Package 'chooseGCM'

November 21, 2024

```
Type Package
Title Selecting General Circulation Models for Species Distribution
Version 1.0.1
Maintainer Luíz Esser < luizesser@gmail.com>
Description Methods to help selecting General Circulation Models (GCMs)
      in the context of projecting models to future scenarios. It is provided
      clusterization algorithms, distance and correlation metrics, as well as
      a tailor-made algorithm to detect the optimum subset of GCMs that recreate
      the environment of all GCMs.
License MIT + file LICENSE
Imports checkmate, cluster, cowplot, factoextra, ggcorrplot, ggplot2,
      graphics, httr, methods, reshape2, stats, terra, usedist, utils
Encoding UTF-8
RoxygenNote 7.3.2
Suggests tictoc, testthat, sf, knitr, raster, rmarkdown, stars
Config/testthat/edition 3
URL https://luizesser.github.io/chooseGCM/,
      https://github.com/luizesser/chooseGCM
VignetteBuilder knitr
BugReports https://github.com/luizesser/chooseGCM/issues
Depends R (>= 2.10)
NeedsCompilation no
Author Dayani Baili [aut] (<a href="https://orcid.org/0000-0002-6954-9902">https://orcid.org/0000-0002-6954-9902</a>),
      Reginaldo Ré [aut] (<https://orcid.org/0000-0001-6452-3466>),
      Marcos Lima [aut] (<https://orcid.org/0000-0002-5901-0911>),
      Luíz Esser [aut, cre, cph] (<a href="https://orcid.org/0000-0003-2982-7223">https://orcid.org/0000-0003-2982-7223</a>)
Repository CRAN
Date/Publication 2024-11-21 17:20:11 UTC
```

2 closestdist_gcms

Contents

close	stdist_gcms Distance between General Circulation Models (GCMs)	
Index		19
	worldclim_data	17
	transform_gcms	
	summary_gcms	
	optk_gcms	
	montecarlo_gcms	
	kmeans_gcms	
	import_gcms	
	hclust_gcms	
	flatten_gcms	
	env_gcms	
	dist_gcms	
	cor_gcms	
	compare_gcms	
	closestdist_gcms	2

Description

This function compares future climate projections from multiple General Circulation Models (GCMs) based on their similarity in terms of bioclimatic variables. It computes distance metrics between GCMs and identifies subsets of GCMs that are similar to the global set.

Usage

```
closestdist_gcms(
    s,
    var_names = c("bio_1", "bio_12"),
    study_area = NULL,
    scale = TRUE,
    k = NULL,
    method = "euclidean",
    minimize_difference = TRUE,
    max_difference = NULL
)
```

Arguments

S	A list of stacks of General Circulation Models (GCMs).	
var_names	Character. A vector with names of the bioclimatic variables to compare, or 'all' to include all available variables.	
study_area	An Extent object, or any object from which an Extent object can be extracted. Defines the study area for cropping and masking the rasters.	

closestdist_gcms 3

scale Logical. Whether to apply centering and scaling to the data. Default is TRUE.

k Numeric. The number of GCMs to include in the subset. If NULL (default),

stopping criteria are applied.

method The distance method to use. Default is "euclidean". Possible values are: "eu-

clidean", "maximum", "manhattan", "canberra", "binary", "minkowski", "pear-

son", "spearman", or "kendall". See ?dist_gcms.

minimize_difference

Logical. If k = NULL, the function will search for the optimal value of k by adding GCMs to the subset until the mean distance starts to diverge from the

global mean distance. Default is TRUE.

max_difference Numeric. A distance threshold to stop searching for the optimal subset. If NULL,

no threshold is set. Default is NULL.

Details

The minimize_difference option searches for the best value of k by progressively adding GCMs to the subset. The function monitors the mean distance between the subset of GCMs and the global mean distance, stopping when the distance begins to increase. The max_difference option sets a maximum distance difference. If the mean distance between the subset GCMs exceeds this threshold, the function stops searching and returns the current subset.

Value

A set of GCMs that have a mean distance closer to the global mean distance of all GCMs provided in s.

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

See Also

```
cor_gcms dist_gcms
```

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
closestdist_gcms(s, var_names, study_area, method = "euclidean")
```

4 compare_gcms

compare_gcms

Compare General Circulation Models (GCMs)

Description

This function compares future climate projections from multiple General Circulation Models (GCMs) based on their similarity in terms of variables. The function uses three clustering algorithms — k-means, hierarchical clustering, and closestdist — to group GCMs, and generates visualizations for the resulting clusters.

Usage

```
compare_gcms(
    s,
    var_names = c("bio_1", "bio_12"),
    study_area = NULL,
    scale = TRUE,
    k = 3,
    clustering_method = "closestdist"
)
```

Arguments

s A list of stacks of General Circulation Models (GCMs).

var_names Character. A vector with the names of the variables to compare, or 'all' to

include all available variables.

study_area An Extent object, or any object from which an Extent object can be extracted.

Defines the study area for cropping and masking the rasters.

scale Logical. Whether to apply centering and scaling to the data. Default is TRUE.

k Numeric. The number of clusters to use for k-means clustering.

clustering_method

Character. The clustering method to use. One of: "kmeans", "hclust", or "closestdist". Default is "closestdist".

Value

A list with two items: suggested_gcms (the names of the GCMs suggested for further analysis) and statistics_gcms (a grid of plots visualizing the clustering results).

Author(s)

cor_gcms 5

Examples

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
compare_gcms(s, var_names, study_area, k = 3, clustering_method = "closestdist")
```

cor_gcms

Compute and Plot Correlation Matrix for a Set of General Circulation Models

Description

This function computes and visualizes the correlation matrix for a set of General Circulation Models (GCMs) based on their variables.

Usage

```
cor_gcms(
    s,
    var_names = c("bio_1", "bio_12"),
    study_area = NULL,
    scale = TRUE,
    method = "pearson"
)
```

Arguments

S	A list of stacks of General Circulation Models (GCMs).	
var_names	Character. A vector with names of the variables to compare, or 'all' to include all variables.	
study_area	An Extent object, or any object from which an Extent object can be extracted. Defines the study area for cropping and masking the rasters.	
scale	Logical. Whether to apply centering and scaling to the data. Default is TRUE.	
method	Character. The correlation method to use. Default is "pearson". Possible values are: "pearson", "kendall", or "spearman".	

Value

A list containing two items: cor_matrix (the calculated correlations between GCMs) and cor_plot (a plot visualizing the correlation matrix).

Author(s)

6 dist_gcms

See Also

transform_gcms flatten_gcms summary_gcms

Examples

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
cor_gcms(s, var_names, study_area, method = "pearson")
```

dist_gcms

Distance Between GCMs

Description

This function compares future climate projections from multiple General Circulation Models (GCMs) based on their similarity in terms of variables. It calculates distance metrics and plots the results on a heatmap.

Usage

```
dist_gcms(
    s,
    var_names = c("bio_1", "bio_12"),
    study_area = NULL,
    scale = TRUE,
    method = "euclidean"
)
```

Arguments

S	A list of stacks of General Circulation Models (GCMs).
var_names	Character. A vector of names of the variables to compare, or 'all' to include all variables.
study_area	An Extent object, or any object from which an Extent object can be extracted. Defines the study area for cropping and masking the rasters.
scale	Logical. Whether to apply centering and scaling to the data. Default is TRUE.
method	Character. The correlation method to use. Default is "euclidean". Possible values are: "euclidean", "maximum", "manhattan", "canberra", "binary", "minkowski", "pearson", "spearman", or "kendall".

Value

A list containing two items: distances (the calculated distances between GCMs) and heatmap (a plot displaying the heatmap).

env_gcms 7

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

See Also

```
transform_gcms flatten_gcms
```

Examples

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
dist_gcms(s, var_names, study_area, method = "euclidean")
```

env_gcms

General Circulation Model (GCM) Environmental Distribution

Description

This function visualizes GCM data in environmental space, with options to highlight clusters or specific GCMs.

Usage

```
env_gcms(
    s,
    var_names = c("bio_1", "bio_12"),
    study_area = NULL,
    highlight = "sum",
    resolution = 25,
    title = NULL
)
```

Arguments

S	A list of stacks of General Circulation Models (GCMs).
var_names	Character. A vector of names of the variables to include, or 'all' to include all variables.
study_area	An Extent object, or any object from which an Extent object can be extracted. Defines the study area for cropping and masking the rasters.
highlight	Character. A vector of GCM names to be highlighted. All other GCMs will appear in grey.
resolution	Numeric. The resolution to be used in the plot. Default is 25.
title	Character. The title of the plot.

8 flatten_gcms

Value

A plot displaying the environmental space for the specified GCMs.

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

See Also

```
summary_gcms
```

Examples

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
env_gcms(s, var_names, study_area, highlight = "sum")
env_gcms(s, var_names, study_area, highlight = c("ae", "ch", "cr"))
```

flatten_gcms

Flatten General Circulation Models (GCMs)

Description

Scale and flatten a list of GCMs data. frames.

Usage

```
flatten_gcms(s)
```

Arguments

S

A list of transformed data. frames representing GCMs.

Value

A data.frame with columns as GCMs and rows as values from each cell to each variable.

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

See Also

```
transform_gcms
```

hclust_gcms 9

Examples

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
s_trans <- transform_gcms(s, var_names, study_area)
flattened_gcms <- flatten_gcms(s_trans)</pre>
```

hclust_gcms

Hierarchical Clustering of GCMs

Description

This function performs hierarchical clustering on a random subset of raster values and produces a dendrogram visualization of the clusters.

Usage

```
hclust_gcms(
    s,
    var_names = c("bio_1", "bio_12"),
    study_area = NULL,
    scale = TRUE,
    k = 3,
    n = NULL
)
```

Arguments

S	A list of stacks of General Circulation Models (GCMs).
var_names	Character. A vector of names of the variables to include, or 'all' to include all variables.
study_area	An Extent object, or any object from which an Extent object can be extracted. Defines the study area for cropping and masking the rasters.
scale	Logical. Should the data be centered and scaled? Default is TRUE.
k	Integer. The number of clusters to identify.
n	Integer. The number of values to use in the clustering. If $NULL$ (default), all data is used.

Value

A dendrogram visualizing the clusters and the suggested GCMs.

Author(s)

import_gcms

See Also

```
transform_gcms flatten_gcms
```

Examples

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
hclust_gcms(s, var_names, study_area, k = 4, n = 500)
```

import_gcms

Import GCM Data to R

Description

This function imports GCM stack files from a folder into R.

Usage

```
import_gcms(
  path = "input_data/WorldClim_data_gcms",
  extension = ".tif",
  recursive = TRUE,
  gcm_names = NULL,
  var_names = NULL
)
```

Arguments

path	Character. A string specifying the path to the GCM files.
extension	Character. The file extension of the stack files. Default is ".tif", the standard extension for WorldClim 2.1 data.
recursive	Logical. Should the function import stacks recursively (i.e., search for files within subfolders)? Default is TRUE.
gcm_names	Character. A vector of names to assign to each GCM.
var_names	Character. A vector of names to assign to each variable.

Value

A list of stacks, where each element corresponds to a GCM from the specified path.

Author(s)

kmeans_gcms 11

See Also

```
worldclim_data
```

Examples

```
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = c("bio1", "bio12"))</pre>
```

kmeans_gcms

Perform K-Means Clustering on GCMs

Description

This function performs k-means clustering on a distance matrix and produces a scatter plot of the resulting clusters.

Usage

```
kmeans_gcms(
    s,
    var_names = c("bio_1", "bio_12"),
    study_area = NULL,
    scale = TRUE,
    k = 3,
    method = NULL
)
```

Arguments

S	A list of stacks of General Circulation Models (GCMs).	
var_names	Character. A vector of names of the variables to include, or 'all' to include all variables.	
study_area	An Extent object, or any object from which an Extent object can be extracted. Defines the study area for cropping and masking the rasters.	
scale	Logical. Should the data be centered and scaled? Default is TRUE.	
k	Integer. The number of clusters to create.	
method	Character. The method for distance matrix computation. Default is "euclidean." Possible values are: "euclidean," "maximum," "manhattan," "canberra," "binary," or "minkowski." If NULL, clustering will be performed on the raw variable data.	

Value

A scatter plot showing the resulting clusters and the suggested GCMs.

montecarlo_gcms

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

See Also

```
transform_gcms flatten_gcms
```

Examples

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
kmeans_gcms(s, var_names, study_area, k = 3)
```

montecarlo_gcms

Perform Monte Carlo Permutations on GCMs

Description

This function performs Monte Carlo permutations on a distance matrix and produces a violin plot showing the mean distance between subsets of the distance matrix.

Usage

```
montecarlo_gcms(
    s,
    var_names = c("bio_1", "bio_12"),
    study_area = NULL,
    scale = TRUE,
    perm = 10000,
    dist_method = "euclidean",
    clustering_method = "closestdist",
    ...
)
```

Arguments

S	A list of stacks of General Circulation Models (GCMs).	
var_names Character. A vector of names of the variables to include, or 'all' to include variables.		
study_area	An Extent object, or any object from which an Extent object can be extracted. Defines the study area for cropping and masking the rasters.	
scale	Logical. Should the data be centered and scaled? Default is TRUE.	
perm	Integer. The number of permutations to perform.	

optk_gcms 13

dist_method

Character. The method for distance matrix computation. Default is "euclidean." Possible values are: "euclidean," "maximum," "manhattan," "canberra," "binary," or "minkowski." If NULL, clustering will be performed on the raw variable data.

clustering_method

Character. The method for clustering. Default is "closestdist." Possible values are: "kmeans," "hclust," or "closestdist."

. . . Additional arguments to pass to the clustering function.

Value

A violin plot showing the results. The dashed red line and red dots represent the mean absolute distance between subsets of GCMs using the clustering approach. The violin plot is generated from Monte Carlo permutations, selecting random subsets of GCMs from the provided set.

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

See Also

hclust_gcms env_gcms kmeans_gcms

Examples

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
montecarlo_gcms(s, var_names, study_area)
```

optk_gcms

Optimize the number of clusters for a dataset

Description

This function performs clustering analysis on a dataset and determines the optimal number of clusters based on a specified method.

Usage

```
optk_gcms(
    s,
    var_names = c("bio_1", "bio_12"),
    study_area = NULL,
    cluster = "kmeans",
    method = "wss",
    n = NULL,
```

optk_gcms

```
nstart = 10,
K.max = 10,
B = 100
)
```

Arguments

S	A list of stacks of General Circulation Models.
var_names	Character. A vector with the names of the variables to compare OR 'all'.
study_area	Extent object, or any object from which an Extent object can be extracted. An object that defines the study area for cropping and masking the rasters.
cluster	A character string specifying the method to build the clusters. Options are 'kmeans' (default) or 'hclust'.
method	A character string specifying the method to use for determining the optimal number of clusters. Options are 'wss' for within-cluster sum of squares, 'silhouette' for average silhouette width, and 'gap_stat' for the gap statistic method. Default is 'wss'.
n	An integer specifying the number of randomly selected samples to use in the clustering analysis. If NULL (default), all data is used.
nstart	Numeric. The number of random sets to be chosen. Default is 10. Argument is passed to 'stats::kmeans()'.
K.max	Numeric. The maximum number of clusters to consider. Default is 10. Argument is passed to 'factoextra::fviz_nbclust()'.
В	Integer. The number of Monte Carlo ("bootstrap") samples. Default is 100. Argument is passed to 'cluster::clusGap()'.

Value

A ggplot object representing the optimal number of clusters.

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

See Also

```
transform_gcms flatten_gcms
```

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
optk_gcms(s, var_names, study_area)
```

summary_gcms 15

summary	/ gcms
Julilliai y	_50113

Summarize General Circulation Model (GCM) Data

Description

This function summarizes GCM data by calculating various statistics for each variable.

Usage

```
summary_gcms(s, var_names = c("bio_1", "bio_12"), study_area = NULL)
```

Arguments

s A list of stacks of Gener	ral Circulation Models (GCMs).
-----------------------------	--------------------------------

var_names Character. A vector of names of the variables to include, or 'all' to include all

variables.

study_area An Extent object, or any object from which an Extent object can be extracted.

Defines the study area for cropping and masking the rasters.

Value

A data frame containing the summary statistics for each variable.

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

See Also

```
transform_gcms
```

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
summary_gcms(s, var_names, study_area)
```

16 transform_gcms

transform_	gcms
------------	------

Transform General Circulation Model (GCM) Stacks

Description

This function transforms a list of GCM stacks by subsetting it to include only the variables specified in var_names, reprojecting it to match the CRS of study_area, cropping and masking it to study_area, and returning a list of data frames.

Usage

```
transform_gcms(s, var_names = c("bio_1", "bio_12"), study_area = NULL)
```

Arguments

S	A list of stacks of General Circulation Models (GCMs).	

var_names Character. A vector of names of the variables to include, or 'all' to include all

variables.

study_area An Extent object, or any object from which an Extent object can be extracted.

Defines the study area for cropping and masking the rasters.

Value

A list of data frames, where each element corresponds to a GCM in the input list.

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

See Also

```
summary_gcms
```

```
var_names <- c("bio_1", "bio_12")
s <- import_gcms(system.file("extdata", package = "chooseGCM"), var_names = var_names)
study_area <- terra::ext(c(-80, -30, -50, 10)) |> terra::vect(crs="epsg:4326")
transform_gcms(s, var_names, study_area)
```

worldclim_data 17

worldclim_data	Download WorldClim v2.1 Bioclimatic Data	
----------------	--	--

Description

This function allows downloading data from WorldClim v2.1 (https://www.worldclim.org/data/index.html) for multiple GCMs, time periods, and SSPs.

Usage

```
worldclim_data(period = 'current', variable = 'bioc', year = '2030',
gcm = 'mi', ssp = '126', resolution = 10, path = NULL)
```

Arguments

period	Character. Can be 'current' or 'future'.
variable	Character. Specifies which variables to retrieve. Possible entries are: 'tmax', 'tmin', 'prec', and/or 'bioc'.
year	Character or vector. Specifies the year(s) to retrieve data for. Possible entries are: '2030', '2050', '2070', and/or '2090'.
gcm	Character or vector. Specifies the $GCM(s)$ to consider for future scenarios. See the table below for available options:
	CODE **GCM**
	CM6-1-HR cr CNRM-ESM2-1 ec EC-Earth3-Veg ev EC-Earth3-Veg
	LR fi FIO-ESM-2-0 gf GFDL-ESM4 gg GISS-E2-1-G gh GISS- E2-1-H hg HadGEM3-GC31-LL in INM-CM4-8 ic INM-CM5-0
	ip IPSL-CM6A-LR me MIROC-ES2L mi MIROC6 mp MPI-ESM1-
	2-HR ml MPI-ESM1-2-LR mr MRI-ESM2-0 uk UKESM1-0-LL
ssp	Character or vector. SSP(s) for future data. Possible entries are: '126', '245', '370', and/or '585'.
resolution	Numeric. Specifies the resolution. Possible values are 10, 5, 2.5, or 30 arcseconds.
path	Character. Directory path to save the downloaded files. Default is NULL.

Details

This function creates a folder in path. All downloaded data will be stored in this folder. Note: While it is possible to retrieve a large volume of data, it is not recommended to do so due to the large file sizes. For example, datasets at 30 arcseconds resolution can exceed 4 GB. If the function fails to retrieve large datasets, consider increasing the timeout by setting options(timeout = 600).

18 worldclim_data

Value

This function does not return any value.

Author(s)

Luíz Fernando Esser (luizesser@gmail.com) https://luizfesser.wordpress.com

References

https://www.worldclim.org/data/index.html

```
# download data from multiple periods:
year <- c("2050", "2090")
worldclim_data("future", "bioc", year, "mi", "126", 10, path=tempdir())
# download data from one specific period:
worldclim_data("future", "bioc", "2070", "mi", "585", 10, path=tempdir())</pre>
```

Index

```
closestdist_gcms, 2
compare_gcms, 4
cor_gcms, 3, 5

dist_gcms, 3, 6
env_gcms, 7, 13

flatten_gcms, 6, 7, 8, 10, 12, 14

hclust_gcms, 9, 13
import_gcms, 10
kmeans_gcms, 11, 13
montecarlo_gcms, 12
optk_gcms, 13
summary_gcms, 6, 8, 15, 16
transform_gcms, 6–8, 10, 12, 14, 15, 16
worldclim_data, 11, 17
```