

# Package ‘sdm’

October 17, 2016

**Type** Package

**Title** Species Distribution Modelling

**Version** 1.0-28

**Date** 2016-10-15

**Author** Babak Naimi, Miguel B. Araujo

**Maintainer** Babak Naimi <naimi.b@gmail.com>

**Depends** methods, sp (>= 1.0-13)

**Imports** raster

**Suggests** rgdal (>= 0.8-12), knitr, shinyBS, shiny, dismo

**Description** An extensible R framework for developing species distribution models using individual and community-based approaches, generate ensembles of models, evaluate the models, and predict species potential distributions in space and time.

**License** GPL (>= 3)

**URL** <http://biogeoinformatics.org>

**VignetteBuilder** knitr

**Repository** CRAN

**Repository/R-Forge/Project** sdm

**Repository/R-Forge/Revision** 53

**Repository/R-Forge/DateTimeStamp** 2016-10-16 16:47:09

**Date/Publication** 2016-10-17 00:46:24

**NeedsCompilation** no

## R topics documented:

add . . . . .	2
as.data.frame . . . . .	4
boxplot . . . . .	5
calibration . . . . .	6

coordinates . . . . .	7
density . . . . .	8
ensemble . . . . .	9
evaluates . . . . .	11
Extract by index . . . . .	13
featuresFrame-class . . . . .	14
get models' outputs . . . . .	15
getVarImp . . . . .	16
gui . . . . .	17
installAll . . . . .	18
names . . . . .	19
predict . . . . .	20
read.sdm . . . . .	22
roc . . . . .	23
sdm . . . . .	24
sdmCorrelativeMethod-class . . . . .	26
sdmData . . . . .	27
sdmdata-class . . . . .	31
sdmModels-classes . . . . .	32
sdmSetting . . . . .	33
subset . . . . .	35

## **Index** **37**

---

add	<i>add a new method to the package</i>
-----	--

---

### **Description**

This function is an interface to extend the package. A user can define a new method band add it to the package. When the method is successfully added, it can be used together with all existing methods. The names of available methods in the package can be seen using `getmethodNames`. It is not limited only to modelling (fitting) methods, but can be a replication method, or one to generate pseudo-absences, etc.

You can get an object of an existing method through `getmethod`.

### **Usage**

```
add(x,w,echo,...)
```

```
getmethod(x,w,...)
```

```
getmethodNames(w,...)
```

**Arguments**

x	Either a list, or an object generated by <code>getmethod</code> function
w	specify which group of methods the new method belongs to. "sdm" (default) can be used for modelling method
echo	logical (default=TRUE), determines whether a message should be printed to report if the adding is successful
...	additional arguments. see details

**Details**

These functions provide flexibility to extend the package by adding new methods. It is also possible to add several instances of an existing method to have, for example, a method with different settings at the same time. Whetevet the new method is, it can also be shared and used by other users.

**Value**

`getmethod` gives an object of an appropriate class depending on w.

`getmethodNames` generate a list (if `alt=TRUE` is provided as additional argument) containing the name of methods and all alternative names (aliases) specified for each method, or a character vector (if `alt=FALSE`) containing the main names.

**Author(s)**

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

**References**

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, 39:368-375, DOI: 10.1111/ecog.01881

**See Also**

#

**Examples**

```
## Not run:  
getmethodNames('sdm')  
  
## End(Not run)
```

---

`as.data.frame`*Get a data.frame with record id values (rID)*

---

## Description

Converts a `sdmdata` object to a `data.frame`. By additional arguments, it is possible to make a query on the dataset (see details).

## Usage

```
## S4 method for signature 'sdmdata'  
as.data.frame(x, ...)
```

## Arguments

<code>x</code>	<code>sdmdata</code> object
<code>...</code>	Additional arguments (optional, see details)

## Details

Following additional arguments optionally can be used to get a subset of data by specifying the record IDs; or using a query by specifying the name of species, and/or the name of data groups, and/or range of times (if time information are available):

`ind`: an integer vector containing the record IDs; `sp`: a character vector of the name of species `grp`: a character vector of the group names (e.g., 'test'; if independent test is available) `time`: a vector of times (an appropriate time class or a character that can be converted into a time format)

## Value

`data.frame`

## Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

## References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, 39:368-375, DOI: 10.1111/ecog.01881

## Examples

```
## Not run:
file <- system.file("external/data.sdd", package="sdm")

d <- read.sdm(file)

d # a sdmdata object

df <- as.data.frame(d)
head(df)

# only the records with rID == c(1,2,3):
as.data.frame(d, ind=1:3)

## End(Not run)
```

---

boxplot

*boxplot*

---

## Description

Make a box plot of model evaluation data, i.e., the model predictions for known presence and absence points.

## Details

Arguments:

x Object of class `sdmEvaluate` names Optional, the x-axis label for the group of data (e.g., 'Absence', 'Presence') ... Additional arguments that can be passed to [boxplot](#)

## Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

## References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

## Examples

```
## Not run:

e <- evaluates(x=c(1,1,0,1,0,0,0,1,1,1,0),
              p=c(0.69,0.04,0.05,0.95,0.04,0.65,0.09,0.61,0.75,0.84,0.15))
```

```

boxplot(e)

## End(Not run)

```

---

calibration

*Calibration*


---

### Description

evaluates for calibration

### Usage

```
calibration(x,p,nbin,weight,...)
```

### Arguments

x	a numeric vector including the observed values; or a <code>sdmEvaluate</code> object
p	a numeric vector including the predicted values
nbin	number of bins to discretize the predicted values into the specified bins (default: 10); instead, it can be the keyword of 'seek' to ask for seeking the best number
weight	logical, specifies whether a weight should be calculated based on the number of records at each bin. The weight will be used to summarize the calibration statistic
...	additional arguments (not implemented yet.)

### Details

The output of this function can be used in the `plot` function to generate Calibration plot. The calibration statistic is calculated using a method developed by the authors of this package (the journal article is not published yet, but in preparation)

### Value

an object of class `.sdmCalibration`

### Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

## References

Naimi, B., Niamir, A., Jimenez-Valverde, A., Araujo, M.B. (In preparation) Measuring calibration capacity of statistical models: a new statistic.

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

## Examples

```
## Not run:

ca <- calibration(x=c(1,1,0,1,0,0,0,1,1,1,0),
                 p=c(0.69,0.04,0.05,0.95,0.04,0.65,0.09,0.61,0.75,0.84,0.15))

ca

plot(ca)

## End(Not run)
```

---

coordinates	<i>get or set spatial coordinates of species data</i>
-------------	---

---

## Description

Get or set spatial coordinates of a `sdmdata` object.

## Usage

```
## S4 method for signature 'sdmdata'
coordinates(obj)

## S4 replacement method for signature 'sdmdata'
coordinates(object)<-value
```

## Arguments

<code>obj</code>	speciesData (either of <code>singleSpecies</code> , <code>multiple Species</code> or <code>SpeciesDataList</code> ) object
<code>object</code>	same as <code>obj</code>
<code>value</code>	spatial coordinates either a matrix, or data.frame, or column name as character, or a formula (e.g., <code>~x+y</code> , or <code>~coords(x+y)</code> )
<code>...</code>	Additional arguments

**Value**

matrix, or if the coordinates set, the sdmdata object is returned.

**Author(s)**

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

**References**

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

**Examples**

```
file <- system.file("external/data.sdd", package="sdm")
d <- read.sdm(file)

d # a sdmdata object

coordinates(d)
```

---

density

*density*

---

**Description**

Create a density plots of presence and absence data

**Value**

A density plot. Presence data are in darkblue, and absence data are in red.

**Methods**

```
density(x, ...)
```

x Object of class 'sdmEvaluate' (or a numeric vector of observed presence/absence)

**Author(s)**

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>



## References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

## Examples

```
## Not run:

e <- evaluates(x=c(1,1,0,1,0,0,0,1,1,1,0),
              p=c(0.69,0.04,0.05,0.95,0.04,0.65,0.09,0.61,0.75,0.84,0.15))

density(e)

## End(Not run)
```

---

ensemble	<i>Ensemble Forecasting of SDMs</i>
----------	-------------------------------------

---

## Description

Make a Raster object with a weighted averaging over all predictions from several fitted model in a sdmModel object.

## Usage

```
## S4 method for signature 'sdmModels'
ensemble(x, newdata, filename="", setting,...)
```

## Arguments

x	a sdmModels object
newdata	Raster* object or data.frame
filename	character, output file name
setting	list, contains the parameters that are used in the ensemble procedure; see details
...	additional arguments passed to the predict function

## Details

ensemble function uses the fitted models in an sdmModels object to generate an ensemble/consensus of predictions by individual models. Several methods do exist for this procedure, that are (or will be) implemented in this function, and can be defined in the method argument. A list can be introduced in the setting argument in which several parameters can be set including:

- method: specify which ensemble method should be used. Currently, 'unweighted' (unweighted averaging), and 'weighted' (weighted averaging) are implemented, but more methods will be added.

- stat: if the method='weighted' is used, this specify which evaluation statistics can be used as weight in the weighted averaging procedure. Alternatively, one may directly introduce weights (see the next argument)
- weights: an optional numeric vector (with a length equal to the models that are successfully fitted) to specify the weights for weighted averaging procedure (if the method='weighted' is specified)
- id: specify the model IDs that should be considered in the ensemble procedure. If missing, all the models that are successfully fitted are considered.
- wtest: specify which test dataset ("training","test.dep","test.indep") should be used to extract the statistic (stat) values as weights (if a relevant method is specified)

### Value

- a Raster object if predictors is a Raster object
- a numeric vector if predictors is a data.frame object

### Author(s)

Babak Naimi <naimi.b@gmail.com>  
<http://biogeoinformatics.org>

### References

#

### See Also

#

### Examples

```
## Not run:

file <- system.file("external/species.shp", package="sdm") # get the location of the species data
species <- shapefile(file) # read the shapefile

path <- system.file("external", package="sdm") # path to the folder contains the data

lst <- list.files(path=path,pattern='asc$',full.names = T) # list the name of the raster files

# stack is a function in the raster package, to read/create a multi-layers raster dataset
preds <- stack(lst) # making a raster object

d <- sdmData(formula=0ccurrence~., train=species, predictors=preds)

d

# fit the models (5 methods, and 10 replications using bootstrapping procedure):
```

```

m <- sdm(Occurrence~,data=d,methods=c('rf','tree','fda','mars','svm'),
        replicatin='boot',n=10)

# ensemble using weighted averaging based on AUC statistic:
p1 <- ensemble(m, newdata=preds, filename='ens.img',setting=list(method='weighted',stat='AUC'))
plot(p1)

# ensemble using weighted averaging based on TSS statistic
# and optimum threshold critesion 2 (i.e., Max(spe+sen)) :
p2 <- ensemble(m, newdata=preds, filename='ens2.img',setting=list(method='weighted',
        stat='TSS',opt=2))

plot(p2)

## End(Not run)

```

---

evaluates

*evaluate for accuracy*


---

### Description

evaluates for accuracy

### Usage

```
evaluates(x,p,...)
```

```
getEvaluation(x,w,wtest,stat,opt,...)
```

### Arguments

x	a numeric vector including the observed values; or a <code>sdmModels</code> object
p	a numeric vector including the predicted values
w	a numeric vector indicates model IDs
wtest	which test, training, dep.test, or indep.test?
stat	statistics that should be extracted from the <code>sdmEvaluate</code> object
opt	a numeric value indicates which threshold optimisation criteria should be considered if a threshold-based statistic is selected in stat
...	additional arguments (see details)

## Details

Evaluates the performance (accuracy) given the observed values, and the predicted values. As additional argument, the distribution of data can be specified (through `distribution`), that can be either of 'binomial', 'gaussian', 'laplace', or 'poisson'. If not specified, it will be guessed by the function!

`getEvaluation` can be used to get the evaluation results from a fitted model (`sdmModels` object that is output of the `sdm` function). Each model in `sdmModels` has a `modelID`, that can be specified in `w` argument. If `w` is not specified or more than a `modelID` is specified, then a `data.frame` is generated that contains the statistics specified in `stat`. For a single model (if length `w` is 1), `stat` can be 1 (threshold\_independent statistics), or 2 (threshold\_based statistics) or `NULL` (both groups). If more than a model is specified (`w` is either `NULL` or has a length greater than 1), `stat` can be the name of statistics such as 'AUC', 'COR', 'Deviance', 'obs.prevalence', 'threshold', 'sensitivity', 'specificity', 'TSS', 'm'. If either of the threshold\_based stats are selected, `opt` can be also specified to select one of the criteria for optimising the threshold. The possible value can be between 1 to 10 for "sp=se", "max(se+sp)", "min(cost)", "m" criteria, respectively.

## Value

an object of class `sdmEvaluate` from `evaluates` function

a list or `data.frame` from `getEvaluation` function

## Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

## References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

## See Also

#

## Examples

```
## Not run:
file <- system.file("external/model.sdm", package="sdm")

m <- read.sdm(file) # a sdmModels Object (fitted using sdm function)

getModelInfo(m)

# there are 4 models in the sdmModels objects

# so let's take a look at all the results for the model with modelID 1
```

```

# evaluation using training data (both threshod_independent and threshold_based groups):
getEvaluation(m,w=1,wtest='training')

getEvaluation(m,w=1,wtest='training',stat=1) # stat=1 (threshold_independent)

getEvaluation(m,w=1,wtest='test.dep',stat=2) # stat=2 (threshold_based)

getEvaluation(m,w=1:3,wtest='test.dep',stat=c('AUC','TSS'),opt=2)

getEvaluation(m,opt=1) # all models

getEvaluation(m,stat=c('TSS','Kappa','AUC'),opt=1) # all models

#####

example for evaluation:

evaluates(x=c(1,1,0,1,0,0,0,1,1,1,0),
          p=c(0.69,0.04,0.05,0.95,0.04,0.65,0.09,0.61,0.75,0.84,0.15))

## End(Not run)

```

---

 Extract by index

*Indexing to extract records of a sdmdata object*


---

## Description

This function extracts records of a sdmdata object and generates a new object of the same type (if drop=FALSE; otherwise a data.frame). In sdmdata, rID is the unique ID for each record.

## Methods

x[i]

Arguments

x a Raster\* object

i an index: record id (rID) in sdmdata object

drop If TRUE, a data.frame is returned, otherwise a sdmdata object is returned.

## Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

## References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, 39:368-375, DOI: 10.1111/ecog.01881

## Examples

```
## Not run:
file <- system.file("external/pa_df.csv", package="sdm")
df <- read.csv(file)
d <- sdmData(sp~b15+NDVI,train=df)

# see the number of records:
d

d2 <- d[1:10]

d2

d3 <- d[1:10,drop=TRUE]

d3

## End(Not run)
```

---

featuresFrame-class    *featureFrame class*

---

## Description

An S4 class contains the information of features used to fit a model

## Slots

`vars` A character vector containing the name of variables from the dataset used to generate the features

`feature.types` A list containing the definition of features

`response.specific` NULL, or a list containing the definition of features that their definitions are according to the response variable (i.e. species)

## Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

## References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

## See Also

#

## Examples

#

---

`get models' outputs`     *get info out of a sdmModels objects*

---

## Description

When SDMs are fitted using `sdm` function, a `sdmModels` object is generated that contains all the information and objects created through fitting and evaluation procedures for all species and methods. These functions facilitate getting these information or specific objects from the `sdmModels`.

## Usage

```
getModelInfo(x, ...)
```

## Arguments

<code>x</code>	a <code>sdmModel</code> object
<code>...</code>	additional arguments. see details

## Details

`getModelInfo` generates a `data.frame` containing the information on the whole information available in the model (i.e., `sdmModels` object), including the name of methods, what setting were used (i.e., replication) to evaluate, whether the fitting or different kinds of evaluations were successful. The first column in the `data.frame` contains the model IDs, the unique code for each model that is used everywhere in the package. That can be used for referring to a specific object in the other functions.

## Value

`data.frame`

**Author(s)**

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

**References**

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, 39:368-375, DOI: 10.1111/ecog.01881

**See Also**

#

**Examples**

```
## Not run:
file <- system.file("external/model.sdm", package="sdm")

m <- read.sdm(file)

getModelInfo(m)

## End(Not run)
```

---

getVarImp

*variable importance*

---

**Description**

Calculates relative importance of different variables in the models using several approaches.

**Usage**

```
getVarImp(x, id, wtest, ...)
```

**Arguments**

x	a numeric vector including the observed values; or a sdmEvaluate object
id	a numeric vector including the predicted values
wtest	specifies which dataset ('training', 'test.dep', 'test.indep') should be used (if exist) to calculate the importance of variables
...	additional arguments (not implemented yet.)



## Details

getVarImp function returns an object including different measures of variable importance, and if be put in plot function, a barplot is generated:

## Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

## References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

## Examples

```
## Not run:
# if m is a sdmModels object (output of sdm function) then:

getVarImp(m,id=1,wtest='training') # variable importance based on training dataset

vi <- getVarImp(m,id=1,wtest='test.dep')

plot(vi, 'auc')

plot(vi, 'cor')

## End(Not run)
```

---

gui

*Graphical User Interface*

---

## Description

Provides the possibility of using functions in the package through an interactive graphical user interface (GUI). Depending on input, different GUIs are opened.

## Usage

```
## S4 method for signature 'sdmModels'
gui(x,...)
```

**Arguments**

x                    a sdm\* object  
...                   not implemented yet.

**Details**

When x is missing, a GUI is opened to facilitate all the steps required to create sdmData, specify the settings for the different steps, and fit sdm models. Specifying x would be useful to interact with sdm\* object. For example, if x is a sdmModels (that is generated by sdm function), a user can interactively explore the results (e.g., to see different plots of model evaluation results).

**Value**

A HTML page in browser is opened.

**Author(s)**

Babak Naimi <naimi.b@gmail.com>  
<http://r-gis.net>  
<http://biogeoinformatics.org>

**Examples**

```
## Not run:  
file <- system.file("external/model.sdm", package="sdm")  
  
m <- read.sdm(file) # a sdmModels Object (fitted using sdm function)  
  
m  
  
gui(m)  
  
## End(Not run)
```

---

installAll

*install all packages that may be required by the package*

---

**Description**

This function facilitates to install the required packages that some functions are dependent on in the sdm package. It first checks whether the package is already installed, and if not, it installs the packages. If update=TRUE is used, the packages re-installed if they were already installed.

**Usage**

```
installAll(pkgs,update,...)
```

**Arguments**

pkgs	optional. the user provided list of packages (not required for the purpose of this function)
update	logical (default=FALSE), specifies whether the packages re-installed if they are already installed on the machine
...	Additional arguments passed to the <a href="#">install.packages</a> function

**Author(s)**

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

**References**

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

**See Also**

#

**Examples**

```
## Not run:
```

```
installAll()
```

```
## End(Not run)
```

---

names

*Names of species*

---

**Description**

Get or set the names of the species of a sdmdata object

**Usage**

```
## S4 method for signature 'sdmdata'
names(x)

## S4 replacement method for signature 'sdmdata'
names(x)<-value
```

**Arguments**

x	A sdm data object (sdmdata)
value	character (vector)

**Value**

For names, a character  
For names<-, the updated object.

**Author(s)**

Babak Naimi <naimi.b@gmail.com>  
<http://r-gis.net>  
<http://biogeoinformatics.org>

**References**

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, 39:368-375, DOI: 10.1111/ecog.01881

**Examples**

```
#
```

---

predict	<i>sdm model prediction</i>
---------	-----------------------------

---

**Description**

Make a Raster or matrix object (depending on input dataset) with predictions from one or several fitted models in sdmModels object.

**Usage**

```
## S4 method for signature 'sdmModels'
predict(object, newdata, filename="", w=NULL, species=NULL
        ,method=NULL,replication=NULL,run=NULL,mean=FALSE,control=NULL,
        overwrite=TRUE,nc=1,obj.size=1, ...)
```

**Arguments**

object	sdmModels object
newdata	Raster* object, or data.frame
filename	character, output file name, if missing, a name starts with sdm_prediction will be generated
w	numeric, specifies which model(s) should be used if the object contains several models; with NULL all models are used
species	character, (optional), specifies which species should be used if the object contains models for multiple species; with NULL all species are used
method	character, names of fitted models, e.g., glm, brt, etc.
replication	character, specifies the names of replication method, if NULL, all available replications are considered
run	numeric, works if replication with multiple runs are used
mean	logical, works if replication with multiple runs are used to fit the models, and specifies whether a mean should be calculated over all predictions of a replication method (e.g., bootstrapping) for each modelling method.
control	not implemented yet!
overwrite	logical, whether the filename should be overwritten if it does exist
nc	number of cores for parallel running of the function
obj.size	the size of object can be kept in memory (default=1 Giga byte). Depending on the available memory, this value can be changed
...	additional arguments, as for <a href="#">writeRaster</a>

**Details**

predict uses the fitted models in the sdmModel to generate the prediction given newdata. A raster (if newdata is Raster object) or data.frame (if newdata is data.frame) will be created.

The predictions can be generated for a specific set of models in the input sdmModels by determining either or a combination of the name of

For each prediction, a name is assigned which is kind of abbreviation or coding that tells which species, which method, which replication method, and which run is the prediction for. If the output is a Raster object, [setZ](#) function can be used to get a full name of each layer.

**Value**

a Raster object or data.frame

**Author(s)**

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

**References**

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, 39:368-375, DOI: 10.1111/ecog.01881

**See Also**

#

**Examples**

#

---

read.sdm	<i>read/write sdm* object from/to a file</i>
----------	--

---

**Description**

Read an sdm object from a file, or write it to a file.

**Usage**

```
read.sdm(filename, ...)
write.sdm(x, filename, overwrite, ...)
```

**Arguments**

filename	Filename (character)
x	a sdm object (e.g., sdmModels, sdmdata or sdmSetting)
overwrite	Logical. If TRUE, "filename" will be overwritten if it exists (default is FALSE)
...	additional arguments

**Details**

read.sdm function reads any files that has been written by write.sdm. These functions use [saveRDS](#) and [readRDS](#) to write and read the sdm objects. Additional arguments ... pass to these functions. An sdmModels object is saved to a file with an extension of ".sdm". The file extensions for sdmdata and sdmSetting object are ".sdd", and "sds", respectively.

**Value**

#

**Author(s)**

Babak Naimi  
 <naimi@r-gis.net>  
<http://r-gis.net>  
<http://biogeoinformatics.org>

**Examples**

```
#
```

---

```
roc                plot ROC curves
```

---

**Description**

Plot the Receiver Operating Characteristics (ROC) curve with AUC statistic in the legend.

**Usage**

```
roc(x,p=NULL,species=NULL,method=NULL,replication=NULL,run=NULL,
    wtest=NULL,smooth=FALSE,legend=TRUE,...)

getRoc(x,p,...)
```

**Arguments**

x	Either <code>sdmModels</code> , or <code>sdmEvaluate</code> object; or a numeric vector including observed binary values of species occurrence
p	if x is <code>sdmModels</code> , p is an optional vector with model ID number(s) that should be plotted (NULL (default means all models)); if x is a numeric vector, p is a vector with the same length including the predicted values
species	the name of species should be specified (required if x is <code>sdmModels</code> containing models for several species)
method	a character vector with the name of modelling methods that one need to get the roc plot for (if NULL [default], all methods in the object are considered); only if x is <code>sdmModels</code>
replication	a character vector with the name of replication methods (i.e., 'sub', 'cv', 'boot') that one need to get the roc plot for
run	if x is <code>sdmModels</code> , and the models are fitted through a replication procedure, run specifies which runs of the partitioning (replications) are required; if NULL, all are considered
wtest	evaluation for which test datasets are required, maximum 2 names from 'training', 'test.dep', 'test.indep' (i.e., evaluation for training data, dependent test dataset, and independent test dataset, respectively)

smooth	logical, specified whether the ROC curves should be smoothed through a spline procedure
legend	logical, specified whether a legend including AUC statistic is required on the plot
...	additional arguments passed to plot function

### Details

roc generates the plots of roc curves, and getRoc generate the values of ROC

### Value

an object of class matrix

### Author(s)

Babak Naimi <naimi.b@gmail.com>  
<http://r-gis.net>  
<http://biogeoinformatics.org>

### References

#

### See Also

#

### Examples

#

---

sdm

*Fit and evaluate species distribution models*

---

### Description

Fits sdm for single or multiple species using single or multiple methods specified by a user in methods argument, and evaluates their performance.

### Usage

sdm(formula, data, methods,...)



## Arguments

formula	Specifies the structure of the model, types of features, etc.
data	a sdmdata object created using <code>sdmData</code> function
methods	Character. Specifies the methods, used to fit the models
...	additional arguments

## Details

sdm fits multiple models and can be used to generate multiple runs (replicates) of each method through partitioning (using one or several partitioning methods including: subsampling, cross-validation, and bootstrapping).

Each model is evaluated against training data, and if available, splitted data (through partitioning; called dependent test data as well, i.e., "dep.test") and/or independent test data ("indep.test").

User should make sure that the methods are available and the required packages for them are installed before putting their names in the function, otherwise, the methods that cannot be run for any reason, are excluded by the function. It is a good practice to call `installAll` function (just one time when the sdm is installed), that tries to install all the packages that may be needed somewhere in the sdm package.

A new method can be adopted and added to the package by a user using `add` function. It is also possible to get an instance of an existing method, override the setting and definition, and then add it with a new name (e.g., `my.glm`).

The output would be a single object (`sdmModels`) that can be read/reproduced everywhere (e.g., on a new machine). A setting object can also be taken (exported) out of the output `sdmModels` object, that can be used to reproduce the same practice but given new conditions (i.e., new dataset, area, etc.)

## Value

an object of class `sdmModels`

## Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

## References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, 39:368-375, DOI: 10.1111/ecog.01881

## See Also

#

**Examples**

```

## Not run:
file <- system.file("external/pa_df.csv", package="sdm")

df <- read.csv(file)

head(df)

d <- sdmData(sp~b15+NDVI,train=df)

d
#----
# Example 1: fit using 3 models, and no evaluation (evaluation based on training dataset):

m <- sdm(sp~b15+NDVI,data=d,methods=c('glm','gam','gbm'))

m

# Example 3: fit using 5 models, and
# evaluates using 10 runs of subsampling replications taking 30 percent as test:

m <- sdm(sp~b15+NDVI,data=d,methods=c('glm','gam','gbm','svm','rf'),
         replication='sub',test.percent=30,n=10)

m

# Example 3: fits using 5 models, and
# evaluates using 10 runs of both 5-folds cross-validation and bootstrapping replication methods

m <- sdm(sp~.,data=d,methods=c('gbm','tree','mars','mda','fda'),
         replication=c('cv','boot'),cv.folds=5,n=10)

m

## End(Not run)

```

---

sdmCorrelativeMethod-class

*sdmCorrelativeMethod class*


---

**Description**

An S4 class representing sdm dataset

**Slots**

**name** Modelling method name  
**aliases** Alternative names for the method  
**dataArgument.names** A list keeps the name of data arguments in both fit and predict functions  
**packages** The required external package by the method  
**modelTypes** Specifies whether the model is presence-absence, presence-only, abundance, or multinomial  
**fitParams** a list of parameters needed by the method  
**fitSettings** a list of setting parameters for the method  
**settingRules** a function that adjust the setting parameters according to data  
**fitFunction** The main function use for fitting the model  
**tuneParams** a list of parameters to be tuned before the final fitting  
**predictParams** a list of parameters needed by predict function  
**predictSettings** a list of setting parameters for prediction  
**predictFunction** The main predict function  
**metadata** a metadata object containing the information about who creates the object, date, etc.  
**.temp.env** an environment object containing the functions defined by a user that is not from a package

---

sdmData

*creating sdm Data object*


---

**Description**

Creates a sdmdata objects that holds species (single or multiple) and explanatory variates. In addition, more information such as spatial coordinates, time, grouping variables, and metadata (e.g., author, date, reference, etc.) can be included.

**Usage**

```
sdmData(formula, train, test, predictors, bg, filename, crs, ...)
```

**Arguments**

<b>formula</b>	Specifies which species and explanatory variables should be taken from the input data. Other information (e.g., spatial coordinates, grouping variables, time, etc.) can be determined as well
<b>train</b>	Training data containing species observations as a data.frame or SpatialPoints or SpatialPointsDataFrames. It may contain predictor variables as well
<b>test</b>	Independent test data with the same structure as the train data

predictors	explanatory variables (predictors), defined as a raster object (RasterStack or RasterBrick). Required if train data only contain species records, or background records (pseudo-absences) should be generated
bg	Background data (pseudo-absence), as a data.frame. It can also be a list contains the settings to generate background data (a Raster object is required in the predictors argument)
filename	filename of the sdm data object to store in the disk
crs	optional, coordinate reference system
...	Additional arguments (optional) that are used to create a metadata object. See details

### Details

sdmData creates a data object, for single or multiple species. It can automatically detect the variables containing species data (if a data.frame is provided in train), but it is recommended to use formula through which all species (in the left hand side, e.g., sp1+sp2+sp3 ~ .), and the explanatory variables (in the right hand side) can be determined. If there are additional information such as spatial coordinates, time, or some variables based on which the observation can be grouped, they can be determined in the right hand side of the formula in a flexible way (e.g., ~ . + coords(x+y) + g(var); This right hand side formula, simply determines all variables (.) + x and y as spatial coordinates + grouping observations based on the variable var; for grouping, the variable (var in this example) should be categorical, i.e., factor ).

Additional arguments can be provided to determine metadata information including: author, website, citation, help, description, date, and license

### Value

an object of class sdmdata

### Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

### References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, 39:368-375, DOI: 10.1111/ecog.01881

### Examples

```
## Not run:
# Example 1: a data.frame containing records for a species (sp) and two predictors (b15 & NDVI):

file <- system.file("external/pa_df.csv", package="sdm")

df <- read.csv(file)
```

```

head(df)

d <- sdmData(sp~b15+NDVI,train=df)

d

# or simply:
d <- sdmData(sp~b15+NDVI,train=df)

d

#-----
# if formula is not specified, function tries to detect species and covariates, it works well only
# if dataset contains no additional columns but species and covariates!

d <- sdmData(train=df)

d

# # only right hand side of the formula is specified (one covariate), so function detects species:
d <- sdmData(~NDVI,train=df)

d

#-----
#####
# Example 2: a data.frame containing presence-absence records for 1 species, 4 covariates, and
# x, y coordinates:

file <- system.file("external/pa_df_with_xy.csv", package="sdm")

df <- read.csv(file)

head(df)

d <- sdmData(sp~b15+NDVI+categoric1+categoric2+coords(x+y),train=df)

d
#----
# categoric1 and categoric2 are categorical variables (factors), if not sure the data.frame has
# them as factor, it can be specified in the formula:
d <- sdmData(sp~b15+NDVI+f(categoric1)+f(categoric2)+coords(x+y),train=df)

d
# more simple forms of the formula:
d <- sdmData(sp~.+coords(x+y),train=df)

d

d <- sdmData(~.+coords(x+y),train=df) # function detects the species

d

```

```
#####
# Example 3: a data.frame containing presence-absence records for 10 species:

file <- system.file("external/multi_pa_df.csv", package="sdm")

df <- read.csv(file)

head(df)

# in the following formula, spatial coordinates columns are specified, and the rest is asked to
# be detected by the function:
d <- sdmData(~.+coords(x+y),train=df)

d

#--- or it can be customized with species and which covariates are needed:
d <- sdmData(sp1+sp2+sp3~b15+NDVI+f(categoric1) + coords(x+y),train=df)

d # 3 species, 3 covariates, and coordinates
# just be careful that if you put "." in the right hand side, while not all species columns or
# additional columns (e.g., coordinates, time) are specified in the formula, then it takes those
# columns as covariates which is NOT right!

#####
# Example 4: Spatial data:

file <- system.file("external/pa_spatial_points.shp", package="sdm") # path to a shapefile

# use a package like rgdal, or maptools, or shapefile function in package raster to read shapefile:
p <- shapefile(file)
class(p) # a "SpatialPointsDataFrame"

plot(p)

head(p) # it contains data for 3 species

# presence-absence plot for the first species (i.e., sp1)
plot(p[p@data$sp1 == 1,],col='blue',pch=16, main='Presence-Absence for sp1')

points(p[p@data$sp1 == 0,],col='red',pch=16)

# Let's read raster dataset containing predictor variables for this study area:

file <- system.file("external/predictors.grd", package="sdm") # path to a raster object

r <- brick(file)

r # a RasterBrick object including 2 rasters (covariates)

plot(r)

# now, we can use the species points and predictor rasters in sdmData function:
```

```

d <- sdmData(sp1+sp2+sp3~b15+NDVI,train=p,predictors = r)

d

#####
# Example 5: presence-only records:

file <- system.file("external/po_spatial_points.shp", package="sdm") # path to a shapefile

# use an appropriate function to read the shapefile (e.g., readOGR in rgdal, readShapeSpatial in
# maptools, or shapefile in raster):

po <- shapefile(file)
class(po) # a "SpatialPointsDataFrame"

head(po) # it contains data for one species (sp4) and the column has only presence records!

d <- sdmData(sp4~b15+NDVI,train=po,predictors = r)

d # as you see in the type, the data is Presence-Only

### we can add another argument (i.e., bg) to generate background (pseudo-absence) records:

#----- in bg, we are going to provide a list containing the setting to generate background
#----- the setting includes n (number of background records), method (the method used for
#----- background generation; gRandom refers to random in geographic space), and remove (whether
#----- points located in presence sites should be removed).

d <- sdmData(sp4~b15+NDVI,train=po,predictors = r,bg=list(n=1000,method='gRandom',remove=TRUE))

d # as you see in the type, the data is Presence-Background

# you can alternatively, put a data.frame including background records in bg!

## End(Not run)

```

---

sdmdata-class

*An S4 class representing sdm dataset*


---

## Description

An S4 class representing sdm dataset sdmdata

## Slots

species.names The names of species

species Contains the species data

`features.name` The names of predictor variables  
`features` A `data.frame` containing predictor variables  
`factors` The names of categorical variables (if any)  
`info` Other information such as coordinates, metadata, etc.  
`groups` A list including information on groups in the dataset  
`sdmFormula` An object of class `sdmFormula` containing the formula and its' terms defined by user  
`errorLog` Reports on errors in the data raised through data cleaning (e.g., NA, duplications, etc.)

### Author(s)

Babak Naimi <naimi.b@gmail.com>  
<http://biogeoinformatics.org>

### References

Naimi, B., Araujo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

### See Also

#

### Examples

#

---

sdmModels-classes      *sdmModels classes*

---

### Description

An S4 class to keep all the information of fitted models as well as their evaluations.

### Slots

Slots for `sdmModels` objects:

`data` a `sdmdata` object  
`recordIDs` Contains the species data  
`setting` A `data.frame` containing predictor variables  
`run.info` a `data.frame` containing info on runs  
`replicates` The names of categorical variables (if any)  
`models` a list contains all fitted objects and relevant information (e.g., evaluation)

**Slots for `sdmEvaluate` objects:**



observed a numeric vector of observed values  
 predicted a numeric vector of predicted values  
 statistics a list of threshold-independent statistics  
 threshold\_based a data.frame of threshold-based statistics

**Slots for sdmFormula objects:**

formula input formula  
 vars character, name of variables  
 model.terms the formula terms used in model fitting  
 data.terms the formula terms used to manipulate data

**Author(s)**

Babak Naimi  
 <naimi.b@gmail.com>  
<http://r-gis.net>  
<http://biogeoinformatics.org>

---

sdmSetting

*creating sdmSetting object*

---

**Description**

Creates sdmSetting object that holds settings to fit and evaluate the models. It can be used to reproduce a study.

**Usage**

```
sdmSetting(formula,data,methods,interaction.depth=1,n=1,replication=NULL,cv.folds=NULL,
  test.percent=NULL,bg=NULL,bg.n=NULL,var.importance=NULL,response.curve=TRUE,
  var.selection=FALSE,ncore=1L,...)
```

**Arguments**

formula	specify the structure of the model
data	sdm data object or data.frame including species and feature data
methods	character, name of the algorithms
interaction.depth	level of interactions between predictors
n	number of replicates (run)
replication	replication method (e.g., 'subsampling', 'bootstrapping', 'cv')

<code>cv.folds</code>	number of folds if cv (cross-validation) is in the selected replication methods
<code>test.percent</code>	test percentage if subsampling is in the selected replication methods
<code>bg</code>	method to generate background
<code>bg.n</code>	number of background records
<code>var.importance</code>	logical, whether variable importance should be calculated
<code>response.curve</code>	method to calculate variable importance
<code>var.selection</code>	logical, whether variable selection should be considered
<code>ncore</code>	number of cores to parallelize processing
<code>...</code>	additional arguments

### Details

using `sdmSetting`, the feature types, `interaction.depth` and all settings of the model can be defined.

### Value

an object of class `sdmSettings`

### Author(s)

Babak Naimi <naimi.b@gmail.com>

<http://r-gis.net>

<http://biogeoinformatics.org>

### References

#

### See Also

#

### Examples

#

---

subset	<i>Subset models in a sdmModels object</i>
--------	--

---

### Description

This function extracts a subset of models from a `sdmModels` object. It generates a new object of the same type as the original object. In `sdmModels`, `modelID` provides the unique IDs.

Instead of using the `subset` function, double brackets `'[[ ]]` can be used.

### Usage

```
x[[i,...]]  
  
## S4 method for signature 'sdmModels'  
subset(x, subset, drop=TRUE, ...)
```

### Arguments

<code>x</code>	<code>sdmModels</code> object
<code>i</code>	integer. Indicates the index/id of the models ( <code>modelID</code> ) should be extracted from <code>sdmModels</code> object
<code>subset</code>	Same as <code>i</code>
<code>drop</code>	If <code>TRUE</code> , new <code>modelIDs</code> are generated, otherwise, the original <code>modelIDs</code> are kept in the new object.
<code>...</code>	additional arguments (not implemented yet!)

### Value

`sdmModels` object

### Author(s)

Babak Naimi <naimi.b@gmail.com>  
<http://r-gis.net>  
<http://biogeoinformatics.org>

### References

Naimi, B., Araujo, M.B. (2016) `sdm`: a reproducible and extensible R platform for species distribution modelling, *Ecography*, DOI: 10.1111/ecog.01881

**Examples**

```
## Not run:
file <- system.file("external/model.sdm", package="sdm")

m <- read.sdm(file)

m

getModelInfo(m)

m1 <- m[[3:4]]

m1

getModelInfo(m1)

m2 <- m[[3:4,drop=FALSE]]

m2

getModelInfo(m2)

#---- the following is the same as previous:

m2 <- subset(m,3:4,drop=FALSE)

m2

getModelInfo(m2)

## End(Not run)
```

# Index

- \*Topic **classes**
  - sdmModels-classes, 32
- \*Topic **methods**
  - as.data.frame, 4
  - Extract by index, 13
- \*Topic **spatial,data**
  - boxplot, 5
  - density, 8
  - getVarImp, 16
- \*Topic **spatial,species,data**
  - add, 2
  - calibration, 6
  - ensemble, 9
  - evaluates, 11
  - featuresFrame-class, 14
  - get models' outputs, 15
  - predict, 20
  - roc, 23
  - sdmData, 27
  - sdmdata-class, 31
  - sdmSetting, 33
- \*Topic **spatial,species,sdm, model**
  - sdm, 24
- \*Topic **spatial**
  - as.data.frame, 4
  - coordinates, 7
  - Extract by index, 13
  - gui, 17
  - names, 19
  - sdmModels-classes, 32
  - subset, 35
- \*Topic **utilities, data**
  - installAll, 18
- \*Topic **write, write sdm**
  - read.sdm, 22
- [, sdmdata, ANY, ANY-method (Extract by index), 13
- [, sdmdata, missing, missing-method (Extract by index), 13
- [[ (subset), 35
- [[, sdmModels, ANY, ANY-method (subset), 35
- add, 2, 25
- add, list, character-method (add), 2
- as.data.frame, 4
- as.data.frame, sdmdata-method (as.data.frame), 4
- boxplot, 5, 5
- boxplot, sdmEvaluate-method (boxplot), 5
- calibration, 6
- calibration, sdmEvaluate, missing-method (calibration), 6
- calibration, vector, vector-method (calibration), 6
- characterORmissing-class (sdmModels-classes), 32
- characterORnull-class (sdmModels-classes), 32
- coordinates, 7
- coordinates, sdmdata-method (coordinates), 7
- coordinates, sdmModels-method (coordinates), 7
- coordinates<- (coordinates), 7
- coordinates<-, sdmdata-method (coordinates), 7
- CRSorNULL-class (sdmModels-classes), 32
- data.frameORnull-class (sdmModels-classes), 32
- density, 8
- density, sdmEvaluate-method (density), 8
- ensemble, 9
- ensemble, sdmModels, Raster-method (ensemble), 9
- ensemble, sdmModels-method (ensemble), 9

- environmentORnull-class  
(sdmModels-classes), 32
- evaluates, 11
- evaluates, vector, vector-method  
(evaluates), 11
- Extract by index, 13
- featuresFrame-class, 14
- formulaORnull-class  
(sdmModels-classes), 32
- functionORcharacter-class  
(sdmModels-classes), 32
- functionORnull-class  
(sdmModels-classes), 32
- get models' outputs, 15
- getEvaluation (evaluates), 11
- getEvaluation, sdmModels-method  
(evaluates), 11
- getmethod (add), 2
- getmethod, character-method (add), 2
- getmethodNames (add), 2
- getmethodNames, ANY-method (add), 2
- getModelInfo (get models' outputs), 15
- getModelInfo, sdmModels-method (get  
models' outputs), 15
- getRoc (roc), 23
- getRoc, sdmModels, ANY-method (roc), 23
- getRoc, vector, vector-method (roc), 23
- getVarImp, 16
- getVarImp, sdmModels-method (getVarImp),  
16
- gui, 17
- gui, sdmModels-method (gui), 17
- install.packages, 19
- installAll, 18, 25
- installAll, ANY-method (installAll), 18
- listORcharacter-class  
(sdmModels-classes), 32
- listORnull-class (sdmModels-classes), 32
- matrixORnull-class (sdmModels-classes),  
32
- names, 19
- names, sdmdata-method (names), 19
- names<- (names), 19
- names<- , sdmdata-method (names), 19
- numericORnull-class  
(sdmModels-classes), 32
- predict, 20
- predict, .maxlikeModel-method (predict),  
20
- predict, sdmModels-method (predict), 20
- read.sdm, 22
- read.sdm, character-method (read.sdm), 22
- readRDS, 22
- roc, 23
- roc, sdmModels, ANY-method (roc), 23
- roc, sdmModels-method (roc), 23
- roc, vector, vector-method (roc), 23
- saveRDS, 22
- sdm, 15, 24
- sdm, formula, sdmdata, character-method  
(sdm), 24
- sdmCorrelativeMethod-class, 26
- sdmData, 25, 27
- sdmData, ANY, data.frame, ANY, missing-method  
(sdmData), 27
- sdmData, ANY, data.frame, missing, missing-method  
(sdmData), 27
- sdmData, ANY, SpatialPoints, ANY, missing-method  
(sdmData), 27
- sdmData, ANY, SpatialPoints, ANY, Raster-method  
(sdmData), 27
- sdmData, data.frame, formula, ANY, missing-method  
(sdmData), 27
- sdmData, data.frame, missing, ANY, missing-method  
(sdmData), 27
- sdmdata-class, 31
- sdmEvaluate-class (sdmModels-classes),  
32
- sdmFormula-class (sdmModels-classes), 32
- sdmModels-class (sdmModels-classes), 32
- sdmModels-classes, 32
- sdmSetting, 33
- sdmSetting, ANY, sdmdata, character-method  
(sdmSetting), 33
- setZ, 21
- show, sdmdata-method (sdmdata-class), 31
- show, sdmModels-method  
(sdmModels-classes), 32
- show, sdmSetting-method (sdmSetting), 33
- subset, 35

subset, sdmModels-method (subset), [35](#)

write.sdm (read.sdm), [22](#)

write.sdm, .sdmCorSetting, character-method  
(read.sdm), [22](#)

write.sdm, sdmdata, character-method  
(read.sdm), [22](#)

write.sdm, sdmModels, character-method  
(read.sdm), [22](#)

writeRaster, [21](#)