Package ‘SDMPlay’

August 9, 2016

Type Package

Title Species Distribution Modelling Playground

Version 1.0

Date 2016-08-03

Description
Functions provided by this pedagogic package allow to compute models with two popular machine learning approaches, BRT (Boosted Regression Trees) and MaxEnt (Maximum Entropy) applied on sets of marine biological and environmental data. They include the possibility of managing the main parameters for the construction of the models. Classic tools to evaluate model performance are provided (Area Under the Curve, omission rate and confusion matrix, map standard deviation) and are completed with tools to perform null models. The biological dataset includes original occurrences of two species of the class Echinoidea (sea urchins) present on the Kerguelen Plateau and that show contrasted ecological niches. The environmental dataset includes the corresponding statistics for 15 abiotic and biotic descriptors summarized for the Kerguelen Plateau and for different periods in a raster format. The package can be used for practicals to teach and learn the basics of species distribution modelling. Maps of potential distribution can be produced based on the example data included in the package, which brings prior observations of the influence of spatial and temporal heterogeneities on modelling performances. The user can also provide his own datasets to use the modelling functions.

License GPL-3

LazyData TRUE

Depends R (>= 3.1.0)

Imports sp, raster, dismo, SDMTools, gbm, stats, base, graphics,
ncdf4, rgdal

Suggests maptools, testthat, grDevices, rJava (>= 0.5-0)

RoxygenNote 5.0.1

NeedsCompilation no

Author Guillaumot Charlene [aut, cre],
Martin Alexis [aut],
Eleaume Marc [aut],
Saucede Thomas [aut]

Maintainer Guillaumot Charlene <charleneguillaumot21@gmail.com>

Repository CRAN

Date/Publication 2016-08-08 19:09:02

R topics documented:

`brisaster.antarcticus` *Records of* Brisaster antarcticus *echinoid presences on the Kerguelen Plateau*

Description

Dataset that contains the presence of the echinoid species *Brisaster antarcticus* reported on the Kerguelen Plateau (63/81W; -46/-56S) during the campaigns of the RV Marion Dufresne MD03 1974 & MD04 1975, POKER 2 (2010) campaigns and the program PROTEKER (2013, 2014, 2015).
Brisaster antarcticus (Doderlein 1906) is distributed from 3.5 to 75.6W and -53.35 to -45.95S in the Southern Ocean. The species is mainly found around Kerguelen and Crozet Islands. *Brisaster antarcticus* commonly lives from 100 to 600 meters depth.
It is a detrivorous species for which reproduction includes dispersal (David et al. 2005).

See Guillaumot et al. (submitted) for more details.

Usage

```
data(brisaster.antarcticus)
```

Format

A data frame containing 43 occurrences and 13 descriptive variables

- *id*
  Occurrence number indicator

- *scientific.name*
  Species scientific name
- *scientific.name.authorship*
  Author of the species description
- *genus*
  Genus scientific name and its associated author
- *family*
  Family scientific name and its associated author
- *order.and.higher.taxonomic.range*
  Order scientific name and its associated author
- *decimal.Longitude*
  Longitude in decimal degrees
- *decimal.Latitude*
  Latitude in decimal degrees
- *depth*
  Depth in meters
- *campaign*
  Campaign origin of the data
- *reference*
  Campaign reference
- *vessel*
  Campaign vessel

References

David B, Chone T, Mooi R, De Ridder C (2005) Antarctic Echinoidea. Synopses of the Antarctic Benthos 10.

Doderlein L (1906) Die Echinoiden der Deutschen Tiefsee-Expedition. Deutsche Tiefsee Expedition 1898-1899. 5: 63-290.

Guillaumot C, A Martin, S Fabri-Ruiz, M Eleaume & T Saucede. Echinoids of the Kerguelen Plateau: Occurrence data and environmental setting for past, present, and future species distribution modelling, Zookeys, Manuscript submitted for publication.

Examples

```
data(brisaster.antarcticus)
x <- brisaster.antarcticus

# plot of the occurrences:
# selecting the species according to the campaigns
brisaster7475 <- subset(x,x$year==1974 | x$year==1975)
brisaster20102015 <- subset(x,x$campaign=='POKER II'| x$campaign=='PROTEKER')

# drawing the background (depth)
library(grDevices)
blue.palette <- colorRampPalette(c('blue','deepskyblue','azure'))(100)
```

```
data('predictors1965_1974')
depth <- raster :: subset(predictors1965_1974, 1)

raster::plot(depth, col=blue.palette,main= "Brisaster antarcticus occurrences")

# adding the occurrence data to the background
points(brisaster7475[,c('decimal.Longitude','decimal.Latitude')],
      col='orange',pch=16)
points(brisaster20102015[,c('decimal.Longitude','decimal.Latitude')],
      col='darkgreen',pch=16)
legend('bottomleft',
       legend=c('Brisaster antarcticus 1974-1975','Brisaster antarcticus 2010-2015'),
       col= c('orange','darkgreen'), pch= c(15, 15),cex=0.9)
```

`compute.brt` *Compute BRT (Boosted Regression Trees) model*

Description

Compute species distribution models with Boosted Regression Trees

Usage

```
compute.brt(x, proj.predictors, tc = 2, lr = 0.001, bf = 0.75,
            n.trees = 50, step.size = n.trees)
```

Arguments

`x`	SDMtab object or dataframe that contains id, longitude, latitude and values of environmental descriptors at corresponding locations
`proj.predictors`	RasterStack of environmental descriptors on which the model will be projected
`tc`	Integer. Tree complexity. Sets the complexity of individual trees
`lr`	Learning rate. Sets the weight applied to individual trees
`bf`	Bag fraction. Sets the proportion of observations used in selecting variables
`n.trees`	Number of initial trees to fit. Set at 50 by default
`step.size`	Number of trees to add at each cycle

Details

The function realizes a BRT model according to the gbm.step function provided by Elith et al.(2008).
See the publication for further information about setting decisions.
The map produced provides species presence probability on the projected area.

Value

A list of 4

model$algorithm "brt" string character

- *model$data* x dataframe that was used to implement the model
- *model$response* Parameters returned by the model object
- *model$raster.prediction* Raster layer that predicts the potential species distribution

Note

See Barbet Massin et al. (2012) for information about background selection to implement BRT models

References

Barbet Massin M, F Jiguet, C Albert & W Thuiller (2012) Selecting pseudo absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2): 327-338.

Elith J, J Leathwick & T Hastie (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4): 802-813.

See Also

gbm.step

Examples

```
## Not run:
#Download the presence data
data('ctenocidaris.nutrix')
occ <- ctenocidaris.nutrix
# select longitude and latitude coordinates among all the information
occ <- ctenocidaris.nutrix[,c('decimal.Longitude','decimal.Latitude')]

#Download the environmental predictors restricted on geographical extent and depth (-1500m)
envi <- raster::stack(system.file('extdata', 'pred.grd',package='SDMPlay'))
envi

#Open SDMtab matrix
x <- system.file(file='extdata/SDMdata1500.csv',package='SDMPlay')
SDMdata <- read.table(x,header=TRUE, sep=';')

#Run the model
model <- SDMPlay:::compute.brt (x=SDMdata, proj.predictors=envi,lr=0.0005)

#Plot the partial dependance plots
dismo::gbm.plot(model$response)

#Get the contribution of each variable for the model
model$response$contributions
```

```
#Get the interaction between variables
dismo::gbm.interactions(model$response)
#Plot the interactions
int <- dismo::gbm.interactions(model$response)
# choose the interaction to plot
dismo::gbm.perspec(model$response,int$rank.list[1,1],int$rank.list[1,3])

#Plot the map prediction
library(grDevices) # add nice colors
palet.col <- colorRampPalette(c('deepskyblue','green','yellow', 'red'))( 80 )
raster::plot(model$raster.prediction, col=palet.col)
#add data
points(occ, col='black',pch=16)


# SECOND EXAMPLE: projecting the model on another period
# Remark: to predict on a different RasterStack, the rasterlayer names of the two
# stacks must be the same and the number of layers must be the same as well.
# Changes have been done in this example by attributing similar names to pred
# and pred2000 stacks and adding extra blank layers (NA layers) to pred2000 stack.
envi2000 <- raster::stack(system.file('extdata', 'pred2000.grd',package='SDMPlay'))

#Run the model
model2 <- SDMPlay:::compute.brt (x=SDMdata, proj.predictors=envi2000,lr=0.0005)

#Plot the new predicting map
raster::plot(model2$raster.prediction, col=palet.col)
#add data
points(occ, col='black',pch=16)
## End(Not run)
```

`compute.maxent` *Compute MaxEnt model*

Description

Compute species distribution models with MaxEnt (Maximum Entropy)

Usage

```
compute.maxent(x, proj.predictors)
```

Arguments

`x`	SDMtab object or dataframe that contains id, longitude, latitude and values of environmental descriptors at corresponding locations.
`proj.predictors`	RasterStack of environmental descriptors on which the model will be projected

Details

MaxEnt species distribution model minimizes the relative entropy between environmental descriptors and presence data. Further information are provided in the references below.

compute.maxent uses the functionalities of the maxent function. This function uses MaxEnt species distribution software, which is a java program that could be downloaded at http://www.cs.princeton.edu/~schapire/maxent/. In order to run compute.maxent, put the 'maxent.jar' file downloaded at this adress in the 'java' folder of the dismo package (path obtained with the system.file('java', package='dismo') command). MaxEnt 3.3.3b version or higher is required.

Value

A list of 4

model$algorithm "maxent" string character

- *model$data* x dataframe that was used to implement the model
- *model$response* Parameters returned by the model object
- *model$raster.prediction* Raster layer that predicts the potential species distribution

Note

To implement MaxEnt models, Phillips & Dudik (2008) advice a large number of background data. You can also find further information about background selection in Barbet Massin et al. (2012).

References

http://www.cs.princeton.edu/~schapire/maxent/

Barbet Massin M, F Jiguet, C Albert & W Thuiller (2012) Selecting pseudo absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2): 327-338.

Elith J, S Phillips, T Hastie, M Dudik, Y Chee & C Yates (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17:43-57. http://dx.doi.org/10.1111/j.1472-4642.2010.00725.x

Phillips S, M Dudik & R Schapire (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning* : 655-662

Phillips S, R Anderson & R Schapire (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231-259.

Phillips S and M Dudik (2008) Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography* 31(2): 161-175.

See Also

maxent .jpackage: initialize dismo for Java

Examples

```
#Download the presence data
data('ctenocidaris.nutrix')
occ <- ctenocidaris.nutrix
# select longitude and latitude coordinates among all the information
occ <- ctenocidaris.nutrix[,c('decimal.Longitude','decimal.Latitude')]

#Download the environmental predictors restricted on geographical extent and depth (-1500m)
envi <- raster::stack(system.file('extdata', 'pred.grd',package='SDMPlay'))
envi

#Open SDMtab matrix
x <- system.file(file='extdata/SDMdata1500.csv',package='SDMPlay')
SDMdata <- read.table(x,header=TRUE, sep=';')

#only run if the maxent.jar file is available, in the right folder
jar <- paste(system.file(package="dismo"), "/java/maxent.jar", sep='')
# checking if maxent can be run (normally not part of your script)
if (file.exists(jar) & require(rJava)) {

# Run the model
model <- SDMPlay:::compute.maxent(x=SDMdata , proj.predictors=envi)

# Plot the map prediction
library(grDevices) # add nice colors
palet.col <- colorRampPalette(c('deepskyblue','green','yellow','red'))(80)
#'raster::plot(model$raster.prediction, col=palet.col)
# add data
points(occ, col='black',pch=16)

# Get the partial dependance curves
dismo::response(model$response)

# Get the percentage of contribution of each variable to the model
#plot(model$response)

# Get all the information provided by the model on a html document
model$response

# SECOND EXAMPLE: projecting the model on another period
# Remark: to predict on a different RasterStack, the rasterlayer names of the two
# stacks must be the same and the number of layers must be the same as well.
# Changes have been done in this example by attributing similar names to pred
# and pred2000 stacks and adding extra blank layers (NA layers) to pred2000 stack.
envi2000 <- raster::stack(system.file('extdata', 'pred2000.grd',package='SDMPlay'))

# run the model
model2 <- SDMPlay:::compute.maxent(x=SDMdata, proj.predictors=envi2000)

# plot the new predicting map
raster::plot(model2$raster.prediction, col=palet.col)
# add data
```

```
points(occ, col='black',pch=16)}
```

ctenocidaris.nutrix *Records of* Ctenocidaris nutrix *echinoid presences on the Kerguelen Plateau*

Description

Dataset that contains the presence of the echinoid species *Ctenocidaris nutrix* reported on the Kerguelen Plateau (63/81W; -46/-56S) during the campaigns of the RV Marion Dufresne MD03 1974 & MD04 1975, POKER 2 (2010) campaigns and the program PROTEKER (2013, 2014, 2015). *Ctenocidaris nutrix* (Thomson 1876) is a broad range species, distributed from -70.5W to 143.7E and -76.13 to -47.18S in the Southern Ocean. The species is mainly found around the Kerguelen Plateau, near Weddel Sea and Scotia Ridge. The species is known from littoral waters down to 800m. It is a carnivorous and direct developer species that breeds its youngs (David et al. 2005). *Ctenocidaris nutrix* is considered as an indicator species of Vulnerable Marine Ecosystems (VME) by the CCAMLR.

See Guillaumot et al. (submitted) for more details

Usage

```
data(ctenocidaris.nutrix)
```

Format

A data frame containing 125 occurrences and 13 descriptive variables

- *id*
  Occurrence number indicator
- *scientific.name*
  Species scientific name
- *scientific.name.authorship*
  Author of the species description
- *genus*
  Genus scientific name and its associated author
- *family*
  Family scientific name and its associated author
- *order.and.higher.taxonomic.range*
  Order scientific name and its associated author
- *decimal.Longitude*
  Longitude in decimal degrees
- *decimal.Latitude*
  Latitude in decimal degrees
- *depth*
  Depth in meters

- *campaign*
  Campaign origin of the data
- *reference*
  Campaign reference
- *vessel*
  Campaign vessel

References

David B, Chone T, Mooi R, De Ridder C (2005) Antarctic Echinoidea. Synopses of the Antarctic Benthos 10.

Guillaumot C, A Martin, S Fabri-Ruiz, M Eleaume & T Saucede. Echinoids of the Kerguelen Plateau: Occurrence data and environmental setting for past, present, and future species distribution modelling, Zookeys, Manuscript submitted for publication.

Thomson CW (1876) Notice of some peculiarities in the mode of propagation of certain echinoderms of the southern seas. J. Linn. Soc. London 13: 55-79.

Examples

```
data(ctenocidaris.nutrix)
x <- ctenocidaris.nutrix
# plot of the occurrences:
# selecting the species according to the campaigns
ctenocidaris7475 <- base::subset(x,x$year==1974 | x$year==1975)
ctenocidaris20102015 <- base::subset(x,x$campaign=='POKER II' | x$campaign=='PROTEKER')

# drawing the background (depth)
library(grDevices)
blue.palette <- colorRampPalette(c('blue','deepskyblue','azure'))(100)
data('predictors1965_1974')
depth <- raster :: subset(predictors1965_1974, 1)

raster::plot(depth, col=blue.palette,main= "Ctenocidaris nutrix occurrences")

# adding the occurrences data to the background
points(ctenocidaris7475[,c('decimal.Longitude','decimal.Latitude')],
      col='orange',pch=16)
points(ctenocidaris20102015[,c('decimal.Longitude','decimal.Latitude')],
      col='darkgreen',pch=16)
legend('bottomleft',
       legend=c('Ctenocidaris nutrix 1974-1975','Ctenocidaris nutrix 2010-2015'),
       col= c('orange','darkgreen'), pch= c(15, 15),cex=0.9)
```

delim.area *RasterStack preparation for modelling*

Description

Delimit the RasterStack of environmental descriptors at a precise extent (latitude, longitude, maximum depth...) before computing species distribution modelling

Usage

```
delim.area(predictors, longmin, longmax, latmin, latmax, interval=NULL,
           crslayer = raster::crs(predictors))
```

Arguments

`predictors`	RasterStack object that contains the environmental predictors used for species distribution models
`longmin`	Expected minimum longitude of the RasterStack
`longmax`	Expected maximum longitude of the RasterStack
`latmin`	Expected minimum latitude of the RasterStack
`latmax`	Expected maximum latitude of the RasterStack
`interval`	Vector of 2. Minimum and maximum values outside of which the pixel values of the RasterStack first layer will be assigned to NA values. Set as NULL by default (no treatment).
`crslayer`	CRS object or character string describing a projection and datum. The crs of the original RasterStack is set by default

Details

interval enable the user to delimit the RasterStack according to an interval of values applied on the **first layer** of the RasterStack. It is often applied on depth in SDM studies.

Missing values contained in the provided RasterStack must be set up as NA values.

Value

RasterLayer object

See Also

stack, raster, origin, extent

Examples

```
data('predictors2005_2012')
envi <- predictors2005_2012

r <- SDMPlay:::delim.area(predictors = envi,
longmin = 70,longmax = 75, latmin = -50,latmax = -40,interval = c(0,-1000))
r

library(grDevices) # plot the result with nice colors
palet.col <- colorRampPalette(c('deepskyblue','green','yellow', 'red'))(80)
raster::plot(r, col=palet.col)
```

`null.model` *Compute null model*

Description

Compute null model. Null models are useful tools to enhance a priori due to occurrence spatial structuration in species distribution predictions (e.g. non homogeneous sampling).
Null model 'type 1' performs a model by randomly sampling data in a matrix that provides the visited longitudes and latitudes. Model null type #1 highlights extra information about prior influences of sampling effort on models.

Null model type #2 samples data in the entire study area, and reflects what should be predicted if occurrences were randomly distributed in the area.

Null models are a compilation of *nb.rep* models. Species distribution models can be compared to density distribution of the null model outputs in order to estimate signification rates.

Usage

```
null.model(predictors, xy = NULL, type = c(1, 2), algorithm = c("brt", "maxent"), nb,
         unique.data = T, same = T, background.nb = nb, nb.rep = 10, tc = 2,
         lr = 0.001, bf = 0.75, n.trees = 50, step.size = n.trees)
```

Arguments

`predictors`	Rasterstack object that contains the predictors that will be used for species distribution models
`xy`	Dataframe that contains the longitude and latitude of the visited pixels. Information required to perform type 1 null model. Default= NULL
`type`	Null model type to perform. type=1 to perform a null model based on visited areas, type=2 to predict random model
`algorithm`	Algorithm to compute the null model. 'brt' or 'maxent'
`nb`	Number of points to randomly sample (among the matrix of visited pixels for 'type=1' model or in the entire geographic space for 'type=2')
`unique.data`	If TRUE (default), pixel duplicates contained in 'xy' are removed

same	If TRUE (default), the number of background data sampled in the area will be 'nb'
background.nb	Number of background data to sample. If this argument is filled, 'same' is set FALSE.
nb.rep	Null models number of replicates. See compute.brt
tc	BRT parameter. Integer. Tree complexity. Sets the complexity of individual trees. See compute.brt
lr	BRT parameter.Learning rate. Sets the weight applied to individual trees. See compute.brt
bf	BRT parameter.Bag fraction. Sets the proportion of observations used in selecting variables. See compute.brt
n.trees	BRT parameter.Number of initial trees to fit. Set at 50 by default. See compute.brt
step.size	BRT parameter.Number of trees to add at each cycle. See compute.brt

Details

Data are sampled without replacement. Each time the model is runned, new data (presence-like and background data) are sampled

Value

List of 6

- *$inputs* Remembers the arguments used to implement null.model function
- *$eval* Evaluation parameters of each model that compose the null model. See SDMeval for further information
- *$eval.null* Evaluation of the mean null model. See SDMeval for further information
- *$pred.stack* RasterStack of all the models produced to build the null model
- *$pred.mean* Raster layer. Null model prediction. Mean of the $pred.stack RasterStack
- *$correlation* Spearman rank test value between the different maps produced

Note

Increasing the number of replications will enhance model null relevance (we advice nb.rep=100 for minimum). Please note that processing may take few minutes to hours.

If you want to build a MaxEnt model, compute.maxent uses the functionalities of the maxent function. This function uses MaxEnt species distribution software, which is a java program that could be downloaded at `http://www.cs.princeton.edu/~schapire/maxent/`. In order to run compute.maxent, put the 'maxent.jar' file downloaded at this adress in the 'java' folder of the dismo package (path obtained with the system.file('java', package='dismo') command). MaxEnt 3.3.3b version or higher is required.

See Also

nicheOverlap: compare prediction maps .jpackage: initialize dismo for Java

Examples

```
## Not run:
library(dismo)
#Download the environmental predictors restricted on geographical extent and depth (-1500m)
envi <-raster::stack(system.file('extdata', 'pred.grd',package='SDMPlay'))

# Realize a null model type #2 with BRT
#--------------------------------------
# NB: the following arguments chosen for the example are not relevant,
# in the scope to minimize running time
modelN2 <- SDMPlay:::null.model(xy=NULL,predictors=envi,type=2,algorithm='brt',
                     nb=300,unique.data=TRUE, same=TRUE, nb.rep=2,lr=0.005)

# Look at the inputs used to implement the model
modelN2$input

# Get the evaluation of the models produced
modelN2$eval

# Get the evaluation of the mean of all these produced models (i.e. evaluation
# of the null model )
modelN2$eval.null

# Get the values of Spearman correlations between the all the prediction maps produced
modelN2$correlation

# Plot the mean null model map with nice colors
library(grDevices)
palet.col <- colorRampPalette(c('deepskyblue','green','yellow', 'red'))(80)
raster::plot(modelN2$pred.mean, col=palet.col)

## End(Not run)
```

`predictors1965_1974` *Environmental descriptors for 1965-1974 on the Kerguelen Plateau*

Description

RasterStack that compiles 15 environmental descriptors on the Kerguelen Plateau (63/81W; -46/-56S). See Guillaumot et al. (2016) for more information

Usage

```
data(predictors1965_1974)
```

Format

RasterStack of 15 environmental descriptors. Grid: nrow= 100, ncol= 179, ncells= 17900 pixels. Spatial resolution: 0.1. Spatial extent: 63/81W; -46/-56S.
Crs : +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0. Origin=0

- **depth**
  Bathymetric grid around the Kerguelen Plateau
  Unit=meter. Reference=Guillaumot et al. (2016), derived from Smith & Sandwell (1997)
  http://topex.ucsd.edu/WWW_html/mar_topo.html
- **seasurface_temperature_mean_1965_1974**
  Mean sea surface temperature over 1965-1974
  Unit=Celsius degrees. Reference= World Ocean Circulation Experiment 2013
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **seasurface_temperature_amplitude_1965_1974**
  Amplitude between mean summer and mean winter sea surface temperature over 1965-1974
  Unit=Celsius degrees. Reference= World Ocean Circulation Experiment 2013
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **seafloor_temperature_mean_1965_1974**
  Mean seafloor temperature over 1965-1974
  Unit=Celsius degrees. Reference=Guillaumot et al. (2016), derived from World Ocean Circulation Experiment 2013 sea surface temperature layers
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **seafloor_temperature_amplitude_1965_1974**
  Amplitude between mean summer and mean winter seafloor temperature over 1965-1974
  Unit=Celsius degrees. Reference=Guillaumot et al. (2016), derived from World Ocean Circulation Experiment 2013 sea surface temperature layers
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **seasurface_salinity_mean_1965_1974**
  Mean sea surface salinity over 1965-1974
  Unit=PSS. Reference= World Ocean Circulation Experiment 2013
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **seasurface_salinity_amplitude_1965_1974**
  Amplitude between mean summer and mean winter sea surface salinity over 1965-1974
  Unit=PSS. Reference= World Ocean Circulation Experiment 2013
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **seafloor_salinity_mean_1965_1974**
  Mean seafloor salinity over 1965-1974
  Unit=PSS. Reference= Guillaumot et al. (2016), derived from World Ocean Circulation Experiment 2013 sea surface salinity layers
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **seafloor_salinity_amplitude_1965_1974**
  Amplitude between mean summer and mean winter seafloor salinity over 1965-1974
  Unit=PSS. Reference= Guillaumot et al. (2016,submitted), derived from World Ocean Circulation Experiment 2013 sea surface salinity layers
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **chlorophyla_summer_mean_2002_2009**
  Surface chlorophyll a concentration. Summer mean over 2002-2009
  Unit=mg/m3.Reference=MODIS AQUA (NASA) 2010
  http://oceandata.sci.gsfc.nasa.gov/MODIS-Aqua/Mapped
- **geomorphology**
  Geomorphologic features
  Unit= 27 categories. Reference= ATLAS ETOPO2 2014 (Douglass et al. 2014)

- **sediments**
  Sediment features
  Unit= 14 categories. Reference= McCoy (1991), updated by Griffiths 2014 (unpublished)
- **slope**
  Bathymetric slope
  Unitless. Reference= Smith & Sandwell (1997)
- **seafloor_oxygen_mean_1955_2012**
  Mean seafloor oxygen concentration over 1955-2012
  Unit=mL/L. Reference= Guillaumot et al. (2016), derived from World Ocean Circulation Experiment 2013 sea surface oxygen concentration layers
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **roughness**
  Rugosity index (difference between minimal and maximal depth values of the 8 neighbour-pixels)
  Unit= meters. Reference=Guillaumot et al.(2016), derived from bathymetric layer

References

Douglass L, Turner J, Grantham HS, Kaiser S, Constable A, Nicoll R, Raymond B, Post A, Brandt A, Beaver D (2014) A hierarchical classification of benthic biodiversity and assessment of protected areas in the Southern Ocean. PloS one 9(7): e100551. doi: 10.1371/journal.pone.0100551.

Guillaumot, C., Martin , A., Fabri-Ruiz, S., Eleaume, M. and Saucede, T. (2016) Environmental parameters (1955-2012) for echinoids distribution modelling on the Kerguelen Plateau. Australian Antarctic Data Centre - doi:10.4225/15/578ED5A08050F

McCoy FW (1991) Southern Ocean sediments: circum-Antarctic to 30S. Marine Geological and Geophysical Atlas of the circum-Antarctic to 30S. (ed. by D.E. Hayes). Antarctic Research Series.

Smith W, Sandwell D (1997) Global seafloor topography from satellite altimetry and ship depth soundings. Science 277(5334): 1957-1962. doi: 10.1126/science.277.5334.1956.

Examples

```
data(predictors1965_1974)
raster::plot(predictors1965_1974)
```

---

predictors2005_2012 *Environmental descriptors for 2005-2012 on the Kerguelen Plateau*

Description

RasterStack that compiles 15 environmental descriptors on the Kerguelen Plateau (63/81W; -46/-56S). See Guillaumot et al. (2016) for more information

Usage

```
data(predictors2005_2012)
```

Format

RasterStack of 15 environmental descriptors. Grid: nrow= 100, ncol= 179, ncells= 17900 pixels. Spatial resolution: 0.1. Spatial extent: 63/81W; -46/-56S.
Crs : +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0. Origin=0

- **depth**
  Bathymetric grid around the Kerguelen Plateau
  Unit=meter. Reference=Guillaumot et al. (2016), derived from Smith & Sandwell (1997)
  `http://topex.ucsd.edu/WWW_html/mar_topo.html`
- **seasurface_temperature_mean_2005_2012**
  Mean sea surface temperature over 2005-2012
  Unit=Celsius degrees. Reference= World Ocean Circulation Experiment 2013
  `https://www.nodc.noaa.gov/OC5/woa13/woa13data.html`
- **seasurface_temperature_amplitude_2005_2012**
  Amplitude between mean summer and mean winter sea surface temperature over 2005-2012
  Unit=Celsius degrees. Reference= World Ocean Circulation Experiment 2013
  `https://www.nodc.noaa.gov/OC5/woa13/woa13data.html`
- **seafloor_temperature_mean_2005_2012**
  Mean seafloor temperature over 2005-2012
  Unit=Celsius degrees. Reference=Guillaumot et al. (2016), derived from World Ocean Circulation Experiment 2013 sea surface temperature layers
  `https://www.nodc.noaa.gov/OC5/woa13/woa13data.html`
- **seafloor_temperature_amplitude_2005_2012**
  Amplitude between mean summer and mean winter seafloor temperature over 2005-2012
  Unit=Celsius degrees. Reference=Guillaumot et al. (2016,submitted), derived from World Ocean Circulation Experiment 2013 sea surface temperature layers
  `https://www.nodc.noaa.gov/OC5/woa13/woa13data.html`
- **seasurface_salinity_mean_2005_2012**
  Mean sea surface salinity over 2005-2012
  Unit=PSS. Reference= World Ocean Circulation Experiment 2013
  `https://www.nodc.noaa.gov/OC5/woa13/woa13data.html`
- **seasurface_salinity_amplitude_2005_2012**
  Amplitude between mean summer and mean winter sea surface salinity over 2005-2012
  Unit=PSS. Reference= World Ocean Circulation Experiment 2013
  `https://www.nodc.noaa.gov/OC5/woa13/woa13data.html`
- **seafloor_salinity_mean_2005_2012**
  Mean seafloor salinity over 2005-2012
  Unit=PSS. Reference= Guillaumot et al. (2016), derived from World Ocean Circulation Experiment 2013 sea surface salinity layers.
  `https://www.nodc.noaa.gov/OC5/woa13/woa13data.html`
- **seafloor_salinity_amplitude_2005_2012**
  Amplitude between mean summer and mean winter seafloor salinity over 2005-2012
  Unit=PSS. Reference= Guillaumot et al. (2016), derived from World Ocean Circulation Experiment 2013 sea surface salinity layers
  `https://www.nodc.noaa.gov/OC5/woa13/woa13data.html`

- **chlorophyla_summer_mean_2002_2009**
  Surface chlorophyll a concentration. Summer mean over 2002-2009
  Unit=mg/m3.Reference=MODIS AQUA (NASA) 2010
  http://oceandata.sci.gsfc.nasa.gov/MODIS-Aqua/Mapped
- **geomorphology**
  Geomorphologic features
  Unit= 27 categories. Reference= ATLAS ETOPO2 2014 (Douglass et al. 2014)
- **sediments**
  Sediment features
  Unit= 14 categories. Reference= McCoy (1991), updated by Griffiths 2014 (unpublished).
- **slope**
  Bathymetric slope
  Unitless. Reference= Smith & Sandwell (1997)
- **seafloor_oxygen_mean_1955_2012**
  Mean seafloor oxygen concentration over 1955-2012
  Unit=mL/L. Reference= Guillaumot et al. (2016), derived from World Ocean Circulation Experiment 2013 sea surface oxygen concentration layers
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **roughness**
  Rugosity index (difference between minimal and maximal depth values of the 8 neighbour-pixels)
  Unit= meters. Reference=Guillaumot et al.(2016), derived from bathymetric layer

References

Douglass L, Turner J, Grantham HS, Kaiser S, Constable A, Nicoll R, Raymond B, Post A, Brandt A, Beaver D (2014) A hierarchical classification of benthic biodiversity and assessment of protected areas in the Southern Ocean. PloS one 9(7): e100551. doi: 10.1371/journal.pone.0100551.

Guillaumot, C., Martin , A., Fabri-Ruiz, S., Eleaume, M. and Saucede, T. (2016) Environmental parameters (1955-2012) for echinoids distribution modelling on the Kerguelen Plateau. Australian Antarctic Data Centre - doi:10.4225/15/578ED5A08050F

McCoy FW (1991) Southern Ocean sediments: circum-Antarctic to 30S. Marine Geological and Geophysical Atlas of the circum-Antarctic to 30S. (ed. by D.E. Hayes). Antarctic Research Series.

Smith W, Sandwell D (1997) Global seafloor topography from satellite altimetry and ship depth soundings. Science 277(5334): 1957-1962. doi: 10.1126/science.277.5334.1956.

Examples

```
data(predictors2005_2012)
raster::plot(predictors2005_2012)
```

predictors2200AIB *IPCC environmental descriptors predicted for 2200 (AIB scenario) on the Kerguelen Plateau*

Description

RasterStack of 10 environmental descriptors modelled by IPCC (scenario AIB, 4th report, 2007) for 2187 to 2196 (described as 2200), on the extent of the Kerguelen Plateau (63/81W; -46/-56S)

Usage

```
data(predictors2200AIB)
```

Format

RasterStack of 10 environmental descriptors. Grid: nrow= 100, ncol= 179, ncells= 17900 pixels.
Spatial resolution: 0.1. Spatial extent: 63/81W; -46/-56S.
Crs : +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0. Origin=0.
See Guillaumot et al.(2016) for more information

- **depth**
  Bathymetric grid around the Kerguelen Plateau
  Unit=meter. Reference=Guillaumot et al. (2016), derived from Smith & Sandwell (1997)
  http://topex.ucsd.edu/WWW_html/mar_topo.html
- **seasurface_salinity_mean_2200_A1B**
  Mean sea surface salinity over 2187 to 2196, A1B scenario
  Unit= PSS. Reference= BIO ORACLE (Tyberghein et al. 2012)
  http://www.oracle.ugent.be/
- **seasurface_temperature_mean_2200_A1B**
  Mean sea surface temperature over 2187-2196, A1B scenario
  Unit=Celsius degrees. Reference= BIO ORACLE (Tyberghein et al. 2012)
  http://www.oracle.ugent.be/
- **seasurface_temperature_amplitude_2200_A1B**
  Amplitude between mean summer and mean winter sea surface temperature. Absolute value interpolated over 2187-2196, scenario A1B
  Unit=Celsius degrees. Reference= BIO ORACLE (Tyberghein et al. 2012)
  http://www.oracle.ugent.be/
- **chlorophyla_summer_mean_2002_2009**
  Surface chlorophyll a concentration. Summer mean over 2002-2009
  Unit=mg/m3. Reference=MODIS AQUA (NASA) 2010
  http://oceandata.sci.gsfc.nasa.gov/MODIS-Aqua/Mapped
- **geomorphology**
  Geomorphologic features
  Unit= 27 categories. Reference= ATLAS ETOPO2 2014 (Douglass et al. 2014)

- **sediments**
  Sediment features
  Unit= 14 categories. Reference= McCoy (1991), updated by Griffiths 2014 (unpublished)
- **slope**
  Bathymetric slope
  Unitless. Reference= Smith & Sandwell (1997)
- **seafloor_oxygen_mean_1955_2012**
  Mean seafloor oxygen concentration over 1955-2012
  Unit=mL/L. Reference= Guillaumot et al. (2016), derived from World Ocean Circulation Experiment 2013 sea surface oxygen concentration layers
  https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
- **roughness**
  Rugosity index (difference between minimal and maximal depth values of the 8 neighbour-pixels)
  Unit= meters. Reference=Guillaumot et al.(2016), derived from bathymetric layer.

References

Douglass LL, Turner J, Grantham HS, Kaiser S, Constable A, Nicoll R, Raymond B, Post A, Brandt A, Beaver D (2014) A hierarchical classification of benthic biodiversity and assessment of protected areas in the Southern Ocean. PloS one 9(7): e100551. doi: 10.1371/journal.pone.0100551.

Guillaumot, C., Martin , A., Fabri-Ruiz, S., Eleaume, M. and Saucede, T. (2016) Environmental parameters (1955-2012) for echinoids distribution modelling on the Kerguelen Plateau. Australian Antarctic Data Centre - doi:10.4225/15/578ED5A08050F

Jueterbock A, Tyberghein L, Verbruggen H, Coyer JA, Olsen JL, Hoarau G (2013) Climate change impact on seaweed meadow distribution in the North Atlantic rocky intertidal. Ecology and Evolution 3(5): 1356-1373. doi: 10.1002/ece3.541.

McCoy FW (1991) Southern Ocean sediments: circum-Antarctic to 30S. Marine Geological and Geophysical Atlas of the circum-Antarctic to 30S. (ed. by D.E. Hayes). Antarctic Research Series.

Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F, De Clerck O (2012) Bio ORACLE: a global environmental dataset for marine species distribution modelling. Global Ecology and Biogeography 21(2): 272-28. doi: 10.1111/j.1466-8238.2011.00656.x.

Smith W, Sandwell D (1997) Global seafloor topography from satellite altimetry and ship depth soundings. Science 277(5334): 1957-1962. doi: 10.1126/science.277.5334.1956.

Examples

```
data(predictors2200AIB)
raster :: plot(predictors2200AIB)
```

SDMdata.quality — *Evaluate dataset quality*

Description

Evaluate the percentage of occurrences that fall on pixels assigned by NA values in the environmental RasterStack. It may provide interesting information to interpret model robustness.

Usage

```
SDMdata.quality(data)
```

Arguments

data	SDMtab object or dataframe that contains id, longitude, latitude and values of environmental descriptors at corresponding locations

Value

prop Dataframe that provides the proportion of NA values on which the presence data fall, for each environmental predictor

See Also

SDMeval

Examples

```
#Open SDMtab object example
x <- system.file ("extdata","SDMdata1500.csv", package="SDMPlay")
SDMdata <- read.table(x,header=TRUE, sep=";")

# Evaluate the dataset
SDMPlay:::SDMdata.quality(data=SDMdata)
```

SDMeval — *Evaluate species distribution models*

Description

Performs model evaluation. Measure of AUC (Area Under the Curve) value, confusion matrix, maxSSS threshold (Maximum Sensitivity plus Specificity), percentage of predicted preferential area based on the MaxSSS value and model stability (standard deviation of pixel values)

Usage

```
SDMeval(model)
```

Arguments

`model` Model produced with compute.maxent or compute.brt functions

Details

Area Under the Curve is a parameter largely refered in the literature and used to test species distribution models performance (Fielding & Bell, 1997). It evaluates the area under the Receiver Operating Curve (ROC), which draws the relationship between 1-specificity (False Positive Rate) and specificity (True Positive Rate). AUC values bordering 1 present models with high True Positive Rate, 0.5 model with random prediction and 0 to models presenting a strong False Positive Rate.

MaxSSS threshold value maximizes the sum of True Positive Rate and True Negative Rate. See Liu et al. (2013) for more information.

Modelling performance can be evaluated with the measure of omission rate, the proportion of occurrences that falls out the area predicted as preferential by the MaxSSS threshold (False Positive Rate). Models stability is evaluated with the mean standard deviation value of the pixel values of the grid predicted by the model.

Value

Dataframe with the following information

- *AUC.value* Returns the AUC (Area Under the Curve) value of the model
- *maxSSS* Maximum Sensitivity plus Sensibility threshold of the model
- *preferential.area* Pixel proportion for which the predicted value is superior to the MaxSSS threshold
- *omission.rate* Proportion of data that fall out of the area predicted as preferential
- *nb.omission* Corresponding number of data that fall out of the predicted preferential area
- *SD.value* Mean standard deviation of the predicted grid

References

Fielding A, & J Bell (1997) A review of methods for the assessment of prediction errors in conservation presence absence models. *Environmental Conservation*, 24(1): 38-49.

Liu C, M White & G Newell (2013) Selecting thresholds for the prediction of species occurrence with presence only data. *Journal of Biogeography*, 40(4): 778-789.

Examples

```
# Model example
load(system.file('extdata','model.RData',package='SDMPlay'))
modelBRT <- model

# Evaluate modelling performance
#SDMPlay:::SDMeval(modelBRT)
```

SDMtab — *Compile species distribution dataset for modelling*

Description

Create a dataframe that contains the required information to implement species distribution models

Usage

```
SDMtab(xydata, predictors, unique.data = TRUE, same = TRUE, background.nb=NULL)
```

Arguments

`xydata`	Dataframe with longitude (column 1) and latitude (column 2) of the presence-only data. Decimal longitude and latitude are required.
`predictors`	Rasterstack of environmental descriptors. Used to extract values of the presence location
`unique.data`	If TRUE (by default), duplicate presence points, that fall in the same grid cell, will be removed
`same`	If TRUE (by default), the number of background data sampled in the area equals the number of presence data
`background.nb`	Set as NULL if same= TRUE.

Details

Background data are sampled randomly (without replacement) among the entire area, on pixels that are not assigned NA. It constitutes a summary of environmental descriptors to improve modelling performance. See Barbet Massin et al. (2012) for further information about background selection.

Value

A dataframe that contains the id (1 for presence, 0 for background data) of data, their longitude, latitude and extracted values of environmental descriptors at the corresponding locations.

xydata for which coordinates fall out of the RasterStack extent are removed from the analysis.

References

Barbet Massin M, F Jiguet, C Albert & W Thuiller (2012) Selecting pseudo absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2): 327-338.

See Also

delim.area to refine the environmental RasterStack before using this function

Examples

```
#Open occurrence data
data('ctenocidaris.nutrix')
occ <- ctenocidaris.nutrix

#Open environmental descriptors RasterStack
r <- raster:: stack(system.file('extdata', 'pred.grd',package='SDMPlay'))

#create the dataframe for modelling
z <- SDMPlay:::SDMtab(xydata=occ[,c('decimal.Longitude','decimal.Latitude')],predictors=r)
head(z)
```

Index