

Package: spThin (via r-universe)

August 20, 2024

Type Package

Title Functions for Spatial Thinning of Species Occurrence Records for Use in Ecological Models

Version 0.2.0

Date 2019-11-14

Description A set of functions that can be used to spatially thin species occurrence data. The resulting thinned data can be used in ecological modeling, such as ecological niche modeling.

BugReports <https://github.com/mlammens/spThin/issues>

Depends spam, grid, fields, knitr

Imports grDevices, graphics, utils

LazyData TRUE

License GPL-3

VignetteBuilder knitr

RoxygenNote 7.0.0

Repository <https://mlammens.r-universe.dev>

RemoteUrl <https://github.com/mlammens/spthin>

RemoteRef HEAD

RemoteSha 731bd8602e67cbc6a593d2d9cd14e4b298a5d28c

Contents

| | |
|--|---|
| Heteromys_anomalus_South_America | 2 |
| plotThin | 2 |
| summaryThin | 3 |
| thin | 3 |
| thin.algorithm | 5 |

| | |
|--------------|----------|
| Index | 6 |
|--------------|----------|

Heteromys_anomalus_South_America

Occurrence record locations for Heteromys anomalus

Description

A dataset containing compiled occurrence record locations for *Heteromys anomalus* in northern coastal South America. These records have been examined to check for accurate species identification.

Format

A data frame with 201 rows and 4 variables

Details

- SPEC. species name assigned to occurrence record
- LAT. decimal degree latitude value
- LONG. decimal degree longitude value
- REGION. region, or island, of occurrence

plotThin

Plot diagnosis for results of thin function

Description

Three plots (selected by `which`) are currently available: a plot of the number of repetitions versus the number of maximum records retained at each repetition ([1] observed values; [2] log transformed) and a histogram of the maximum records retained [3].

Usage

```
plotThin(
  thinned,
  which = c(1:3),
  ask = prod(par("mfcol")) < length(which) && dev.interactive(),
  ...
)
```

Arguments

| | |
|----------------------|--|
| <code>thinned</code> | A list of data.frames returned by <code>thin</code> function. |
| <code>which</code> | if a subset of the plots is required, specify a subset of the numbers 1:3. |
| <code>ask</code> | logical; if TRUE, the user is asked before each plot, see <code>par(ask=)</code> . |
| <code>...</code> | other parameters to be passed through to plotting functions. |

See Also[thin.algorithm](#)[thin](#)

summaryThin*Summary method for results of thin function*

Description

Summarize the results of thin function.

Usage

```
summaryThin(thinned, show = TRUE)
```

Arguments

thinned A list of data.frames returned by [thin](#) function.

show logical; if TRUE, the summary values are printed at the console.

Value

Returns a list with the (1) maximum number of records, (2) number of data frames with maximum number of records and (3) a table with the number of data frames per number of records.

See Also[thin.algorithm](#)[thin](#)

thin*Spatially thin species occurrence data*

Description

`thin` returns spatially thinned species occurrence data sets. A randomization algorithm ([thin.algorithm](#)) is used to create a data set in which all occurrence locations are at least `thin.par` distance apart. Spatial thinning helps to reduce the effect of uneven, or biased, species occurrence collections on spatial model outcomes.

Usage

```
thin(
  loc.data,
  lat.col = "LAT",
  long.col = "LONG",
  spec.col = "SPEC",
  thin.par,
  reps,
  locs.thinned.list.return = FALSE,
  write.files = TRUE,
  max.files = 5,
  out.dir,
  out.base = "thinned_data",
  write.log.file = TRUE,
  log.file = "spatial_thin_log.txt",
  verbose = TRUE
)
```

Arguments

| | |
|--------------------------|---|
| loc.data | A data.frame of occurrence locations. It can include several columns, but must include at minimum a column of latitude values, a column of longitude values, and a column of species names. |
| lat.col | Name of column of latitude values. Caps sensitive. |
| long.col | Name of column of longitude values. Caps sensitive. |
| spec.col | Name of column of species name. Caps sensitive. |
| thin.par | Thinning parameter - the distance (in kilometers) that you want records to be separated by. |
| reps | The number of times to repeat the thinning process. Given the random process of removing nearest-neighbors there should be 'rep' number of different sets of coordinates. |
| locs.thinned.list.return | TRUE/FALSE - If true, the 'list' of the data.frame of thinned locs resulting from each replication is returned (see Returns below). |
| write.files | TRUE/FALSE - If true, new *.csv files will be written with the thinned locs data |
| max.files | The maximum number of *.csv files to be written based on the thinned data |
| out.dir | Directory to write new *.csv files to |
| out.base | A file basename to give to the thinned datasets created |
| write.log.file | TRUE/FALSE create/append log file of thinning run |
| log.file | Text log file |
| verbose | TRUE/FALSE - If true, running details of the function are print at the console. |

Value

locs.thinned.dfs A list of data.frames, each data.frame the spatially thinned locations of the algorithm for a single replication. This list will have 'reps' elements.

See Also

[thin.algorithm](#)

| | |
|-----------------------------|---|
| <code>thin.algorithm</code> | <i>Implements random spatial thinning algorithm</i> |
|-----------------------------|---|

Description

`thin.algorithm` implements a randomization approach to spatially thinning species occurrence data. This function is the algorithm underlying the `thin` function.

Usage

```
thin.algorithm(rec.df.orig, thin.par, reps)
```

Arguments

- | | |
|--------------------------|---|
| <code>rec.df.orig</code> | A data frame of long/lat points for each presence record. The data.frame should be a two-column data frame, one column of long and one of lat |
| <code>thin.par</code> | Thinning parameter - the distance (in kilometers) that you want records to be separated by. |
| <code>reps</code> | The number of times to repete the thinning process. Given the random process of removing nearest-neighbors there should be 'rep' number of different sets of coordinates. |

Value

`reduced.rec.dfs`: A list object of length 'rep'. Each list element is a different data.frame of spatially thinned presence records.

Index

* datasets

Heteromys_anomalus_South_America,
[2](#)

Heteromys_anomalus_South_America, [2](#)

plotThin, [2](#)

summaryThin, [3](#)

thin, [2](#), [3](#), [3](#), [5](#)

thin.algorithm, [3](#), [5](#), [5](#)