Package ‘SSDM’

February 28, 2020

Type Package
Title Stacked Species Distribution Modelling
Version 0.2.8
Author Sylvain Schmitt, Robin Pouteau, Dimitri Justeau, Florian de Boissieu, Lukas Baum-bach, Philippe Birnbaum
Maintainer Sylvain Schmitt <sylvain.m.schmitt@gmail.com>
URL https://github.com/sylvainschmitt/SSDM
BugReports https://github.com/sylvainschmitt/SSDM/issues
Description Allows to map species richness and endemism based on stacked species distribution models (SSDM). Individuals SDMs can be created using a single or multiple algorithms (ensemble SDMs). For each species, an SDM can yield a habitat suitability map, a binary map, a between-algorithm variance map, and can assess variable importance, algorithm accuracy, and between-algorithm correlation. Methods to stack individual SDMs include summing individual probabilities and thresholding then summing. Thresholding can be based on a specific evaluation metric or by drawing repeatedly from a Bernoulli distribution. The SSDM package also provides a user-friendly interface.
License GPL (>= 3) | file LICENSE
LazyData TRUE
Imports sp (>= 1.2.0), raster (>= 2.9-5), methods (>= 3.2.2), mgcv (>= 1.8.7), earth (>= 4.4.3), rpart (>= 4.1.10), gbm (>= 2.1.1), randomForest (>= 4.6.10), dismo (>= 1.0.12), nnet (>= 7.3.10), e1071 (>= 1.6.7), ggplot2 (>= 3.1.1), reshape2 (>= 1.4.3), scales (>= 1.0.0), shiny (>= 0.12.2), shinydashboard (>= 0.5.1), shinyFiles (>= 0.7.0), spThin (>= 0.1.0), poibin (>= 1.3.0)
Depends R (>= 3.2.2)
Collate 'SDM.R' 'Algorithm.SDM.R' 'Ensemble.SDM.R' 'Env.R' 'Occurrences.R' 'PA.select.R' 'SSDM.R' 'Stacked.SDM.R' 'accuracy.R' 'checkargs.R' 'data.values.R' 'ensemble.R' 'modelling.R' 'ensemble_modelling.R' 'optim.thresh.R' 'evaluate.R' 'evaluate.axes.R' 'get_PA.R' 'get_model.R' 'gui.R'
Algorithm.SDM-class

An S4 class to represent an SDM based on a single algorithm

Description

This is an S4 class to represent an SDM based on a single algorithm (including generalized linear model, general additive model, multivariate adaptive splines, generalized boosted regression model, classification tree analysis, random forest, maximum entropy, artificial neural network, and support vector machines). This S4 class is obtained with modelling.
**Slots**

- name: character. Name of the SDM (by default Species.SDM).
- projection: raster. Habitat suitability map produced by the SDM.
- binary: raster. Presence/Absence binary map produced by the SDM.
- evaluation: data frame. Evaluation of the SDM (available metrics include AUC, Kappa, sensitivity, specificity and proportion of correctly predicted occurrences) and identification of the optimal threshold to convert the habitat suitability map into a binary presence/absence map.
- variable.importance: data frame. Relative importance of each variable in the SDM.
- data: data frame. Data used to build the SDM.
- parameters: data frame. Parameters used to build the SDM.

**See Also**

- **Ensemble.SDM** an S4 class for ensemble SDMs, and **Stacked.SDM** an S4 class for SSDMs.

---

**Description**

This is a method to assemble several algorithms in an ensemble SDM. The function takes as inputs several S4 **Algorithm.SDM** class objects returned by the `modelling` function. The function returns an S4 **Ensemble.SDM** class object containing the habitat suitability map, the binary map, and the uncertainty map (based on the between-algorithm variance) and the associated evaluation tables (model evaluation, algorithm evaluation, algorithm correlation matrix and variable importance).

**Usage**

```r
ensemble(x, ..., name = NULL, ensemble.metric = c("AUC"),
        ensemble.thresh = c(0.75), weight = TRUE, thresh = 1001,
        uncertainty = TRUE, verbose = TRUE, GUI = FALSE)

## S4 method for signature 'Algorithm.SDM'
ensemble(x, ..., name = NULL,
        ensemble.metric = c("AUC"), ensemble.thresh = c(0.75),
        weight = TRUE, thresh = 1001, uncertainty = TRUE, verbose = TRUE,
        GUI = FALSE)

## S4 method for signature 'Algorithm.SDM'
sum(x, ..., name = NULL,
      ensemble.metric = c("AUC"), ensemble.thresh = c(0.75),
      weight = TRUE, thresh = 1001, format = TRUE, verbose = TRUE,
      na.rm = TRUE)
```
Arguments

- **x, ...**: SDMs. SDMs to be assembled.
- **name**: character. Optional name given to the final Ensemble.SDM produced (by default 'Ensemble.SDM').
- **ensemble.metric**: character. Metric(s) used to select the best SDMs that will be included in the ensemble SDM (see details below).
- **ensemble.thresh**: numeric. Threshold(s) associated with the metric(s) used to compute the selection.
- **weight**: logical. If TRUE, SDMs are weighted using the ensemble metric or, alternatively, the mean of the selection metrics.
- **thresh**: numeric. A integer value specifying the number of equal interval threshold values between 0 and 1.
- **uncertainty**: logical. If TRUE, generates an uncertainty map and an algorithm correlation matrix.
- **verbose**: logical. If set to true, allows the function to print text in the console.
- **GUI, format, na.rm**: logical. Do not take those arguments into account (parameters for the user interface and sum function).

Details

- **ensemble.metric** (metric(s) used to select the best SDMs that will be included in the ensemble SDM) can be chosen from among:
  - **AUC**: Area under the receiver operating characteristic (ROC) curve
  - **Kappa**: Kappa from the confusion matrix
  - **sensitivity**: Sensitivity from the confusion matrix
  - **specificity**: Specificity from the confusion matrix
  - **prop.correct**: Proportion of correctly predicted occurrences from the confusion matrix

Value

- an S4 **Ensemble.SDM** class object viewable with the **plot.model** function.

See Also

- **ensemble_modelling** to build an ensemble SDM from multiple algorithms.

Examples

```r
## Not run:
# Loading data
data(Env)
data(Occurrences)
```
Ensemble.SDM-class

An S4 class to represent an ensemble SDM

Description

This is an S4 class to represent an ensemble SDM from multiple algorithms (including generalized linear model, general additive model, multivariate adaptive splines, generalized boosted regression model, classification tree analysis, random forest, maximum entropy, artificial neural network, and support vector machines). This S4 class is returned by `ensemble_modelling` or `ensemble`.

Slots

- **uncertainty** raster. Between-algorithm variance map.
- **algorithm.correlation** data frame. Between-algorithm correlation matrix.
- **algorithm.evaluation** data frame. Evaluation of the ensemble SDM (available
  - **sdms** list. Individual SDMs used to create the ESDM. metrics include AUC, Kappa, sensitivity, specificity and proportion of correctly predicted occurrences) and identification of the optimal threshold to convert the habitat suitability map into a binary presence/absence map.

See Also

- **Algorithm.SDM** an S4 class to represent an SDM based on a single algorithm, and **Stacked.SDM** an S4 class for SSDMs.

- **ensemble_modelling** Build an ensemble SDM that assembles multiple algorithms

Description

Build an ensemble SDM that assembles multiple algorithms for a single species. The function takes as inputs an occurrence data frame made of presence/absence or presence-only records and a raster object for data extraction and projection. The function returns an S4 **Ensemble.SDM** class object containing the habitat suitability map, the binary map, the between-algorithm variance map and the associated evaluation tables (model evaluation, algorithm evaluation, algorithm correlation matrix and variable importance).
ensemble_modelling

Usage

ensemble_modelling(algorithms, Occurrences, Env, Xcol = "Longitude",
Ycol = "Latitude", Pcol = NULL, rep = 10, name = NULL,
save = FALSE, path = getwd(), PA = NULL, cv = "holdout",
cv.param = c(0.7, 1), thresh = 1001, metric = "SES",
axes.metric = "Pearson", uncertainty = TRUE, tmp = FALSE,
ensemble.metric = c("AUC"), ensemble.thresh = c(0.75),
weight = TRUE, verbose = TRUE, GUI = FALSE, ...)

Arguments

- **algorithms**: character. A character vector specifying the algorithm name(s) to be run (see details below).
- **Occurrences**: data frame. Occurrences table (can be processed first by `load_occ`).
- **Env**: raster object. RasterStack object of environmental variables (can be processed first by `load_var`).
- **Xcol**: character. Name of the column in the occurrence table containing Latitude or X coordinates.
- **Ycol**: character. Name of the column in the occurrence table containing Longitude or Y coordinates.
- **Pcol**: character. Name of the column in the occurrence table specifying whether a line is a presence or an absence. A value of 1 is presence and value of 0 is absence. If NULL, presence-only dataset is assumed.
- **rep**: integer. Number of repetitions for each algorithm.
- **name**: character. Optional name given to the final Ensemble.SDM produced (by default 'Ensemble.SDM').
- **save**: logical. If TRUE, the ensemble SDM is automatically saved.
- **path**: character. If save is TRUE, the path to the directory in which the ensemble SDM will be saved.
- **PA**: list(nb, strat) defining the pseudo-absence selection strategy used in case of presence-only dataset. If PA is NULL, recommended PA selection strategy is used depending on the algorithm (see details below).
- **cv**: character. Method of cross-validation used to evaluate the ensemble SDM (see details below).
- **cv.param**: numeric. Parameters associated to the method of cross-validation used to evaluate the ensemble SDM (see details below).
- **thresh**: numeric. A single integer value representing the number of equal interval threshold values between 0 and 1.
- **metric**: character. Metric used to compute the binary map threshold (see details below).
- **axes.metric**: Metric used to evaluate variable relative importance (see details below).
- **uncertainty**: logical. If TRUE, generates an uncertainty map and an algorithm correlation matrix.
tmp logical. If set to true, the habitat suitability map of each algorithm is saved in a temporary file to release memory. But beware: if you close R, temporary files will be deleted. To avoid any loss, you can save your ensemble SDM with `save.model`. Depending on number, resolution, and extent of models, temporary files can take a lot of disk space. Temporary files are written in R environment temporary folder.

ensemble.metric character. Metric(s) used to select the best SDMs that will be included in the ensemble SDM (see details below).

ensemble.thresh numeric. Threshold(s) associated with the metric(s) used to compute the selection.

weight logical. If TRUE, SDMs are weighted using the ensemble metric or, alternatively, the mean of the selection metrics.

verbose logical. If TRUE, allows the function to print text in the console.

GUI logical. Do not take this argument into account (parameter for the user interface).

... additional parameters for the algorithm modelling function (see details below).

Details

**algorithms** 'all' calls all the following algorithms. Algorithms include Generalized linear model (GLM), Generalized additive model (GAM), Multivariate adaptive regression splines (MARS), Generalized boosted regressions model (GBM), Classification tree analysis (CTA), Random forest (RF), Maximum entropy (MAXENT), Artificial neural network (ANN), and Support vector machines (SVM). Each algorithm has its own parameters settable with the ... (see each algorithm section below to set their parameters).

"PA" list with two values: nb number of pseudo-absences selected, and strat strategy used to select pseudo-absences: either random selection or disk selection. We set default recommendation from Barbet-Massin et al. (2012) (see reference).

**cv** Cross-validation method used to split the occurrence dataset used for evaluation: holdout data are partitioned into a training set and an evaluation set using a fraction (cv.param[1]) and the operation can be repeated (cv.param[2]) times, k-fold data are partitioned into k (cv.param[1]) folds being k-1 times in the training set and once the evaluation set and the operation can be repeated (cv.param[2]) times, LOO (Leave One Out) each point is successively taken as evaluation data.

**metric** Choice of the metric used to compute the binary map threshold and the confusion matrix (by default SES as recommended by Liu et al. (2005), see reference below): Kappa maximizes the Kappa, CCR maximizes the proportion of correctly predicted observations, TSS (True Skill Statistic) maximizes the sum of sensitivity and specificity, SES uses the sensitivity-specificity equality, LW uses the lowest occurrence prediction probability, ROC minimizes the distance between the ROC plot (receiving operating characteristic curve) and the upper left corner (1,1).

**axes.metric** Metric used to evaluate the variable relative importance (difference between a full model and one with each variable successively omitted): Pearson computes a simple Pearson’s correlation r between predictions of the full model and the one without a variable, and returns the score 1-r: the highest the value, the more influence the variable has on the model),
AUC, Kappa, sensitivity, specificity, and prop.correct (proportion of correctly predicted occurrences).

ensemble.metric Ensemble metric(s) used to select SDMs: AUC, Kappa, sensitivity, specificity, and prop.correct (proportion of correctly predicted occurrences).

"..." See algorithm in detail section

Value

an S4 Ensemble.SDM class object viewable with the plot.model function.

Generalized linear model (GLM)

Uses the glm function from the package 'stats', you can set the following parameters (see glm for more details):

test character. Test used to evaluate the SDM, default 'AIC'.
epsilon numeric. Positive convergence tolerance eps; the iterations converge when \(|dev - dev_{old}|/(|dev| + 0.1) < eps\). By default, set to 10e-08.
maxit numeric. Integer giving the maximal number of IWLS (Iterative Weighted Last Squares) iterations, default 500.

Generalized additive model (GAM)

Uses the gam function from the package 'mgcv', you can set the following parameters (see gam for more details):

test character. Test used to evaluate the model, default 'AIC'.
epsilon numeric. This is used for judging conversion of the GLM IRLS (Iteratively Reweighted Least Squares) loop, default 10e-08.
maxit numeric. Maximum number of IRLS iterations to perform, default 500.

Multivariate adaptive regression splines (MARS)

Uses the earth function from the package 'earth', you can set the following parameters (see earth for more details):

degree integer. Maximum degree of interaction (Friedman’s mi); 1 meaning build an additive model (i.e., no interaction terms). By default, set to 2.

Generalized boosted regressions model (GBM)

Uses the gbm function from the package 'gbm,' you can set the following parameters (see gbm for more details):

trees integer. The total number of trees to fit. This is equivalent to the number of iterations and the number of basis functions in the additive expansion. By default, set to 2500.
final.leave integer. minimum number of observations in the trees terminal nodes. Note that this is the actual number of observations not the total weight. By default, set to 1.
algcov integer. Number of cross-validations, default 3.
ensemble_modelling

thresh.shrink integer. Number of cross-validation folds to perform. If cv.folds>1 then gbm, in addition to the usual fit, will perform a cross-validation. By default, set to 1e-03.

Classification tree analysis (CTA)

Uses the rpart function from the package 'rpart', you can set the following parameters (see rpart for more details):

final.leave integer. The minimum number of observations in any terminal node, default 1.
algo cv integer. Number of cross-validations, default 3.

Random Forest (RF)

Uses the randomForest function from the package 'randomForest', you can set the following parameters (see randomForest for more details):

trees integer. Number of trees to grow. This should not be set to a too small number, to ensure that every input row gets predicted at least a few times. By default, set to 2500.

final.leave integer. Minimum size of terminal nodes. Setting this number larger causes smaller trees to be grown (and thus take less time). By default, set to 1.

Maximum Entropy (MAXENT)

Uses the maxent function from the package 'dismo'. Make sure that you have correctly installed the maxent.jar file in the folder ~\R\library\version\dismo\java available at https://www.cs.princeton.edu/~schapire/maxent/ (see maxent for more details).

Artificial Neural Network (ANN)

Uses the nnet function from the package 'nnet', you can set the following parameters (see nnet for more details):

maxit integer. Maximum number of iterations, default 500.

Support vector machines (SVM)

Uses the svm function from the package 'e1071', you can set the following parameters (see svm for more details):

epsilon float. Epsilon parameter in the insensitive loss function, default 1e-08.
algo cv integer. If an integer value k>0 is specified, a k-fold cross-validation on the training data is performed to assess the quality of the model: the accuracy rate for classification and the Mean Squared Error for regression. By default, set to 3.

Warning

Depending on the raster object resolution the process can be more or less time and memory consuming.
References


See Also

`modelling` to build SDMs with a single algorithm, `stack_modelling` to build SSDMs.

Examples

```r
## Not run:
# Loading data
data(Env)
data(Occurrences)
Occurrences <- subset(Occurrences, Occurrences$SPECIES == 'elliptica')

# ensemble SDM building
ESDM <- ensemble_modelling(c('CTA', 'MARS'), Occurrences, Env, rep = 1,
                            Xcol = 'LONGITUDE', Ycol = 'LATITUDE',
                            ensemble.thresh = c(0.6))

# Results plotting
plot(ESDM)
## End(Not run)
```

### Env

**A stack of three environmental variables**

**Description**

A stack of three 30 arcsec-resolution rasters covering the north part of the main island of New Caledonia ‘Grande Terre’. CRAINFALL and TEMPERATURE rasters are climatic variables from the WorldClim database, and SUBSTRATE raster is from the IRD Atlas of New Caledonia (2012) (see reference below).

**Usage**

Env
evaluate

Format

A stack of three rasters:

- **RAINFALL**  Annual mean rainfall (mm)
- **TEMPERATURE**  Annual mean temperature (x10 degree Celsius)
- **SUBSTRATE**  Substrate type (categorical variable)

References


---

evaluate  Evaluate

description

Evaluation of SDM or ESDM habitat suitability predictions or evaluation of SSDM floristic composition with Pottier et al, 2013 method (see reference below)

Usage

evaluate(obj, ...)

### S4 method for signature 'Algorithm.SDM'

evaluate(obj, cv, cv.param, thresh = 1001,
metric = "SES", Env, ...)

### S4 method for signature 'MAXENT.SDM'

evaluate(obj, cv, cv.param, thresh = 1001,
metric = "SES", Env, ...)

### S4 method for signature 'Stacked.SDM'

evaluate(obj, ...)

Arguments

- **obj**  Stacked.SDM. SSDM to evaluate
- **...**  unused argument
- **cv**  character. Method of cross-validation used to evaluate the SDM (see details below).
- **cv.param**  numeric. Parameters associated to the method of cross-validation used to evaluate the SDM (see details below).
thresh numeric. A single integer value representing the number of equal interval threshold values between 0 and 1.

metric character. Metric(s) used to select the best SDMs that will be included in the ensemble SDM (see details below).

Env raster object. Stacked raster object of environmental variables (can be processed first by `load_var`).

Value

SDM/ESDM/SSDM evaluation in a data.frame

References


Examples

```r
## Not run:
# Loading data
data(Env)
data(Occurrences)
# SSDM building
SSDM <- stack_modelling(c('CTA', 'SVM'), Occurrences, Env, rep = 1,
                        Xcol = 'LONGITUDE', Ycol = 'LATITUDE',
                        Spcol = 'SPECIES')

# Evaluation
evaluate(SSDM)

## End(Not run)
```

gui SSDM package Graphic User Interface

description

User interface of the SSDM package.

Usage

```r
gui(port = getOption("shiny.port"), host = getOption("shiny.host",
               "127.0.0.1"), working.directory = getwd())
```
load.model

Arguments

- **port** char. The TCP port that the application should listen on (see `runApp` for more details).
- **host** char. The IPv4 address that the application should listen on (see `runApp` for more details).
- **working.directory** char. Directory in which the application will run.

Details

If your environmental variables have an important size, you should give enough memory to the interface with the (`maxmem` parameter). Note that only one instance of gui can be run at a time.

Value

Open a window with a shiny app to use the SSDM package with an user-friendly interface.

Examples

```r
## Not run:
gui()
## End(Not run)
```

---

load.model Load ensemble SDMs and SSDMs.

Description

Load S4 `Ensemble.SDM` and `Stacked.SDM` objects saved with their respective save function.

Usage

```r
load_esdm(name, path = getwd())
load_stack(name = "Stack", path = getwd(), GUI = FALSE)
```

Arguments

- **name** character. Name of the folder containing the model to be loaded.
- **path** character. Path to the directory containing the model to be loaded, by default the path to the current directory.
- **GUI** logical. Do not take this argument into account (parameter for the user interface).
load_occ

Load occurrence data

Description

Load occurrence data from CSV file to perform modelling, ensemble_modelling or stack_modelling.

Usage

load_occ(path = getwd(), Env, file = NULL, ..., Xcol = "Longitude",
Ycol = "Latitude", Spcol = NULL, GeoRes = TRUE,
reso = max(res(Env@layers[[1]])), verbose = TRUE, GUI = FALSE)

Arguments

path character. Path to the directory that contains the occurrence table.
Env raster stack. Environmental variables in the form of a raster stack used to perform spatial thinning (can be the result of the load_var function).
file character. File containing the occurrence table, if NULL (default) the .csv file located in the path will be loaded.
... additional parameters given to read.csv.
Xcol character. Name of the Latitude or X coordinate variable.
Ycol character. Name of the Longitude or Y coordinate variable.
Spcol character. Name of the column containing species names or IDs.
GeoRes logical. If TRUE, performs geographical thinning on occurrences to limit geographical biases in the SDMs.
reso numeric. Resolution used to perform the geographical thinning, default is the resolution of Env.
verbose logical. If TRUE, allows the function to print text in the console.
GUI logical. Parameter reserved for graphical interface.

Value

A data frame containing the occurrence dataset (spatially thinned or not).

See Also

load_var to load environmental variables.
load_var

Examples

```r
## Not run:
load_occ(path = system.file('extdata', package = 'SSDM'), Env,
         Xcol = 'LONGITUDE', Ycol = 'LATITUDE',
         file = 'Occurrences.csv', sep = ',')
## End(Not run)
```

load_var  Load environmental variables

Description

Function to load environmental variables in the form of rasters to perform modelling, ensemble_modelling or stack_modelling.

Usage

```r
load_var(path = getwd(), files = NULL, format = c(".grd", ".tif", ".asc", ".sdat", ".rst", ".nc", ".envi", ".bil", ".img"),
         categorical = NULL, Norm = TRUE, tmp = TRUE, verbose = TRUE,
         GUI = FALSE)
```

Arguments

- `path` character. Path to the directory that contains the environmental variables files.
- `files` character. Files containing the environmental variables If NULL (default) all files present in the path in the selected format will be loaded.
- `format` character. Format of environmental variables files (including .grd, .tif, .asc, .sdat, .rst, .nc, .tif, .envi, .bil, .img).
- `categorical` character. Specify whether an environmental variable is a categorical variable.
- `Norm` logical. If set to true, normalizes environmental variables between 0 and 1.
- `tmp` logical. If set to true, rasters are read in temporary file avoiding to overload the random access memory. But beware: if you close R, temporary files will be deleted.
- `verbose` logical. If set to true, allows the function to print text in the console.
- `GUI` logical. Do not take that argument into account (parameter for the user interface).

Value

A stack containing the environmental rasters (normalized or not).
mapDiversity

See Also

load_occ to load occurrences.

Examples

## Not run:
load_var(system.file('extdata', package = 'SSDM'))

## End(Not run)

mapDiversity Map Diversity

Description

Methods for Stacked.SDM or SSDM to map diversity and communities composition.

Usage

mapDiversity(obj, ...)

## S4 method for signature 'Stacked.SDM'
mapDiversity(obj, method, rep.B = 1000,
  verbose = TRUE, Env = NULL, ...)

Arguments

obj Stacked.SDM. SSDM to map diversity with.
...
other arguments pass to the method.
method character. Define the method used to create the local species richness map (see
details below).
rep.B integer. If the method used to create the local species richness is the random
Bernoulli (Bernoulli), rep.B parameter defines the number of repetitions used
to create binary maps for each species.
verbose logical. If set to true, allows the function to print text in the console.
Env raster object. Stacked raster object of environmental variables (can be processed
first by load_var). Needed only for stacking method using probability ranking
from richness (PRR).
Details

Methods: Choice of the method used to compute the local species richness map (see Calabrese et al. (2014) and D’Amen et al (2015) for more informations, see reference below):

- **pSSDM** sum probabilities of habitat suitability maps
- **Bernoulli** draw repeatedly from a Bernoulli distribution
- **bSSDM** sum the binary map obtained with the thresholding (depending on the metric of the ESDM).
- **MaximumLikelihood** adjust species richness of the model by linear regression
- **PRR.MEM** model richness with a macroecological model (MEM) and adjust each ESDM binary map by ranking habitat suitability and keeping as much as predicted richness of the MEM
- **PRR.pSSDM** model richness with a pSSDM and adjust each ESDM binary map by ranking habitat suitability and keeping as much as predicted richness of the pSSDM

Value

a list with a diversity map and eventually ESDMs for stacking method using probability ranking from richness (PPR).

References


See Also

- stacking to build SSDMs.

Examples

```
## Not run:
# Loading data
data(Env)
data(Occurrences)
# SSDM building
SSDM <- stack_modelling(c('CTA', 'SVM'), Occurrences, Env, rep = 1,
                        Xcol = 'LONGITUDE', Ycol = 'LATITUDE',
                        Spcol = 'SPECIES')

# Diversity mapping
mapDiversity(SSDM, method = 'pSSDM')
```
modelling

## End(Not run)

---

**Build an SDM using a single algorithm**

### Description
This is a function to build an SDM with one algorithm for a single species. The function takes as inputs an occurrence data frame made of presence/absence or presence-only records and a raster object for data extraction and projection. The function returns an S4 *Algorithm.SDM* class object containing the habitat suitability map, the binary map and the evaluation table.

### Usage

```r
modelling(algorithm, Occurrences, Env, Xcol = "Longitude", Ycol = "Latitude", Pcol = NULL, name = NULL, PA = NULL, cv = "holdout", cv.param = c(0.7, 2), thresh = 1001, metric = "SES", axes.metric = "Pearson", select = FALSE, select.metric = c("AUC"), select.thresh = c(0.75), verbose = TRUE, GUI = FALSE, ...)
```

### Arguments

- **algorithm** character. Choice of the algorithm to be run (see details below).
- **Occurrences** data frame. Occurrence table (can be processed first by `load_occ`).
- **Env** raster object. Raster object of environmental variable (can be processed first by `load_var`).
- **Xcol** character. Name of the column in the occurrence table containing Latitude or X coordinates.
- **Ycol** character. Name of the column in the occurrence table containing Longitude or Y coordinates.
- **Pcol** character. Name of the column in the occurrence table specifying whether a line is a presence or an absence. A value of 1 is presence and value of 0 is absence. If NULL presence-only dataset is assumed.
- **name** character. Optional name given to the final SDM produced (by default 'Algorithm.SDM').
- **PA** list(nb, strat) defining the pseudo-absence selection strategy used in case of presence-only dataset. If PA is NULL, recommended PA selection strategy is used depending on the algorithms (see details below).
- **cv** character. Method of cross-validation used to evaluate the SDM (see details below).
- **cv.param** numeric. Parameters associated to the method of cross-validation used to evaluate the SDM (see details below).
thresh numeric. A single integer value representing the number of equal interval threshold values between 0 and 1.

metric character. Metric used to compute the binary map threshold (see details below).

axes.metric Metric used to evaluate variable relative importance (see details below).

select logical. If set to true, models are evaluated before being projected, and not kept if they don’t meet selection criteria (see details below).

select.metric character. Metric(s) used to pre-select SDMs that reach a sufficient quality (see details below).

select.thresh numeric. Threshold(s) associated with the metric(s) used to compute the selection.

verbose logical. If set to true, allows the function to print text in the console.

GUI logical. Don’t take that argument into account (parameter for the user interface).

... additional parameters for the algorithm modelling function (see details below).

Details

algorithm 'all' allows to call directly all available algorithms. Currently, available algorithms include Generalized linear model (GLM), Generalized additive model (GAM), Multivariate adaptive regression splines (MARS), Generalized boosted regressions model (GBM), Classification tree analysis (CTA), Random forest (RF), Maximum entropy (MAXENT), Artificial neural network (ANN), and Support vector machines (SVM). Each algorithm has its own parameters settable with the ... (see each algorithm section below to set their parameters).

'PA' list with two values: nb number of pseudo-absences selected, and strat strategy used to select pseudo-absences: either random selection or disk selection. We set default recommendation from Barbet-Massin et al. (2012) (see reference).

cv Cross-validation method used to split the occurrence dataset used for evaluation: holdout data are partitioned into a training set and an evaluation set using a fraction (cv.param[1]) and the operation can be repeated (cv.param[2]) times, k-fold data are partitioned into k (cv.param[1]) folds being k-1 times in the training set and once the evaluation set and the operation can be repeated (cv.param[2]) times, LOO (Leave One Out) each point is successively taken as evaluation data.

metric Choice of the metric used to compute the binary map threshold and the confusion matrix (by default SES as recommended by Liu et al. (2005), see reference below): Kappa maximizes the Kappa, CCR maximizes the proportion of correctly predicted observations, TSS (True Skill Statistic) maximizes the sum of sensitivity and specificity, SES uses the sensitivity-specificity equality, LW uses the lowest occurrence prediction probability, ROC minimizes the distance between the ROC plot (receiving operating curve) and the upper left corner (1,1).

axes.metric Choice of the metric used to evaluate the variable relative importance (difference between a full model and one with each variable successively omitted): Pearson (computes a simple Pearson’s correlation $r$ between predictions of the full model and the one without a variable, and returns the score $1-r$: the highest the value, the more influence the variable has on the model), AUC, Kappa, sensitivity, specificity, and prop.correct (proportion of correctly predicted occurrences).

select.metric Selection metric(s) used to select SDMs: AUC, Kappa, sensitivity, specificity, and prop.correct (proportion of correctly predicted occurrences).

'...' See algorithm in detail section
Value

an S4 Algorithm.SDM Class object viewable with the `plot.model` method.

**Generalized linear model (GLM)**

Uses the `glm` function from the package ‘stats’, you can set the following parameters (see `glm` for more details):

- **test** character. Test used to evaluate the SDM, default ‘AIC’.
- **epsilon** numeric. Positive convergence tolerance eps ; the iterations converge when \(|\text{dev} - \text{dev}_\text{old}|/(\text{dev} + 0.1) < \text{eps}\). By default, set to 10e-08.
- **maxit** numeric. Integer giving the maximal number of IWLS (Iterative Weighted Last Squares) iterations, default 500.

**Generalized additive model (GAM)**

Uses the `gam` function from the package ‘mgcv’, you can set the following parameters (see `gam` for more details):

- **test** character. Test used to evaluate the model, default ‘AIC’.
- **epsilon** numeric. This is used for judging conversion of the GLM IRLS (Iteratively Reweighted Least Squares) loop, default 10e-08.
- **maxit** numeric. Maximum number of IRLS iterations to perform, default 500.

**Multivariate adaptive regression splines (MARS)**

Uses the `earth` function from the package ‘earth’, you can set the following parameters (see `earth` for more details):

- **degree** integer. Maximum degree of interaction (Friedman’s mi) ; 1 meaning build an additive model (i.e., no interaction terms). By default, set to 2.

**Generalized boosted regressions model (GBM)**

Uses the `gbm` function from the package ‘gbm,’ you can set the following parameters (see `gbm` for more details):

- **trees** integer. The total number of trees to fit. This is equivalent to the number of iterations and the number of basis functions in the additive expansion. By default, set to 2500.
- **final.leave** integer. minimum number of observations in the trees terminal nodes. Note that this is the actual number of observations, not the total weight. By default, set to 1.
- **algocv** integer. Number of cross-validations, default 3.
- **thresh.shrink** integer. Number of cross-validation folds to perform. If cv.folds>1 then gbm, in addition to the usual fit, will perform a cross-validation. By default, set to 1e-03.
modelling

Classification tree analysis (CTA)

Uses the \texttt{rpart} function from the package \texttt{rpart}, you can set the following parameters (see \texttt{rpart} for more details):

- \textbf{final.leave} integer. The minimum number of observations in any terminal node, default 1.
- \textbf{algocv} integer. Number of cross-validations, default 3.

Random Forest (RF)

Uses the \texttt{randomForest} function from the package \texttt{randomForest}, you can set the following parameters (see \texttt{randomForest} for more details):

- \textbf{trees} integer. Number of trees to grow. This should not be set to a too small number, to ensure that every input row gets predicted at least a few times. By default, set to 2500.
- \textbf{final.leave} integer. Minimum size of terminal nodes. Setting this number larger causes smaller trees to be grown (and thus take less time). By default, set to 1.

Maximum Entropy (MAXENT)

Uses the \texttt{maxent} function from the package \texttt{dismo}. Make sure that you have correctly installed the maxent.jar file in the folder \texttt{~R/library/version/dismo/java} available at \url{https://www.cs.princeton.edu/~schapire/maxent/} (see \texttt{maxent} for more details).

Artificial Neural Network (ANN)

Uses the \texttt{nnet} function from the package \texttt{nnet}, you can set the following parameters (see \texttt{nnet} for more details):

- \textbf{maxit} integer. Maximum number of iterations, default 500.

Support vector machines (SVM)

Uses the \texttt{svm} function from the package \texttt{e1071}, you can set the following parameters (see \texttt{svm} for more details):

- \textbf{epsilon} float. Epsilon parameter in the insensitive loss function, default 1e-08.
- \textbf{algocv} integer. If an integer value k>0 is specified, a k-fold cross-validation on the training data is performed to assess the quality of the model: the accuracy rate for classification and the Mean Squared Error for regression. By default, set to 3.

Warning

Depending on the raster object resolution the process can be more or less time and memory consuming.
Occurrences

References


See Also

ensemble_modelling to build ensemble SDMs, stack_modelling to build SSDMs.

Examples

# Loading data
data(Env)
data(Occurrences)
Occurrences <- subset(Occurrences, Occurrences$SPECIES == 'elliptica')

# SDM building
SDM <- modelling('GLM', Occurrences, Env, Xcol = 'LONGITUDE', Ycol = 'LATITUDE')

# Results plotting
## Not run:
plot(SDM)
## End(Not run)

Occurrences

Plant occurrences data frame

Description

A dataset containing plant occurrences of five Cryptocarya species native to New Caledonia. Occurrence data come from the Noumea Herbarium (NOU) and NC-PIPPN network (see Ibanez et al. (2014) in reference below).

Usage

Occurrences
**Format**

A data frame with 57 rows and 3 variables:

- **SPECIES**: Species of the occurrence
- **LONGITUDE**: Longitude of the occurrence
- **LATITUDE**: Latitude of the occurrence

**References**


---

**plot.model**

*Plot SDMs, ensemble SDMs, and SSDMs*

**Description**

Allows to plot S4 Algorithm.SDM, Ensemble.SDM and Stacked.SDM class objects.

**Usage**

```r
## S4 method for signature 'Stacked.SDM,ANY'
plot(x, y, ...)

## S4 method for signature 'SDM,ANY'
plot(x, y, ...)
```

**Arguments**

- `x`: Object to be plotted (S4 Algorithm.SDM, Ensemble.SDM or Stacked.SDM object).
- `y, ...`: Plot-based parameter not used.

**Value**

Open a window with a shiny app rendering all the results (habitat suitability map, binary map, evaluation table, variable importance and/or between-algorithm variance map, and/or algorithm evaluation, and/or algorithm correlation matrix and/or local species richness map) in a user-friendly interface.
Description

This is a collection of methods to project SDMs, ESDMs or SSDMs into the supplied environment. The function is used internally to calculate the input for the projection slot of .SDM classes but can also be used to project existing .SDM objects (see Details).

Usage

```r
project(obj, Env, ...)
## S4 method for signature 'Algorithm.SDM'
project(obj, Env, ...)
## S4 method for signature 'MAXENT.SDM'
project(obj, Env, ...)
## S4 method for signature 'Ensemble.SDM'
project(obj, Env, ...)
## S4 method for signature 'Stacked.SDM'
project(obj, Env, ...)
```

Arguments

- `obj` Object of class Algorithm.SDM, Ensemble.SDM or Stacked.SDM. Model(s) to be projected.
- `Env` Raster stack. Updated environmental rasters to be used for projection.
- `...` Additional arguments for internal use.

Details

The function uses any S4 .SDM class object and a raster stack of environmental layers of the variables the model was trained with.

Value

Either returns the original .SDM object with updated projection slots or if minimal.outputs = TRUE only returns the projections as Raster* objects. Depending on the object class this may be: a raster (Algorithm.SDM), a raster stack (Ensemble.SDM), a biodiversity map/mean raster (Stacked.SDM).
save.model

---

**Save ensemble SDMs and SSDMs**

**Description**

Allows to save S4 `Ensemble.SDM` and `Stacked.SDM` class objects.

**Usage**

```r
save.esdm(esdm, name = strsplit(esdm@name, ".", fixed = TRUE)[[1]][1],
           path = getwd(), verbose = TRUE, GUI = FALSE)
```

```r
# S4 method for signature 'Ensemble.SDM'
save.esdm(esdm, name = strsplit(esdm@name, ".Ensemble.SDM", fixed = TRUE)[[1]][1], path = getwd(),
           verbose = TRUE, GUI = FALSE)
```

```r
save.stack(stack, name = "Stack", path = getwd(), verbose = TRUE,
           GUI = FALSE)
```

```r
# S4 method for signature 'Stacked.SDM'
save.stack(stack, name = "Stack",
           path = getwd(), verbose = TRUE, GUI = FALSE)
```

**Arguments**

- `esdm` Ensemble.SDM. Ensemble SDM to be saved.
- `name` character. Folder name of the model to save.
- `path` character. Path to the directory chosen to save the SDM, by default the path to the current directory.
- `verbose` logical. If set to true, allows the function to print text in the console.
- `GUI` logical. Don’t take that argument into account (parameter for the user interface).
- `stack` Stacked.SDM. SSDM to be saved.

**Value**

Nothing in R environment. Creates folders, tables and rasters associated to the SDM. Tables are in .csv and rasters in .grd/.gri.

**See Also**

`load.model`
SSDM: Stacked species distribution modelling

Description

SSDM is a package to map species richness and endemism based on Stacked Species Distribution Models (SSDM). It provides tools to build SDM, i.e. a single species fitted with a single algorithm, Ensemble SDM (ESDM), i.e. a single species fitted with multiple algorithms, SSDM several species with one or more algorithms. The package includes numerous modelling algorithms, and specifiable ensemble aggregating and stacking methods. This package also provides tools to evaluate and explore models such as variable importance, algorithm accuracy, and between-algorithm correlation, and tools to map results such as habitat suitability maps, binary maps, between-algorithm variance maps. For ease of use, the SSDM package provides a user-friendly graphical interface (gui).

Details

The SSDM package provides five categories of functions (that you can find in details below): Data preparation, Modelling main functions, Model main methods, Model classes, and Miscellaneous.

Data preparation

  load_occ  Load occurrence data
  load_var  Load environmental variables

Modelling main functions

  modelling  Build an SDM using a single algorithm
  ensemble_modelling  Build an SDM that assembles multiple algorithms
  stack_modelling  Build an SSDMs that assembles multiple algorithms and species

Model main methods

  ensemble,Algorithm.SDM-method  Build an ensemble SDM
  stacking,Ensemble.SDM-method  Build an SSDM
  update,Stacked.SDM-method  Update a previous SSDM with new occurrence data

Model classes

  Algorithm.SDM  S4 class to represent SDMs
  Ensemble.SDM  S4 class to represent ensemble SDMs
  Stacked.SDM  S4 class to represent SSDMs
Stacked.SDM-class

Miscellaneous

- **gui** User-friendly interface for SSDM package
- **plot.model** Plot SDMs
- **save.model** Save SDMs
- **load.model** Load SDMs

---

**Stacked.SDM-class**  
**An S4 class to represent SSDMs**

---

**Description**

This is an S4 class to represent SSDMs that assembles multiple algorithms (including generalized linear model, general additive model, multivariate adaptive splines, generalized boosted regression model, classification tree analysis, random forest, maximum entropy, artificial neural network, and support vector machines) built for multiple species. It is obtained with `stack_modelling` or `stacking`.

**Slots**

- **name** character. Name of the SSDM (by default 'Species.SSDM').
- **diversity.map** raster. Local species richness map produced by the SSDM.
- **endemism.map** raster. Endemism map produced by the SSDM (see Crisp et al (2011) in references).
- **uncertainty** raster. Between-algorithm variance map.
- **evaluation** data frame. Evaluation of the SSDM (AUC, Kappa, omission rate, sensitivity, specificity, proportion of correctly predicted occurrences).
- **variable.importance** data frame. Relative importance of each variable in the SSDM.
- **algorithm.correlation** data frame. Between-algorithm correlation matrix.
- **esdms** list. List of ensemble SDMs used in the SSDM.
- **parameters** data frame. Parameters used to build the SSDM.
- **algorithm.evaluation** data frame. Evaluation of the algorithm averaging the metrics of all SDMs (AUC, Kappa, omission rate, sensitivity, specificity, proportion of correctly predicted occurrences).

**References**


**See Also**

- **Ensemble.SDM** an S4 class to represent ensemble SDMs, and **Algorithm.SDM** an S4 class to represent SDMs.
Stack different ensemble SDMs in an SSDM

Description

This is a function to stack several ensemble SDMs in an SSDM. The function takes as inputs several S4 Ensemble.SDM class objects produced with `ensemble_modelling` or `ensemble` functions. The function returns an S4 Stacked.SDM class object containing the local species richness map, the between-algorithm variance map, and all evaluation tables coming with (model evaluation, algorithm evaluation, algorithm correlation matrix and variable importance), and a list of ensemble SDMs for each species (see `ensemble_modelling`).

Usage

```
stacking(esdm, ..., name = NULL, method = "pSSDM", rep.B = 1000,
          Env = NULL, range = NULL, endemism = c("WEI", "Binary"),
          eval = TRUE, verbose = TRUE, GUI = FALSE)
```

## S4 method for signature 'Ensemble.SDM'

```
stacking(esdm, ..., name = NULL,
          method = "pSSDM", rep.B = 1000, Env = NULL, range = NULL,
          endemism = c("WEI", "Binary"), eval = TRUE, verbose = TRUE,
          GUI = FALSE)
```

Arguments

- `esdm, ...` character. Ensemble SDMs to be stacked.
- `name` character. Optional name given to the final SSDM produced (by default 'Species.SDM').
- `method` character. Define the method used to create the local species richness map (see details below).
- `rep.B` integer. If the method used to create the local species richness is the random bernoulli (`Bernoulli`), rep.B parameter defines the number of repetitions used to create binary maps for each species.
- `Env` raster object. Stacked raster object of environmental variables (can be processed first by `load_var`). Needed only for stacking method using probability ranking from richness (`PRR`).
- `range` integer. Set a value of range restriction (in pixels) around presences occurrences on habitat suitability maps (all further points will have a null probability, see Crisp et al (2011) in references). If NULL, no range restriction will be applied.
- `endemism` character. Define the method used to create an endemism map (see details below).
- `eval` logical. If set to true, disable stack evaluation.
- `verbose` logical. If set to true, allows the function to print text in the console.
- `GUI` logical. Don’t take that argument into account (parameter for the user interface).
Details

Methods: Choice of the method used to compute the local species richness map (see Calabrese et al. (2014) and D’Amen et al. (2015) for more informations, see reference below):

- **pSSDM** sum probabilities of habitat suitability maps
- **Bernoulli** draw repeatedly from a Bernoulli distribution
- **bSSDM** sum the binary map obtained with the thresholding (depending on the metric of the ESDM).
- **MaximumLikelihood** adjust species richness of the model by linear regression
- **PRR.MEM** model richness with a macroecological model (MEM) and adjust each ESDM binary map by ranking habitat suitability and keeping as much as predicted richness of the MEM
- **PRR.pSSDM** model richness with a pSSDM and adjust each ESDM binary map by ranking habitat suitability and keeping as much as predicted richness of the pSSDM

Endemism: Choice of the method used to compute the endemism map (see Crisp et al. (2001) for more information, see reference below):

- **NULL** No endemism map
- **WEI** (Weighted Endemism Index) Endemism map built by counting all species in each cell and weighting each by the inverse of its range
- **CWEI** (Corrected Weighted Endemism Index) Endemism map built by dividing the weighted endemism index by the total count of species in the cell.

First string of the character is the method either WEI or CWEI, and in those cases second string of the vector is used to precise range calculation, whether the total number of occurrences 'NbOcc' whether the surface of the binary map species distribution 'Binary'.

Value

an S4 Stacked.SDM class object viewable with the `plot.model` function.

References


See Also

stack_modelling to build SSDMs.

Examples

```r
## Not run:
# Loading data
data(Env)
data(Occurrences)
Occ1 <- subset(Occurrences, Occurrences$SPECIES == 'elliptica')
Occ2 <- subset(Occurrences, Occurrences$SPECIES == 'gracilis')

# SSDM building
ESDM1 <- ensemble_modelling(c('CTA', 'SVM'), Occ1, Env, rep = 1,
                           Xcol = 'LONGITUDE', Ycol = 'LATITUDE',
                           name = 'elliptica', ensemble.thresh = c(0.6))
ESDM2 <- ensemble_modelling(c('CTA', 'SVM'), Occ2, Env, rep = 1,
                           Xcol = 'LONGITUDE', Ycol = 'LATITUDE',
                           name = 'gracilis', ensemble.thresh = c(0.6))
SSDM <- stacking(ESDM1, ESDM2)

# Results plotting
plot(SSDM)
## End(Not run)
```

---

**stack_modelling**

Build an SSDM that assembles multiple algorithms and species.

**Description**

This is a function to build an SSDM that assembles multiple algorithm and species. The function takes as inputs an occurrence data frame made of presence/absence or presence-only records and a raster object for data extraction and projection. The function returns an S4 Stacked.SDM class object containing the local species richness map, the between-algorithm variance map, and all evaluation tables coming with (model evaluation, algorithm evaluation, algorithm correlation matrix and variable importance), and a list of ensemble SDMs for each species (see ensemble_modelling).

**Usage**

```r
stack_modelling(algorithms, Occurrences, Env, Xcol = "Longitude",
               Ycol = "Latitude", Pcol = NULL, Spcol = "SpeciesID", rep = 10,
               name = NULL, save = FALSE, path = getwd(), PA = NULL,
               cv = "holdout", cv.param = c(0.7, 1), thresh = 1001,
               axes.metric = "Pearson", uncertainty = TRUE, tmp = FALSE,
               ensemble.metric = c("AUC"), ensemble.thresh = c(0.75),
               weight = TRUE, method = "pSSDM", metric = "SES", rep.B = 1000,
               range = NULL, endemism = c("WEI", "Binary"), verbose = TRUE,
               GUI = FALSE, cores = 1, ...)
```
Arguments

- `algorithms` character. Choice of the algorithm(s) to be run (see details below).
- `Occurrences` data frame. Occurrence table (can be processed first by `load_occ`).
- `Env` raster object. Raster object of environmental variables (can be processed first by `load_var`).
- `Xcol` character. Name of the column in the occurrence table containing Latitude or X coordinates.
- `Ycol` character. Name of the column in the occurrence table containing Longitude or Y coordinates.
- `Pcol` character. Name of the column in the occurrence table specifying whether a line is a presence or an absence. A value of 1 is presence and value of 0 is absence. If NULL, presence-only dataset is assumed.
- `Spcol` character. Name of the column containing species names or IDs.
- `rep` integer. Number of repetitions for each algorithm.
- `name` character. Optional name given to the final Ensemble.SDM produced.
- `save` logical. If set to true, the SSDM is automatically saved.
- `path` character. If save is true, the path to the directory in which the ensemble SDM will be saved.
- `PA` list(nb, strat) defining the pseudo-absence selection strategy used in case of presence-only dataset. If PA is NULL, recommended PA selection strategy is used depending on the algorithm (see details below).
- `cv` character. Method of cross-validation used to evaluate the ensemble SDM (see details below).
- `cv.param` numeric. Parameters associated with the method of cross-validation used to evaluate the ensemble SDM (see details below).
- `thresh` numeric. A single integer value representing the number of equal interval thresholds values between 0 and 1.
- `axes.metric` Metric used to evaluate variable relative importance (see details below).
- `uncertainty` logical. If set to true, generates an uncertainty map and an algorithm correlation matrix.
- `tmp` logical. If set to true, the habitat suitability map of each algorithms is saved in a temporary file to release memory. But beware: if you close R, temporary files will be deleted. To avoid any loss you can save your SSDM with `save.model`. Depending on number, resolution and extent of models, temporary files can take a lot of disk space. Temporary files are written in R environment temporary folder.
- `ensemble.metric` character. Metric(s) used to select the best SDMs that will be included in the ensemble SDM (see details below).
- `ensemble.thresh` numeric. Threshold(s) associated with the metric(s) used to compute the selection.
weight logical. Choose whether or not you want the SDMs to be weighted using the selection metric or, alternatively, the mean of the selection metrics.

method character. Define the method used to create the local species richness map (see details below).

metric character. Metric used to compute the binary map threshold (see details below.)

rep.B integer. If the method used to create the local species richness is the random bernoulli (Bernoulli), rep.B parameter defines the number of repetitions used to create binary maps for each species.

range integer. Set a value of range restriction (in pixels) around presences occurrences on habitat suitability maps (all further points will have a null probability, see Crisp et al (2011) in references). If NULL, no range restriction will be applied.

derendism character. Define the method used to create an endemism map (see details below).

verbose logical. If set to true, allows the function to print text in the console.

GUI logical. Don’t take that argument into account (parameter for the user interface).

cores integer. Specify the number of CPU cores used to do the computing. You can use detectCores to automatically use all the available CPU cores.

... additional parameters for the algorithm modelling function (see details below).

Details

algorithms 'all' allows you to call directly all available algorithms. Currently, available algorithms include Generalized linear model (GLM), Generalized additive model (GAM), Multivariate adaptive regression splines (MARS), Generalized boosted regressions model (GBM), Classification tree analysis (CTA), Random forest (RF), Maximum entropy (MAXENT), Artificial neural network (ANN), and Support vector machines (SVM). Each algorithm has its own parameters settable with the ... (see each algorithm section below to set their parameters).

"PA" list with two values: nb number of pseudo-absences selected, and strat strategy used to select pseudo-absences: either random selection or disk selection. We set default recommendation from Barbet-Massin et al. (2012) (see reference).

cv Cross-validation method used to split the occurrence dataset used for evaluation: holdout data are partitioned into a training set and an evaluation set using a fraction (cv.param[1]) and the operation can be repeated (cv.param[2]) times, k-fold data are partitioned into k (cv.param[1]) folds being k-1 times in the training set and once the evaluation set and the operation can be repeated (cv.param[2]) times, LOO (Leave One Out) each point is successively taken as evaluation data.

metric Choice of the metric used to compute the binary map threshold and the confusion matrix (by default SES as recommended by Liu et al. (2005), see reference below): Kappa maximizes the Kappa, CCR maximizes the proportion of correctly predicted observations, TSS (True Skill Statistic) maximizes the sum of sensitivity and specificity, SES uses the sensitivity-specificity equality, LW uses the lowest occurrence prediction probability, ROC minimizes the distance between the ROC plot (receiving operating curve) and the upper left corner (1,1).

axes.metric Choice of the metric used to evaluate the variable relative importance (difference between a full model and one with each variable successively omitted): Pearson (computes a simple Pearson’s correlation r between predictions of the full model and the one without a
variable, and returns the score $1-r$: the highest the value, the more influence the variable has on the model). **AUC, Kappa, sensitivity, specificity**, and **prop.correct** (proportion of correctly predicted occurrences).

**ensemble.metric** Ensemble metric(s) used to select SDMs: **AUC, Kappa, sensitivity, specificity**, and **prop.correct** (proportion of correctly predicted occurrences).

**method** Choice of the method used to compute the local species richness map (see Calabrese et al. (2014) and D’Amen et al (2015) for more informations, see reference below): **pSSDM** sum probabilities of habitat suitability maps, **Bernoulli** drawing repeatedly from a Bernoulli distribution, **bSSDM** sum the binary map obtained with the thresholding (depending on the metric, see metric parameter), **MaximumLikelihood** adjust species richness using maximum likelihood parameter estimates on the logit-transformed occurrence probabilities (see Calabrese et al. (2014)). **PRR.MEM** model richness with a macroecological model (MEM) and adjust each ESDM binary map by ranking habitat suitability and keeping as much as predicted richness of the MEM, **PRR.pSSDM** model richness with a pSSDM and adjust each ESDM binary map by ranking habitat suitability and keeping as much as predicted richness of the pSSDM

**endemism** Choice of the method used to compute the endemism map (see Crisp et al. (2001) for more information, see reference below): **NULL** No endemism map, **WEI** (Weighted Endemism Index) Endemism map built by counting all species in each cell and weighting each by the inverse of its range, **CWEI** (Corrected Weighted Endemism Index) Endemism map built by dividing the weighted endemism index by the total count of species in the cell. First string of the character is the method either WEI or CWEI, and in those cases second string of the vector is used to precise range calculation, whether the total number of occurrences 'NbOCC' whether the surface of the binary map species distribution 'Binary'.

... See algorithm in detail section

**Value**

an S4 **Stacked.SDM** class object viewable with the **plot.model** function.

**Generalized linear model (GLM)**

Uses the **glm** function from the package 'stats', you can set the following parameters (see **glm** for more details):

**test** character. Test used to evaluate the SDM, default 'AIC'.

**epsilon** numeric. Positive convergence tolerance eps ; the iterations converge when $|\text{dev} - \text{dev}_\text{old}|/(|\text{dev}| + 0.1) < \text{eps}$. By default, set to 10e-08.

**maxit** numeric. Integer giving the maximal number of IWLS (Iterative Weighted Last Squares) iterations, default 500.

**Generalized additive model (GAM)**

Uses the **gam** function from the package 'mgcv', you can set the following parameters (see **gam** for more details):

**test** character. Test used to evaluate the model, default 'AIC'.

**epsilon** numeric. This is used for judging conversion of the GLM IRLS (Iteratively Reweighted Least Squares) loop, default 10e-08.

**maxit** numeric. Maximum number of IRLS iterations to perform, default 500.
Multivariate adaptive regression splines (MARS)

Uses the earth function from the package ‘earth’, you can set the following parameters (see earth for more details):

- **degree** integer. Maximum degree of interaction (Friedman’s mi); 1 meaning build an additive model (i.e., no interaction terms). By default, set to 2.

Generalized boosted regressions model (GBM)

Uses the gbm function from the package ‘gbm,’ you can set the following parameters (see gbm for more details):

- **trees** integer. The total number of trees to fit. This is equivalent to the number of iterations and the number of basis functions in the additive expansion. By default, set to 2500.
- **final.leave** integer. Minimum number of observations in the trees terminal nodes. Note that this is the actual number of observations not the total weight. By default, set to 1.
- **algocv** integer. Number of cross-validations, default 3.
- **thresh.shrink** integer. Number of cross-validation folds to perform. If cv.folds>1 then gbm, in addition to the usual fit, will perform a cross-validation. By default, set to 1e-03.

Classification tree analysis (CTA)

Uses the rpart function from the package ‘rpart’, you can set the following parameters (see rpart for more details):

- **final.leave** integer. The minimum number of observations in any terminal node, default 1.
- **algocv** integer. Number of cross-validations, default 3.

Random Forest (RF)

Uses the randomForest function from the package ‘randomForest’, you can set the following parameters (see randomForest for more details):

- **trees** integer. Number of trees to grow. This should not be set to a too small number, to ensure that every input row gets predicted at least a few times. By default, set to 2500.
- **final.leave** integer. Minimum size of terminal nodes. Setting this number larger causes smaller trees to be grown (and thus take less time). By default, set to 1.

Maximum Entropy (MAXENT)

Uses the maxent function from the package ‘dismo’. Make sure that you have correctly installed the maxent.jar file in the folder ~R/library/version/dismo/java available at https://www.cs.princeton.edu/~schapire/maxent/ (see maxent for more details).

Artificial Neural Network (ANN)

Uses the nnet function from the package ‘nnet’, you can set the following parameters (see nnet for more details):

- **maxit** integer. Maximum number of iterations, default 500.
**Support vector machines (SVM)**

Uses the `svm` function from the package `e1071`. You can set the following parameters (see `svm` for more details):

- **epsilon** float. Epsilon parameter in the insensitive loss function, default 1e-08.
- **algocv** integer. If an integer value k>0 is specified, a k-fold cross-validation on the training data is performed to assess the quality of the model: the accuracy rate for classification and the Mean Squared Error for regression. By default, set to 3.

**Warning**

Depending on the raster object resolution the process can be more or less time and memory consuming.

**References**


**See Also**

`modelling` to build simple SDMs.

**Examples**

```r
## Not run:
# Loading data
data(Env)
data(Occurrences)

# SSDM building
SSDM <- stack_modelling(c('CTA', 'SVM'), Occurrences, Env, rep = 1)
```
update.Stacked.SDM-method

Update a previous SSDM

Description
Update a previous SSDM with new occurrence data. The function takes as inputs updated or new occurrence data from one species, previous environmental variables, and an S4 Stacked.SDM class object containing a previously built SSDM.

Usage
## S4 method for signature 'Stacked.SDM'
update(object, Occurrences, Env, 
  Xcol = "Longitude", Ycol = "Latitude", Pcol = NULL,
  Spname = NULL, name = stack@name, save = FALSE, path = getwd(), 
  thresh = 1001, tmp = FALSE, verbose = TRUE, GUI = FALSE, ...)

Arguments
- **object**: Stacked.SDM. The previously built SSDM.
- **Occurrences**: data frame. New or updated occurrence table (can be processed first by `load_occ`).
- **Env**: raster object. Environment raster object (can be processed first by `load_var`).
- **Xcol**: character. Name of the column in the occurrence table containing Latitude or X coordinates.
- **Ycol**: character. Name of the column in the occurrence table containing Longitude or Y coordinates.
- **Pcol**: character. Name of the column in the occurrence table specifying whether a line is a presence or an absence. A value of 1 is presence and value of 0 is absence. If NULL presence-only dataset is assumed.
- **Spname**: character. Name of the new or updated species.
- **name**: character. Optional name given to the final SSDM produced, by default it’s the name of the previous SSDM.
- **save**: logical. If set to true, the model is automatically saved.
- **path**: character. Name of the path to the directory to contain the saved SSDM.
thresh numeric. A single integer value representing the number of equal interval threshold values between 0 and 1.

tmp logical. If set to true, the habitat suitability map of each algorithm is saved in a temporary file to release memory. But beware: if you close R, temporary files will be deleted To avoid any loss you can save your model with `save.model`.

verbose logical. If set to true, allows the function to print text in the console.

GUI logical. Don’t take that argument into account (parameter for the user interface).

... additional parameters for the algorithm modelling function (see details below).

Value

an S4 `Stacked.SDM` class object viewable with the `plot.model` function.

See Also

`stack_modelling` to build SSDMs.

Examples

```r
## Not run:
update(stack, Occurrences, Env, Spname = 'NewSpecie')

## End(Not run)
```
Index

*Topic datasets
   Env, 10
   Occurrences, 22
Algorithm.SDM, 3, 5, 18, 20, 23, 26, 27
Algorithm.SDM-class, 2
detectCores, 32
earth, 8, 20, 34
ensemble, 3, 5, 28
ensemble, Algorithm.SDM-method, 26
ensemble, Algorithm.SDM-method (ensemble), 3
Ensemble.SDM, 3–5, 8, 13, 23, 25–28
Ensemble.SDM-class, 5
ensemble_modelling, 4, 5, 5, 14, 15, 22, 26, 28, 30
Env, 10
evaluate, 11
evaluate, Algorithm.SDM-method (evaluate), 11
evaluate, MAXENT.SDM-method (evaluate), 11
evaluate, Stacked.SDM-method (evaluate), 11
gam, 8, 20, 33
gbm, 8, 20, 34
glm, 8, 20, 33
gui, 12, 26, 27
load.model, 13, 25, 27
load.esdm (load.model), 13
load_occ, 6, 14, 16, 18, 26, 31, 36
load_stack (load.model), 13
load_var, 6, 12, 14, 15, 16, 18, 26, 28, 31, 36
mapDiversity, 16
mapDiversity, Stacked.SDM-method (mapDiversity), 16
maxent, 9, 21, 34
modelling, 2, 3, 10, 14, 15, 18, 26, 35
nnet, 9, 21, 34
 Occurrences, 22
plot, SDM, ANY-method (plot.model), 23
plot, Stacked.SDM, ANY-method (plot.model), 23
plot.model, 4, 8, 20, 23, 27, 29, 33, 37
project, 24
project, Algorithm.SDM-method (project), 24
project, Ensemble.SDM-method (project), 24
project, MAXENT.SDM-method (project), 24
project, Stacked.SDM-method (project), 24
randomForest, 9, 21, 34
read.csv, 14
rpart, 9, 21, 34
runApp, 13
save.esdm (save.model), 25
save.esdm, Ensemble.SDM-method (save.model), 25
save.model, 7, 14, 25, 27, 31, 37
save.stack (save.model), 25
save.stack, Stacked.SDM-method (save.model), 25
SSDM, 26
SSDM-package (SSDM), 26
stack_modelling, 10, 14, 15, 22, 26, 27, 30, 30, 37
Stacked.SDM, 3, 5, 13, 23, 25, 26, 28–30, 33, 36, 37
Stacked.SDM-class, 27
stacking, 17, 27, 28
stacking, Ensemble.SDM-method, 26
INDEX

stacking, Ensemble.SDM-method (stacking), 28
sum, Algorithm.SDM-method (ensemble), 3
svm, 9, 21, 35

update, Stacked.SDM-method, 26, 36