Package 'biasProxyR'

November 18, 2026

Title Sampling Effort Proxies for Species Distribution Models

Version 0.1.0

Description Addresses geographic sampling bias in Species Distribution Models (SDMs) for data-poor regions. It generates sampling effort proxies using open-source infrastructure data, allowing researchers to sample bias-aware background points. This prevents models from learning sampling artifacts instead of true ecological niches.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 4.1.0)

VignetteBuilder knitr

Imports osmdata, rnaturalearth, sf, terra

Suggests knitr, rmarkdown, testthat (>= 3.0.0), curl, rnaturalearthdata

Config/roxygen2/version 8.1.0

NeedsCompilation no

Author Aamna Naveed [aut, cre]

Maintainer Aamna Naveed <naamna588@gmail.com>

Repository CRAN

Date/Publication 2026-11-18 10:00:21 UTC

Contents

diagnose_sampling_bias 2
generate_effort_surface 2
sample_bias_aware_background 3

Index **4**

diagnose_sampling_bias *Diagnose Sampling Bias in Occurrence Data*

Description

Plots occurrence points over a sampling effort proxy surface and calculates a "Bias Correlation Score" to quantify how much the data is skewed toward human infrastructure.

Usage

```
diagnose_sampling_bias(occurrences, effort_surface)
```

Arguments

occurrences	An sf object containing occurrence points.
effort_surface	A SpatRaster object representing sampling effort.

Value

A list containing the bias score (percentage) and a diagnostic message.

Examples

```
if (requireNamespace("rnaturalearthdata", quietly = TRUE) &&
    requireNamespace("osmdata", quietly = TRUE) &&
    curl::has_internet()) {
  try({
    effort <- generate_effort_surface("Luxembourg", resolution = 1000)
    bg_points <- sample_bias_aware_background(effort, n_points = 100)
    diagnose_sampling_bias(bg_points, effort)
  }, silent = TRUE)
}
```

generate_effort_surface *Generate a Sampling Effort Proxy Surface*

Description

Downloads road data for a given country and calculates distance to roads to create a proxy for human sampling effort.

Usage

```
generate_effort_surface(country = "Pakistan", resolution = 10000)
```

Arguments

country	Character. Name of the country (e.g., "Pakistan").
resolution	Numeric. Output raster resolution in meters (default = 10000).

Value

A SpatRaster object where higher values indicate higher probability of human sampling effort (closer to roads).

---

sample_bias_aware_background

Sample Bias-Aware Background Points

Description

Generates background points for Species Distribution Models (SDMs) that are weighted by a sampling effort proxy surface. This cancels out geographic sampling bias in data-poor regions.

Usage

```
sample_bias_aware_background(effort_surface, n_points = 10000)
```

Arguments

effort_surface	A SpatRaster object representing sampling effort (e.g., from generate_effort_surface).
n_points	Numeric. The number of background points to generate (default = 10000).

Value

An sf object containing the sampled background points.

Examples

```
if (requireNamespace("rnaturalearthdata", quietly = TRUE) &&
    requireNamespace("osmdata", quietly = TRUE) &&
    curl::has_internet()) {
  try({
    effort <- generate_effort_surface("Luxembourg", resolution = 1000)
    bg_points <- sample_bias_aware_background(effort, n_points = 1000)
  }, silent = TRUE)
}
```

Index

diagnose_sampling_bias, 2

generate_effort_surface, 2

sample_bias_aware_background, 3