Package 'assignR'

June 3, 2020

Type Package

Title Infer Geographic Origin from Isotopic Data

Version 1.2.1

Author Chao Ma, Gabe Bowen

Maintainer Gabe Bowen <gabe.bowen@utah.edu>

Description Routines for re-scaling isotope maps using known-origin tissue isotope data, assigning origin of unknown samples, and summarizing and assessing assignment results. Methods are adapted from Wunder (2010, in ISBN:9789048133536) and Vander Zanden, H. B. et al. (2014) <doi:10.1111/2041-210X.12229> as described in Ma, C. et al. (2020) <doi:10.1111/2041-210X.13426>.

Imports raster, sp, rgdal, maptools

Depends R (>= 3.5)

Suggests knitr, rmarkdown, testthat, covr

VignetteBuilder knitr

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

NeedsCompilation no

Repository CRAN

Date/Publication 2020-06-03 21:20:02 UTC

R topics documented:

---

assignR   *Tools for inferring geographic origin from isotopic data*

Description

Routines for rescaling isoscapes using known-origin tissue isotope data, assigning origin of unknown samples, and summarizing and assessing assignment results.

Author(s)

Maintainer: Gabe Bowen <gabe.bowen@utah.edu> **Authors**: Chao Ma, Gabe Bowen

See Also

https://spatial-lab.github.io/assignR/

---

calRaster   *Rescale isoscape using linear regression*

Description

This function uses known-origin tissue data to rescale a map of environmental isotope values to a map of tissue value (and associated uncertainty) using a linear regression model.

Usage

```
calRaster(known, isoscape, mask = NULL, interpMethod = 2, NA.value = NA,
  ignore.NA = TRUE, genplot = TRUE, outDir =  NULL, verboseLM = TRUE)
```

Arguments

known	SpatialPointsDataFrame. Known-origin data that should contain only one feature: tissue isotope value. Known-origin data can be queried using knownOrig.
isoscape	raster. Isoscape raster with two layers. The first one is the mean and the second is one standard deviation.
mask	SpatialPolygonsDataFrame. Constrains the area of the output rasters. If this is not provided, the entire area of isoscape is returned.
interpMethod	numeric. 1 or 2. Designate one of two methods for extracting values from isoscape. If 1, values for the cell in which a point falls are returned. If 2 the returned values are interpolated (bilinear interpolation) from the values of the four adjacent grid cells.
NA.value	NA or numeric. Value representing the absence of data in isoscape. Commonly used values include NA, -9999, and 9999.
ignore.NA	logical. If NA values are extracted from isoscape at the location of samples in known, ignore these values and proceed with calRaster.
genplot	logical. Plot the results.
outDir	character string. Directory path to which output will be saved. If NULL no files are written.
verboseLM	logical. Print out the linear regression results.

Value

isoscape.rescale	
	RasterBrick. isoscape rescaled using a linear regression model between values of known and isoscape. The first layer is the mean prediction and the second the standard deviation.
lm.data	data.frame. Known origin data and extracted isoscape values used for linear regression modeling.
lm.model	list. Linear regression model.

See Also

pdRaster

Examples

```
# load North America boundary and global isoscape
data("naMap")
data("d2h_world")

# load hydrogen isotope data for human hair in North America
d = subOrigData(taxon = "Homo sapiens", reference = "Ehleringer et al. 2008", mask = naMap)

# rescale from environmental isoscape to tissue isoscape
r = calRaster(known = d, isoscape = d2h_world, mask = naMap)
```

`d18o_world`	*Global growing season H isoscape*

Description

Global growing season precipitation O isoscape from waterisotopes.org

Usage

```
data("d18o_world")
```

Format

RasterStack with two layers. The first layer is the mean prediction and the second is 1 standard deviation

References

Bowen, G. J. (2018) Gridded maps of the isotopic composition of meteoric waters. http://www.waterisotopes.org.

Bowen, G. J., Wassenaar, L. I. and Hobson, K. A. (2005) Global application of stable hydrogen and oxygen isotopes to wildlife forensics. *Oecologia*, **143**, 337–348.

IAEA/WMO (2018) Global Network of Isotopes in Precipitation. The GNIP Database. https://nucleus.iaea.org/wiser.

Examples

```
library(raster)

data("d18o_world")
plot(d18o_world)
```

`d2h_lrNA`	*Low resolution North American growing season H isoscape*

Description

Interpolated growing season precipitation H isoscape from waterisotopes.org

Usage

```
data("d2h_lrNA")
```

Format

RasterBrick with two layers. The first layer is the mean prediction and the second is 1 standard deviation

References

Bowen, G. J. (2018) Gridded maps of the isotopic composition of meteoric waters. http://www.waterisotopes.org.

Bowen, G. J., Wassenaar, L. I. and Hobson, K. A. (2005) Global application of stable hydrogen and oxygen isotopes to wildlife forensics. *Oecologia*, **143**, 337–348.

IAEA/WMO (2018) Global Network of Isotopes in Precipitation. The GNIP Database. https://nucleus.iaea.org/wiser.

Examples

```
library(raster)

data("d2h_lrNA")
plot(d2h_lrNA)
```

---

d2h_world *Global growing season H isoscape*

Description

Interpolated global growing season precipitation H isoscape from waterisotopes.org

Usage

```
data("d2h_world")
```

Format

RasterStack with two layers. The first layer is the mean prediction and the second is 1 standard deviation

References

Bowen, G. J. (2018) Gridded maps of the isotopic composition of meteoric waters. http://www.waterisotopes.org.

Bowen, G. J., Wassenaar, L. I. and Hobson, K. A. (2005) Global application of stable hydrogen and oxygen isotopes to wildlife forensics. *Oecologia*, **143**, 337–348.

IAEA/WMO (2018) Global Network of Isotopes in Precipitation. The GNIP Database. https://nucleus.iaea.org/wiser.

Examples

```
library(raster)

data("d2h_world")
plot(d2h_world)
```

---

jointP   *Joint probability of origin*

Description

Joint probability for individuals of common origin (product of probabilities)

Usage

```
jointP(pdR)
```

Arguments

pdR   RasterStack or RasterBrick of probability density maps, e.g., as produced by
                   `calRaster`. All layers must have common spatial extent and projection.

Examples

```
# load North America boundary and global isoscape
data("naMap")
data("d2h_world")

# load hydrogen isotope data for human hair in North America
d = subOrigData(taxon = "Homo sapiens",
  reference = "Ehleringer et al. 2008", mask = naMap)

# rescale from environmental isoscape to tissue isoscape
r = calRaster(known = d, isoscape = d2h_world, mask = naMap)

# four unknown-origin examples
id = c("A", "B", "C", "D")
d2H = c(-110, -90, -105, -102)
un = data.frame(id,d2H)

# assignment
asn = pdRaster(r, unknown = un, mask = naMap)

# joint probability for individuals of common origin
jointP(asn)
```

---

knownOrig   *Isotope values for tissues of known origin*

Description

This dataset consists of hydrogen and oxygen isotope values of human hair, insect chitin and bird
feathers with location and sample information.

Usage

```
data("knownOrig")
```

Format

SpatialPointsDataFrame with 8 variables.

[, 1] d2H: Hydrogen isotope ratio (permil, VSMOW)

[, 2] d18O: Oxygen isotope ratio (permil, VSMOW)

[, 3] Taxon: Genus and species name

[, 4] Group: Taxonomic group (informal)

[, 5] Source_quality:Code indicating level of certainty in geographic origin

[, 6] Age_code: Code for age of individual

[, 7] Reference: Bibliographic citation of source

[, 8] ID: Unique ID

References

Bowen, G. J. et al. (2009) Dietary and physiological controls on the hydrogen and oxygen isotope ratios of hair from mid-20th century indigenous populations. *American Journal of Physical Anthropology*, **139**, 494–504.

Clark, R. G., Hobson, K. A. and Wassenaar, L. I. (2006) Geographic variation in the isotopic (dD, d13C, d15N, d34S) composition of feathers and claws from lesser scaup and northern pintail: implications for studies of migratory connectivity. *Canadian Journal of Zoology* **84** 1395–1401.

Ehleringer, J. R. et al. (2008) Hydrogen and oxygen isotope ratios in human hair are related to geography. *Proceedings of the National Academy of Sciences* **105** 2788–2793.

Hobson, K. A., Bowen, G. J., Wassenaar, L. I., Ferrand, Y. and Lormee, H. (2004) Using stable hydrogen and oxygen isotope measurements of feathers to infer geographical origins of migrating European birds. *Oecologia* **141** 477–488.

Hobson, K. A. and Koehler, G. (2015) On the use of stable oxygen isotope (d18O) measurements for tracking avian movements in North America. *Ecology and Evolution* **5** 799–806.

Hobson, K. A., Van Wilgenburg, S. L., Wassenaar, L. I. and Larson, K. (2012) Linking hydrogen (d2H) isotopes in feathers and precipitation: sources of variance and consequences for assignment to isoscapes. *PLoS ONE* **7** e35137.

Hobson, K. A. and Wassenaar, L. I. (1997) Linking breeding and wintering grounds of neotropical migrant songbirds using stable hydrogen isotopic analysis of feathers. *Oecologia* **109** 142–148.

Hobson, K. A., Wassenaar, L. I. and Taylor, O. R. (1999) Stable isotopes (dD and d13C) are geographic indicators of natal origins of monarch butterflies in eastern North America. *Oecologia* **120** 397–404.

Lott, C. A. and Smith, J. P. (2006) A geographic-information-system approach to estimating the origin of migratory raptors in North America using stable hydrogen isotope ratios in feathers. *The Auk* **123** 822–835.

Thompson, A. H. et al. (2010) Stable isotope analysis of modern human hair collected from Asia (China, India, Mongolia and Pakistan). *American Journal of Physical Anthropology* **141** 440–451.

Wunder, M.B., 2007. Geographic structure and dynamics in Mountain Plover. Doctoral dissertation, Colorado State University.

See Also

[subOrigData](#)

Examples

```
data("knownOrig")
class(knownOrig)
summary(knownOrig@data)
```

---

naMap   *data: naMap*

Description

Simplified North American boundaries polygon

Usage

```
data("naMap")
```

Format

SpatialPolygonsDataFrame

Examples

```
library(sp)

data("naMap")
plot(naMap)
```

---

oddsRatio   *Odds ratio of points or regions*

Description

Calculate ratio of odds for two locations (points or polygons)

Usage

```
oddsRatio(pdR, inputP)
```

Arguments

pdR	RasterStack or RasterBrick of probability density maps, e.g., as produced by calRaster. All layers must have common projection.
inputP	SpatialPoints or SpatialPolygons (or *DataFrame equivalent) of length 2

Examples

```
library(sp)

# load North America boundary and global isoscape
data("naMap")
data("d2h_world")

# load hydrogen isotope data for human hair in North America
d = subOrigData(taxon = "Homo sapiens",
  reference = "Ehleringer et al. 2008", mask = naMap)

# rescale from environmental isoscape to tissue isoscape
r = calRaster(known = d, isoscape = d2h_world, mask = naMap)

# four unknown-origin examples
id = c("A", "B", "C", "D")
d2H = c(-110, -90, -105, -102)
un = data.frame(id,d2H)

# assignment
asn = pdRaster(r, unknown = un, mask = naMap)

# SpatialPolygons for two regions of interest
data("states")
s1 = states[states$STATE_ABBR == "UT",]
s2 = states[states$STATE_ABBR == "NM",]
plot(naMap)
plot(s1, border = "red", add=TRUE)
plot(s2, border = "blue", add=TRUE)

# Get odds ratio for two regions using SpatialPolygon method
s12 = rbind(s1, s2)
oddsRatio(asn, s12)

# Create SpatialPoints for two points of interest
pp1 = c(-112,40)
pp2 = c(-105,33)
pp12 = SpatialPoints(coords = rbind(pp1,pp2),
  proj4string=CRS("+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"))
points(pp12, pch = 21, bg = "light blue")

# Get odds ratio for two points using SpatialPoints method
oddsRatio(asn, pp12)
```

---

pdRaster   *Probability of origin surfaces*

Description

Calculate posterior probabilities of origin for a sample based on its isotope ratio.

Usage

```
pdRaster(r, unknown, prior = NULL, mask = NULL, genplot = TRUE, outDir = NULL)
```

Arguments

r   RasterStack or RasterBrick with two layers, e.g., as created by calRaster. The
                    first layer is the tissue-specific isoscape and the second the isoscape prediction
                    uncertainty (1 standard deviation).

unknown   data.frame. First column should be unique sample IDs, and second column
                    should be tissue isotope values for the individuals for which the assignments
                    will be made.

prior   raster. Optional raster layer with prior probabilities, which has the same projec-
                    tion, resolution and extent as r.

mask   SpatialPolygonsDataFrame. This mask will constrain the assignment area. If
                    this is not provided, a default of mask of the extent of r is used.

genplot   logical. Plot results in R.

outDir   character string. Directory path to which output will be saved. If NULL no files
                    are written.

Value

RasterStack or RasterLayer including a probability density surface for each individual in unknown.
If saveFiles == TRUE writes individual rasters in GeoTIFF format and a single PDF file with
images for each probability density raster to the working directory.

See Also

calRaster

Examples

```
# load North America boundary and global isoscape
data("naMap")
data("d2h_world")

# load hydrogen isotope data for human hair in North America
d = subOrigData(taxon = "Homo sapiens",
  reference = "Ehleringer et al. 2008", mask = naMap)
```

```
# rescale from environmental isoscape to tissue isoscape
r = calRaster(known = d, isoscape = d2h_world, mask = naMap)

# sample to assign
id = "smile"
d2H = -80
un = data.frame(id, d2H)

# posterior probability surface
asn = pdRaster(r, unknown = un, mask = naMap)
```

---

plot.QA   *Plot quality metrics for geographic assignments*

Description

Plot the output from QA, including spatial precision, bias, sensitivity and odds ratio of known locations for validation samples.

Usage

```
   ## S3 method for class 'QA'
plot(x, ..., outDir = NULL)
```

Arguments

x	One or more QA objects
...	Other arguments to be passed to plot
outDir	character string. Directory path to which output will be saved. If NULL no files are written.

References

Ma, C. et al. (2020) assignR : An R package for isotope-based geographic assignment. *Methods in Ecology and Evolution* doi:10.1111/2041-210X.13426.

Vander Zanden, H. B. et al. (2014) Contrasting assignment of migratory organisms to geographic origins using long-term versus year-specific precipitation isotope maps. *Methods in Ecology and Evolution* **5** 891–900.

See Also

QA

Examples

```
# load data
data("naMap") # North America
data("d2h_lrNA") # precipitation hydrogen isoscape of North America
data("knownOrig") # isotopes of known-origin samples

# extract some known-origin data
d1 = subOrigData(taxon = "Buteo lagopus")

# run quality assessment based on precipitation hydrogen isotopes and
# known-origin samples; small values of valiStation and valiTime
# are used in example to reduce run time

# first with one example
# gives warning because a small number of samples are available
qa1 = QA(isoscape = d2h_lrNA, known = d1, valiStation = 1,
         valiTime = 2, by = 10, mask = naMap, name = "Buteo")

# plot the qa result
plot(qa1)

# now compare with a second data set
d2 = subOrigData(taxon = "Charadrius montanus")
qa2 = QA(isoscape = d2h_lrNA, known = d2, valiStation = 1,
         valiTime = 2, by = 10, mask = naMap, name = "Charadrius")
plot(qa1, qa2)
```

QA   *Quality assessment of geographic assignments*

Description

How well does a given isoscape and known origin data set constrain the geographic origin of samples? Uses iterative re-sampling of the known origin data set to evaluate sample assignments and reports a suite of quality metrics.

Usage

```
QA(isoscape, known, valiStation = ceiling(length(known)*0.1), valiTime = 50,
  by = 2, mask = NULL, setSeed = TRUE, name = NULL)
```

Arguments

isoscape   RasterStack or RasterBrick with two layers. The first layer is mean isoscape
            prediction and the second the isoscape prediction uncertainty (1 standard devia-
            tion).

known   SpatialPointsDataFrame. Known-origin data that should contain only one fea-
            ture: tissue isotope value. Its length must be larger or equal to 3. Known-origin
            data can be queried using knownOrig.

valiStation	numeric. How many samples from known are withheld for validation? Must be two or more smaller than the length of known.
valiTime	numeric. How many times do you want to randomly draw validation stations and run the validation? Must be an integer equal to or greater than one.
by	integer. Threshold increment to use in evaluating assignment performance. Must be between 1 and 25.
mask	SpatialPolygonsDataFrame. Constrains the area of the output rasters. If this is not provided, the entire area of isoscape is returned.
setSeed	logical. Do you want to set.seed() when you randomly draw validation stations? Yes gives the same sequence of random draws each time the function is called.
name	character. Useful for identifying the QA output in subsequent plotting.

Value

val_stations	numeric. An X*Y data.frame of validation station IDs for all valiTime. X = valiTime and Y = valiStation.
pd_val	numeric. An X*Y data.frame containing the posterior probability density for the validation stations. X = valiTime and Y = valiStation.
prption_byArea	numeric. An X*Y data.frame showing the proportion of validation individuals for which the known origin is contained within the top 0.00 to 1.00 area quantile (with increment of 0.01; Y = 101). X = valiTime.
prption_byProb	numeric. An X*Y data.frame showing the proportion of validation individuals for which the known origin is contained within the top 0.00 to 1.00 probability quantile (with increment of 0.01; Y = 101). X = valiTime.
precision	list. The length of the list is valiTime. Each element is an X*Y matrix showing the proportional area of the total assignment surface covered by the assignment region at a given probability quantile from 0.00 to 1.00 *with increment of 0.01; X = 101) for each validation sample (Y = valiStation).
random_prob_density	
	Random probability of assignment to any given grid cell on the assignment surface(i.e. 1 divided by the total number of grid cells).
name	character. Name assigned to the QA object.
by	integer. Value of by used.

Note

Please see Ma et al., 2019 for methodological details.

References

Ma, C. et al. (2020) assignR: An R package for isotope-based geographic assignment. *Methods in Ecology and Evolution* doi:10.1111/2041-210X.13426.

Vander Zanden, H. B. et al. (2014) Contrasting assignment of migratory organisms to geographic origins using long-term versus year-specific precipitation isotope maps. *Methods in Ecology and Evolution* **5** 891–900.

See Also

plot.QA

Examples

```
# load data
data("naMap") # North America
data("d2h_lrNA") # precipitation hydrogen isoscape for North America
data("knownOrig") # hydrogen isotopes of known-origin samples

# extract some known-origin data
d1 = subOrigData(taxon = "Buteo lagopus")

# run quality assessment based on precipitation hydrogen isotopes and
# known-origin samples; small values of valiStation and valiTime
# are used in example to reduce run time

# first with one example
# gives warning because a small number of samples are available
qa1 = QA(isoscape = d2h_lrNA, known = d1, valiStation = 1,
         valiTime = 2, by = 10, mask = naMap, name = "Buteo")

# plot the qa result
plot(qa1)

# now compare with a second data set
d2 = subOrigData(taxon = "Charadrius montanus")
qa2 = QA(isoscape = d2h_lrNA, known = d2, valiStation = 1,
         valiTime = 2, by = 10, mask = naMap, name = "Charadrius")
plot(qa1, qa2)
```

---

qtlRaster   *Sample assignment using thresholds*

Description

Selects the grid cells of probability density rasters with the highest probability and returns rasters
with these cell values set to 1. Cells are selected based on the user-specified quantile threshold
so that the most-probable cells representing a given fraction of the assignment area or posterior
probability are returned.

Usage

```
qtlRaster(pdR, threshold, thresholdType = "area", genplot = TRUE, outDir = NULL)
```

Arguments

pdR	RasterLayer, RasterStack, or RasterBrick. Probability density maps for individual samples, e.g., as output by `pdRaster`.
threshold	numeric from 0 to 1. Quantile to be selected.
thresholdType	character. Either "area" (default) or "prob". If "area", the most probable cells constituting `threshold` percent of the assignment area are selected. If "prob", the most probable cells constituting `threshold` percent of the posterior probability are selected.
genplot	logical.Plot results in R.
outDir	character string. Directory path to which output will be saved. If NULL no files are written.

Examples

```
# load North America boundary and global isoscape
data("naMap")
data("d2h_world")

# load hydrogen isotope data for human hair in North America
d = subOrigData(taxon = "Homo sapiens",
  reference = "Ehleringer et al. 2008", mask = naMap)

# rescale from environmental isoscape to tissue isoscape
r = calRaster(known = d, isoscape = d2h_world, mask = naMap)

# four unknown-origin examples
id = c("A", "B", "C", "D")
d2H = c(-110, -90, -105, -102)
un = data.frame(id, d2H)

# posterior probability surface
asn = pdRaster(r, unknown=un, mask=naMap)

# assign to most probable 10 percent of area
## Not run: qtlRaster(asn, threshold = 0.1)

# assign to most probable 10 percent of proabability distribution
qtlRaster(asn, threshold = 0.1, thresholdType = "prob")
```

---

states	*data: states*

Description

Spatial Polygon Data Frame of lower 48 U.S. states

Usage

```
data("states")
```

Format

SpatialPolygonsDataFrame

Examples

```
library(sp)

data("states")
plot(states)
```

---

subOrigData   *Extract known origin sample data*

Description

This function subsets the known-origin isotope dataset included in this package.

Usage

```
subOrigData(marker = "d2H", taxon = NULL, group = NULL, reference = NULL,
    age_code = NULL, mask = NULL)
```

Arguments

marker	character string. Column name for isotopic data to be extracted, either "d2H" or "d18O".
taxon	character string or string vector. Species name(s) for data to be extracted.
group	character string or string vector. Taxonomic groups for data to be extracted.
reference	character string or string vector. Bibliographic reference for data to be extracted.
age_code	character string or string vector. Animal age code for data to be extracted.
mask	SpatialPolygons or SpatialPolygonsDataFrame. Constrains the geographic area from which data are extracted. If not provided, global.

Value

SpatialPointsDataFrame including the location and a single feature corresponding to the marker for each selected sample. Output is formatted for use in calRaster or QA functions.

Examples

```
## WITHOUT mask
# extract d2H data for Jackdaw, Partridge and Willow Grouse
d1 = subOrigData(taxon = c("Danaus plexippus", "Setophaga ruticilla", "Turdus migratorius"))
summary(d1)

# extract d2H data for insects and birds
d2 = subOrigData(group = c("insect","bird"))
```

```
summary(d2)

# extract d18O data for humans
d3 = subOrigData(marker = "d18O", group = "human")
summary(d3)

# extract d2H data for humans using taxon
d4 = subOrigData(marker = "d2H", taxon = "Homo sapiens")
summary(d4)

## WITH mask
# error - no samples found
## Not run: d5 = subOrigData(taxon = "Turdus philomelos", mask = naMap)
# these work OK
d6 = subOrigData(taxon = c("Danaus plexippus", "Setophaga ruticilla",
  "Turdus migratorius"), mask = naMap)
d7 = subOrigData(group = "human", mask = naMap)
```

---

unionP   *Union probability of origin*

Description

Probabilities that at least one individual came from each location in the assignment area (union of
probabilities)

Usage

```
unionP(pdR)
```

Arguments

pdR   RasterStack or RasterBrick of probability density maps, e.g., as produced by
                calRaster. All layers must have common spatial extent and projection.

Examples

```
# load North America boundary and global isoscape
data("naMap")
data("d2h_world")

# load hydrogen isotope data for human hair in North America
d = subOrigData(taxon = "Homo sapiens",
  reference = "Ehleringer et al. 2008", mask = naMap)

# rescale from environmental isoscape to tissue isoscape
r = calRaster(known = d, isoscape = d2h_world, mask = naMap)

# four unknown-origin examples
id = c("A", "B", "C", "D")
```

```
d2H = c(-110, -90, -105, -102)
un = data.frame(id,d2H)

# assignment
asn = pdRaster(r, unknown = un, mask = naMap)

# probability that one or more individuals are from a given location
unionP(asn)
```

Index