Package: sampbias (via r-universe)

November 6, 2024

Type Package

Title Evaluating Geographic Sampling Bias in Biological Collections

Version 2.0.0

Description Evaluating the biasing impact of geographic features such as airports, cities, roads, rivers in datasets of coordinates based biological collection datasets, by Bayesian estimation of the parameters of a Poisson process. Enables also spatial visualization of sampling bias and includes a set of convenience functions for publication level plotting. Also available as shiny app.

Language en-gb

License GPL-3

URL https://github.com/azizka/sampbias

BugReports https://github.com/azizka/sampbias/issues

Depends R(>= 3.5.0)

Imports cowplot, dplyr, forcats, ggplot2, graphics, magrittr, methods, rlang, tidyr, viridis, terra, sf, rnaturalearth

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

LazyData TRUE

LazyDataCompression xz

RoxygenNote 7.2.3

VignetteBuilder knitr

Encoding UTF-8

Config/testthat/edition 3

Config/pak/sysreqs libgdal-dev gdal-bin libgeos-dev libicu-dev libssl-dev libproj-dev libsqlite3-dev libudunits2-dev

Repository https://azizka.r-universe.dev

RemoteUrl https://github.com/azizka/sampbias

RemoteRef HEAD

RemoteSha d4f7610c0578c42be79e2a75453be45c0e7aeeee

Contents

`area_example` *Example Dataset for a Custom Study Area*

Description

An example of the format needed to provide custom areas for `calculate_bias`

Usage

```
area_example
```

Format

An object of class `sf` (inherits from `data.frame`) with 2 rows and 1 columns.

Examples

```
data(area_example)
```

`borneo` *Borneo*

Description

The outline of Borneo, as example data for the user-defined study area option of `calculate_bias`. From `https://www.naturalearthdata.com`.

Usage

```
borneo
```

Format

An object of class sf (inherits from data.frame) with 1 rows and 3 columns.

Examples

```
data(borneo)
```

---

calculate_bias *Evaluating Sampling Bias in Species Distribution Data*

Description

The major function of the package, calculating the bias effect of sampling bias due to geographic structures, such as the vicinity to cities, airports, rivers and roads. Results are projected to space, and can be compared numerically.

Usage

```
calculate_bias(
  x,
  gaz = NULL,
  res = 1,
  buffer = NULL,
  restrict_sample = NULL,
  terrestrial = TRUE,
  inp_raster = NULL,
  mcmc_rescale_distances = 1000,
  mcmc_iterations = 1e+05,
  mcmc_burnin = 20000,
  mcmc_outfile = NULL,
  prior_q = c(1, 0.01),
  prior_w = c(1, 1),
  plot_raster = FALSE,
  verbose = FALSE,
  run_null_model = FALSE,
  use_hyperprior = TRUE
)
```

Arguments

x	an object of the class data.frame, with one species occurrence record per line, and at least three columns, named 'species', 'decimalLongitude', and 'decimalLatitude'.
gaz	a list of geographic gazetteers as SpatVector or sf. If NULL, a set of default gazetteers, representing large scale occurrence of airports, cities, rivers, and roads is used. See Details.

res	numerical. The raster resolution for the distance calculation to the geographic features and the data visualization, in decimal degrees. The default is to one degree, but higher resolution will be desirable for most analyses. res together with the extent of the input data determine computation time and memory requirements.
buffer	numerical. The size of the geographic buffer around the extent of ras for the distance calculations in degrees, to account for geographic structures neighbouring the study area (such as a road right outside the study area). Should be a multiple of res. Default is to res * 10. See Details.
restrict_sample	a SpatVector object. If provided the area for the bias test will be restricted to raster cells within these polygons (and the extent of the sampled points in x). Make sure to use adequate values for res. Default = NULL.
terrestrial	logical. If TRUE, the empirical distribution (and the output maps) are restricted to terrestrial areas. Uses the rnaturalearth:::ne_countries to define what is terrestrial. Default = TRUE.
inp_raster	an object of class SpatRaster. A template raster for the counts and distance calculation. Can be used to provide a special resolution, or for different coordinate reference systems. See vignette.
mcmc_rescale_distances	numerical. rescaling factor for the distance calculation
mcmc_iterations	numerical. the number of iterations for the MCMC, by default 100,000
mcmc_burnin	numerical. the burn-in for the MCMC, default is to 20,000
mcmc_outfile	character string. the path on where to write the results of the MCMC, optional.
prior_q	the gamma prior for the sampling rate q, which represents the expected number of occurrences per cell in the absence of biases. In the format c(shape,rate).
prior_w	the gamma prior for the steepness of the Poisson rate decline, such that w approximating 0 results in a null model of uniform sampling rate q across cells. In the format c(shape,rate).
plot_raster	logical. If TRUE, a plot of the occurrence raster is shown for diagnostic purposes. Default = FALSE
verbose	logical. If TRUE, progress is reported. Default = FALSE.
run_null_model	logical. Run a null model with bias weights set to zero.
use_hyperprior	logical. If TRUE a hyperprior on the bias weights is used for regularization to avoid over-parametrization.

Details

The default gazetteers delivered with the package are simplified from http://www.naturalearthdata.com/downloads/. They include major features, and for small scale analyses custom gazetteers should be used.

For computational convenience, the gazetteers are cropped to the extent of the point occurrence data sets. To account for the fact, that, relevant structures might lay directly outside this extent, but still influencing the distribution of samples in the study area, the buffer option, gives the area, around the extent that should be included in the distance calculation.

Visit https://github.com/azizka/sampbias/wiki for more information on distance calculation and the algorithm behind sampbias.

Value

An object of the S3-class 'sampbias', which is a list including the following objects:

summa	A list of summary statistics for the sampbias analyses, including the total number of occurrence points in x, the total number of species in x, the extent of the output rasters as well as the settings for res, binsize, and convexhull used in the analyses.
occurrences	a SpatRaster indicating occurrence records per grid cell, with resolution res.
species	a SpatRaster with indicating the number of species per grid cell, with resolution res.
biasmaps	a list of SpatRaster, with the same length as gaz. Each element is the spatial projection of the bias effect for a sources of bias in gaz. The last raster in the list is the average over all bias sources.
biastable	a data.frame, with the estimated bias effect for each bias source in gaz, at the distances specified by biasdist.

Note

Check https://github.com/azizka/sampbias/wiki for a tutorial on sampbias.

See Also

summary.sampbias is.sampbias plot.sampbias

Examples

```
## Not run:
  #simulate data
  x <- data.frame(species = rep(sample(x = LETTERS, size = 5), times = 20),
                  decimalLongitude = runif(n = 100, min = 0, max = 20),
                  decimalLatitude = runif(n = 100, min = -4, max = 4))

  out <- calculate_bias(x, terrestrial = TRUE, buffer = 0)
  summary(out)
  plot(out)



## End(Not run)
```

dis_rast *Distance Rasters from a List of Geographic Gazetteers*

Description

Creates a list of distances rasters based on a list of geographic gazetteers, as SpatVector objects, and a template SpatRaster, indicating the desired extent and resolution.

Usage

```
dis_rast(gaz, ras, buffer = NULL)
```

Arguments

gaz	an object of the class list, including one or more geographic gazetteers of the class SpatVector.
ras	an object of the class SpatRaster. Defining the extent and resolution of the distances rasters.
buffer	numerical. The size of the geographic buffer around the extent of ras for the distance calculations in degrees, to account for geographic structures neighbouring the study area (such as a road right outside the study area) Default is to the resolution of ras.

Value

a list of SpatRaster objects of the same length as gaz. The values in each raster correspond to the planar geographic distance to the next feature in gaz, given the resolution of ras

Note

Check https://github.com/azizka/sampbias/wiki for a tutorial on sampbias.

See Also

calculate_bias

Examples

```
#create raster for resolution and extent
ras <- terra::rast(terra::ext(-5,5,-4,4), res = 1)

#create point gazeteer
pts <- data.frame(long = runif(n = 5, min = -5, max = 5),
                  lat = runif(n = 5, min = -4, max = 4),
                  dat = rep("A", 5))

pts <- terra::vect(pts, geom = c("long", "lat"))
```

```
lin <- as.matrix(data.frame(long = seq(-5, 5, by = 1),
                 lat = rep(2, times = 11)))
lin <- terra::vect(lin, type = "line")

gaz <- list(point.structure = pts, lines.strucutre = lin)

out <- dis_rast(gaz, ras)

## Not run: plot(out[[1]])
```

---

ea_raster *Equal Area Raster*

Description

An example for an global equal area raster (in Behrmann projection) for the format needed for a custom grid provided to `calculate_bias`.

Usage

```
ea_raster
```

Format

An object of class `PackedSpatRaster` of length 1.

Examples

```
data(ea_raster)
ea_raster <- terra::unwrap(ea_raster)
```

---

ecoregion_example *Detailed Example for a Custom Study Area*

Description

An example of the format needed to provide custom areas for `calculate_bias` based on a publicly available set of global ecoregions.

Usage

```
ecoregion_example
```

Format

An object of class `sf` (inherits from `data.frame`) with 7 rows and 22 columns.

Source

https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world

Examples

```
data(ecoregion_example)
```

`is.sampbias` *Is Method for Class sampbias*

Description

Check class of `sampbias` objects.

Usage

```
## S3 method for class 'sampbias'
is(object, class2 = "sampbias")
```

Arguments

`object`	an object of the class `sampbias`
`class2`	the names of the class to which is relations are to be examined defined, or (more efficiently) the class definition objects for the classes.

Details

With two arguments, tests whether object can be treated as from class2. With one argument, returns all the super-classes of this object's class.

Examples

```
## Not run:
  #simulate data
  occ <- data.frame(species = rep(sample(x = LETTERS, size = 5), times = 10),
                    decimalLongitude = runif(n = 50, min = 12, max = 20),
                    decimalLatitude = runif(n = 50, min = -4, max = 4))

  out <- calculate_bias(x = occ, terrestrial = TRUE)
  is(out)

## End(Not run)
```

map_bias *Mapping Projected Bias Effects*

Description

A plotting function to visualize the effect of accessibility bias caused by different biasing factors in space.

Usage

```
map_bias(x, gaz = NULL, sealine = TRUE, type = "sampling_rate")
```

Arguments

x	a raster stack as generate by project_bias
gaz	a list of SpatialObjects, to be printed on the maps. Should be the same objects provided to calculate_bias when creating the Object. If gaz is not supplied, the sampbias package standard gazetteers are used.
sealine	logical. Should the coastline be added to the plots? Default is to TRUE.
type	character vector. One of c("sampling_rate", "log_sampling_rate", "diff_to_max"). If "sampling_rate". the plot shows the raw projected sampling rate depending on the biasing factors, if "log_sampling_rate", the plot shows the log10 transformed sampling rate, and if "diff_to_max", the relative deviation of sampling rate from the maximum rate as calculated using calculate_bias and projected using project_bias. For instance, a value of -25 indicates a drop of 25 (e.g. in a road on river flowing through the city airport).

Value

A series of R plots based on ggplot2.

See Also

calculate_bias, project_bias

Examples

```
## Not run:
  #simulate data
  occ <- data.frame(species = rep(sample(x = LETTERS, size = 5), times = 10),
                    decimalLongitude = runif(n = 50, min = 12, max = 20),
                    decimalLatitude = runif(n = 50, min = -4, max = 4))

  out <- calculate_bias(x = occ, terrestrial = TRUE)
  proj <- project_bias(out)
  map_bias(proj)

## End(Not run)
```

plot.sampbias *Plotting the Posterior Estimates of the Bias Weights*

Description

Plotting method for class sampbias, generating a box-whiskers-plot showing the bias weights for all biasing factors indicating the effect strength for each gazetteer provided to calculate_bias.

Usage

```
## S3 method for class 'sampbias'
plot(x, ...)
```

Arguments

x	an object of the class sampbias.
...	Additional arguments passed to summary.

Value

A plot

See Also

calculate_bias, summary.sampbias

Examples

```
## Not run:
  #simulate data
  occ <- data.frame(species = rep(sample(x = LETTERS, size = 5), times = 10),
                    decimalLongitude = runif(n = 50, min = 12, max = 20),
                    decimalLatitude = runif(n = 50, min = -4, max = 4))

  out <- calculate_bias(x = occ, terrestrial = TRUE)
  summary(out)
  plot(out)

## End(Not run)
```

project_bias *Projecting Bias Effects in Space*

Description

Uses the the estimated bias weights from a sampbias object to project the bias through space, using the same raster as used for the distance calculation.#'

Usage

```
project_bias(x, factors = NULL)
```

Arguments

x	an object of the class sampbias.
factors	a character vector indicating which biasing factors to project

Value

A raster stack, with the same length as the number of biasing factors used in calculate_bias. The names indicate the factors included for each layer.

See Also

calculate_bias, summary.sampbias

Examples

```
## Not run:
  #simulate data
  occ <- data.frame(species = rep(sample(x = LETTERS, size = 5), times = 10),
                    decimalLongitude = runif(n = 50, min = 12, max = 20),
                    decimalLatitude = runif(n = 50, min = -4, max = 4))

  out <- calculate_bias(x = occ, terrestrial = TRUE)
  proj <- project_bias(out)

## End(Not run)
```

summary.sampbias *Summary Method for Class sampbias*

Description

Summary method for objects of the class sampbias.

Usage

```
## S3 method for class 'sampbias'
summary(object, ...)
```

Arguments

object	An object of the class sampbias
...	Additional arguments passed to summary.

Value

Summary printed to screen.

See Also

calculate_bias is.sampbias plot.sampbias

Examples

```
## Not run:
  #simulate data
  occ <- data.frame(species = rep(sample(x = LETTERS, size = 5), times = 10),
                    decimalLongitude = runif(n = 50, min = 12, max = 20),
                    decimalLatitude = runif(n = 50, min = -4, max = 4))

  out <- calculate_bias(x = occ, terrestrial = TRUE)
  summary(out)

## End(Not run)
```

Index