miss,16)
dac.com

### check that 10 were removed
nrow(dac.com)/16

## should be 63
```

---

crocs

*Crocodile morphometrics*

---

## 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

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

## See Also

[obliterator](#), [byclade](#), [missing.data](#), [crocs.landmarks](#)

---

crocs.landmarks	<i>Coordinate data for a crocodilian reference skull</i>
-----------------	--

---

### Description

Landmark data for the measurements points on a reference crocodilian skull, for use with the [obliterator](#) function

### Usage

```
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

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

### See Also

[obliterator](#), [byclade](#), [missing.data](#), [crocs](#)

---

dacrya	<i>Landmark data from Guianacara dacrya</i>
--------	---

---

### Description

Sixteen landmarks taken from the lateral profile of 73 specimens from the Essequibo and rio Branco drainages, used in the description of *Guianacara dacrya*

### Usage

```
data(dacrya)
```

### Format

A n \* l X 2 matrix, with n being the number of specimens and l being the number of landmarks



**Source**

Arbour, J. and Lopez-Fernandez, H. 2011. *Guiancara dacrya*, a new species from the rio Branco and Essequibo River drainages of the Guiana Shield (Perciformes: Cichlidae). *Neotropical Ichthyology* 9:87-96.

**See Also**

[align.missing](#), [MissingGeoMorph](#)

---

 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**

```
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**

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

**See Also**

[best.reg](#)

## Examples

```
data(crocs)

## remove 30% of data points
croc.miss<-missing.data(crocs,0.3)
croc.miss

## assume col 1 is the size variable
croc.new<-est.reg(croc.miss,1)
croc.new
```

---

how.many.missing	<i>Calculate the percentage of missing morphometric data</i>
------------------	--

---

## 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 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

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

## See Also

[missing.data](#)

**Examples**

```
data(crocs)

## remove 30% of data points
croc.miss<-missing.data(crocs,0.3)

## should return 0.3
how.many.missing (croc.miss)
```

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. (In Press).

**Usage**

```
missing.data(x, remperc)
```

**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)

**Value**

Returns a n X m matrix 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 Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

**See Also**

[byclade](#), [obliterator](#)

---

missing.specimens	<i>Simulate incomplete specimens</i>
-------------------	--------------------------------------

---

### Description

Randomly selects a pre-selected number of specimens from a landmark dataset and removes some of its landmarks.

### Usage

```
missing.specimens(dataset, nsremove, nldremove, nlandmarks)
```

### Arguments

dataset	A $n \times l \times 2$ 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.
nsremove	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  $n \times l \times 2$  matrix with some complete and some incomplete specimens.

### Author(s)

J. Arbour

### References

Arbour, J. and Brown, C. *In Press*. Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution*

### See Also

[align.missing](#), `{codeMissingGeoMorph}`

### Examples

```
data(dacrya)

#### remove 1 to 6 landmarks from 10 specimens
dac.miss<-missing.specimens(dacrya,10,c(1,2,3,4,5,6),16)
dac.miss
```

---

MissingGeoMorph*Estimate missing landmark data*

---

**Description**

This function provides several options for estimating landmark data (details of which can be found in the references below). The function first aligned the landmarks using Procrustes superimposition ([align.missing](#)).

**Usage**

```
MissingGeoMorph(x, nlandmarks, method = "BPCA")
```

**Arguments**

x	A $n \times l \times 2$ 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. Missing values should be given as NA
nlandmarks	The number of landmarks per specimen.
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 highest correlated variable available, and 4) "TPS" - thin plate spline interpolation. See Arbour and Brown (In Press) for a comparison of the performance of each of these methods.

**Value**

Returns an  $n \times l \times 2$  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**

Arbour, J. and Brown, C. *In Press*. Incomplete specimens in Geometric Morphometric Analyses. *Methods in Ecology and Evolution*

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

**See Also**

[align.missing](#), [missing.specimens](#)

## Examples

```
data(dacrya)

#### make some specimens incomplete
dac.miss<-missing.specimens(dacrya,10,c(1,2,3,4,5,6),16)
dac.miss

####estimate missing landmarks using mean substitution
dac.new<-MissingGeoMorph(dac.miss,16,method="mean")
dac.new
```

---

 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 deformation or 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 the distribution of missing data produced by [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. (In Press).

## 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 <a href="#">crocs.landmarks</a>
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**

Brown, C., Arbour, J. and Jackson, D. 2012. Testing of the Effect of Missing Data Estimation and Distribution in Morphometric Multivariate Data Analyses. *Systematic Biology* 61(6):941-954.

**See Also**

[missing.data,byclade](#)

**Examples**

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

### remove 20% of data points, with anatomic bias
croc.miss<-obliterator(crocs,0.2,crocs.landmarks)
croc.miss
```

# Index

`align.missing`, [2](#), [6](#), [9](#), [12](#), [13](#)

`best.reg`, [4](#), [9](#)

`byclade`, [5](#), [7](#), [8](#), [10](#), [11](#), [15](#)

`complete.specimens`, [6](#)

`crocs`, [7](#), [8](#)

`crocs.landmarks`, [7](#), [8](#), [14](#)

`dacrya`, [8](#)

`est.reg`, [4](#), [9](#)

`how.many.missing`, [10](#)

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

LOST-package, [2](#)

`missing.data`, [5](#), [7](#), [8](#), [10](#), [11](#), [14](#), [15](#)

`missing.specimens`, [12](#), [13](#)

MissingGeoMorph, [3](#), [6](#), [9](#), [12](#), [13](#)

`obliterator`, [5](#), [7](#), [8](#), [10](#), [11](#), [14](#)