Package ‘SSP’

May 7, 2026

Title Simulated Sampling Procedure for Community Ecology

Version 1.1.0

Date 2025-11-20

Maintainer Edlin Guerra-Castro <`edlinguerra@gmail.com`>

Description The Simulation-based Sampling Protocol (SSP) is an R package designed to estimate sampling effort in studies of ecological communities. It is based on the concept of pseudo-multivariate standard error (MultSE) (Anderson & Santana-Garcon, 2015, <doi:10.1111/ele.12385>) and the simulation of ecological data. The theoretical background is described in Guerra-Castro et al. (2020, <doi:10.1111/ecog.05284>).

Depends R (>= 3.5.0)

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

Suggests knitr, rmarkdown, testthat, roxygen2

VignetteBuilder knitr

URL `https://github.com/edlinguerra/SSP`

BugReports `https://github.com/edlinguerra/SSP/issues`

Imports vegan, stats, sampling, ggplot2

NeedsCompilation no

Author Edlin Guerra-Castro [aut, cre],
Maite Mascaro [aut],
Nuno Simoes [aut],
Juan Cruz-Motta [aut],
Juan Cajas [aut]

Repository CRAN

Date/Publication 2025-11-21 06:10:45 UTC

Contents

`SSP-package` *SSP: Simulated Sampling Procedure for Community Ecology*

Description

SSP is an R package designed to estimate sampling effort in studies of ecological communities based on the definition of pseudo multivariate standard error (MultSE) (Anderson & Santana-Garcon 2015) and simulation of data (Guerra-Castro et al., 2021).

Details

The protocol in SSP consists in simulating several extensive data matrices that mimic some of the relevant ecological features of the community of interest using a pilot data set. For each simulated data, several sampling efforts are repeatedly executed and MultSE is calculated to each one. The mean value, 0.025 and 0.975 quantiles of MultSE for each sampling effort across all simulated data are then estimated and plotted. The mean values are standardized in relation to the lowest sampling effort (consequently, the worst precision), and an optimal sampling effort can be identified as that in which the increase in sample size do not improve the precision beyond a threshold value (e.g. 2.5%).

SSP includes seven functions: `assempar` for extrapolation of assemblage parameters using pilot data; `simdata` for simulation of several data sets based on extrapolated parameters; `datquality` for evaluation of plausibility of simulated data; `sampsd` for repeated estimations of MultSE for different sampling designs in simulated data sets; `summary_ssp` for summarizing the behavior of MultSE for each sampling design across all simulated data sets, `ioptimum` for identification of the optimal sampling effort, and `plot_ssp` to plot sampling effort vs MultSE of simulated data.

The SSP package is developed at GitHub (`https://github.com/edlinguerra/SSP/`).

Author(s)

The SSP development team is Edlin Guerra-Castro, Maite Mascaro, Nuno Simoes, Juan Cruz-Motta and Juan Cajas

References

-Anderson, M.J., & Santana-Garcon, J. (2015). Measures of precision for dissimilarity-based multivariate analysis of ecological communities. Ecology Letters 18(1), 66-73. doi: doi:10.1111/ele.12385

-Guerra-Castro, E.J., Cajas, J.C., Simões, N., Cruz-Motta, J.J., & Mascaró, M. (2021). SSP: an R package to estimate sampling effort in studies of ecological communities. Ecography 44(4), 561-573. doi: doi:10.1111/ecog.05284

Examples

```
###To speed up the simulation of these examples, the cases, sites and N were set small.

##Single site: micromollusk from Cayo Nuevo (Yucatan, Mexico)
data(micromollusk)

#Estimation of parameters of pilot data
par.mic<-assempar (data = micromollusk,
                    type= "P/A",
                    Sest.method = "average")

#Simulation of 3 data sets, each one with 20 potential sampling units from a single site
sim.mic<-simdata(par.mic, cases= 3, N = 20, sites = 1)

#Sampling and estimation of MultSE for each sample size (few repetitions
#to speed up the example)

sam.mic<-sampsd(dat.sim = sim.mic,
               Par = par.mic,
               transformation = "P/A",
               method = "jaccard",
               n = 10,
               m = 1,
               k = 3)

#Summary of MultSE for each sampling effort
summ.mic<-summary_ssp(results = sam.mic, multi.site = FALSE)

#Cut-off points to identify optimal sampling effort
opt.mic<-ioptimum(xx = summ.mic, multi.site = FALSE)

#Plot
plot_ssp(xx = summ.mic, opt = opt.mic, multi.site = FALSE)

##Multiple sites: Sponges from Alacranes National Park (Yucatan, Mexico).
data(sponges)

#Estimation of parameters of pilot data
par.spo<-assempar(data = sponges,
                  type= "counts",
                  Sest.method = "average")
```

```
#Simulation of 3 data sets, each one with 10 potential sampling units in 3 sites.
sim.spo<-simdata(par.spo, cases= 3, N = 10, sites = 3)

#Sampling and estimation of MultSE for each sampling design (few repetitions
#to speed up the example)

sam.spo<-sampsd(dat.sim = sim.spo,
                Par = par.spo,
                transformation = "square root",
                method = "bray",
                n = 10,
                m = 3,
                k = 3)

#Summary of MultSE for each sampling effort
summ.spo<-summary_ssp(results = sam.spo, multi.site = TRUE)

#Cut-off points to identify optimal sampling effort
opt.spo<-ioptimum(xx = summ.spo, multi.site = TRUE)

#Plot
plot_ssp(xx = summ.spo, opt = opt.spo, multi.site = TRUE)
```

`assempar` *Estimation of Ecological Parameters of the Assemblage*

Description

This function extracts the main parameters of the pilot data using base R functions, as well as functions like `specpool` and `dispweight`.

Usage

```
assempar(data, type = c("P/A", "counts", "cover"), Sest.method = "average")
```

Arguments

`data`	Data frame with species names (columns) and samples (rows). The first column should indicate the site to which the sample belongs, regardless of whether a single site has been sampled.
`type`	Nature of the data to be processed. It may be presence/absence ("P/A"), counts of individuals ("counts"), or coverage ("cover").
`Sest.method`	Method for estimating species richness. The function `specpool` is used. Available methods are "chao", "jack1", "jack2", and "boot". By default, the "average" of the four estimates is used.

Details

The expected number of species in the assemblage is estimated using non-parametric methods (Gotelli et al. 2011). Due to variability in the estimates of each approximation (Reese et al. 2014), we recommend using the average. The probability of detection of each species is estimated among and within sites. Among-site detection is calculated as the frequency of occurrences of each species across sampled sites; within-site detection is calculated as the weighted average of frequencies in sites where the species are present. Spatial aggregation (only for count data) is evaluated using the index of dispersion D (Clarke et al. 2006). Properties of unseen species are approximated using information from observed species, assuming their detection probabilities match those of the rarest observed species. Abundance distributions are simulated using random Poisson values with lambda as the overall mean of observed abundances.

Value

A list (class `list`) containing the estimated parameters of the assemblage, to be used by `simdata`.

Note

Important: The first column should indicate the site ID of each sample (as character or numeric), even when only a single site was sampled.

References

Clarke, K. R., Chapman, M. G., Somerfield, P. J., & Needham, H. R. (2006). Dispersion-based weighting of species counts in assemblage analyses. Journal of Experimental Marine Biology and Ecology, 320, 11–27.

Gotelli, N. J., & Colwell, R. K. (2011). Estimating species richness. In A. E. Magurran & B. J. McGill (Eds.), Biological diversity: frontiers in measurement and assessment (pp. 39–54). Oxford University Press.

Guerra-Castro, E.J., Cajas, J.C., Simões, N., Cruz-Motta, J.J., & Mascaró, M. (2021). SSP: an R package to estimate sampling effort in studies of ecological communities. Ecography 44(4), 561-573. doi: doi:10.1111/ecog.05284

Reese, G. C., Wilson, K. R., & Flather, C. H. (2014). Performance of species richness estimators across assemblage types and survey parameters. Global Ecology and Biogeography, 23(5), 585–594.

See Also

`dispweight`, `specpool`, `simdata`

Examples

```
## Single site: micromollusk from Cayo Nuevo (Yucatan, Mexico)
data(micromollusk)
par.mic <- assempar(data = micromollusk, type = "P/A", Sest.method = "average")
par.mic

## Multiple sites: Sponges from Alacranes National Park (Yucatan, Mexico)
data(sponges)
```

```
par.spo <- assempar(data = sponges, type = "counts", Sest.method = "average")
par.spo
```

`datquality` *Diversity Metrics of Simulated and Original Data*

Description

Estimates the average number of species and the Simpson diversity index per sampling unit, as well as the total multivariate dispersion of both the original (pilot) and simulated datasets.

Usage

```
datquality(data, dat.sim, Par, transformation, method)
```

Arguments

`data`	Data frame with species as columns and samples as rows. The first column should indicate the site to which the sample belongs, regardless of whether a single site was sampled.
`dat.sim`	List of simulated data sets generated by `simdata`.
`Par`	List of parameters generated by `assempar`.
`transformation`	Mathematical transformation to reduce the weight of dominant species: one of "square root", "fourth root", "Log (X+1)", "P/A", or "none".
`method`	Dissimilarity metric used for multivariate dispersion, passed to `vegdist`.

Details

The quality of the simulated data sets is evaluated by statistical similarity to the pilot data. This includes: (i) the average number of species per sampling unit, (ii) the average Simpson diversity index, and (iii) the multivariate dispersion (MVD), defined as the average dissimilarity of each sampling unit to the group centroid in the dissimilarity space (Anderson 2006). For simulated datasets, mean and standard deviation are reported for (i) and (ii), and the 0.95 quantile of the MVD distribution is used to describe its variability.

Value

A data frame containing the mean and standard deviation of richness and diversity per sampling unit, and the MVD for original data, as well as the 0.95 quantile of MVD from the simulated data.

Note

It is desirable that simulated data resemble observed data in species richness and diversity per sampling unit.

References

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. Biometrics, 62, 245–253.

Guerra-Castro, E.J., Cajas, J.C., Simões, N., Cruz-Motta, J.J., & Mascaró, M. (2021). SSP: an R package to estimate sampling effort in studies of ecological communities. Ecography 44(4), 561-573. doi: doi:10.1111/ecog.05284

See Also

`vegdist`, `diversity`

Examples

```
## Single site: micromollusk from Cayo Nuevo (Yucatan, Mexico)
data(micromollusk)
par.mic <- assempar(data = micromollusk, type = "P/A", Sest.method = "average")
sim.mic <- simdata(par.mic, cases = 2, N = 10, sites = 1)
qua.mic <- datquality(data = micromollusk, dat.sim = sim.mic, Par = par.mic,
                      transformation = "none", method = "jaccard")
qua.mic

## See the full multi-site workflow in the vignette
```

`epibionts` *Epibionts on Caribbean mangrove roots*

Description

Data corresponds to epibenthic organisms in mangrove roots from Laguna de La Restinga National Park, Venezuela (Guerra-Castro et al. 2016).

Usage

```
data("epibionts")
```

Format

A data frame with 96 observations on the following 152 variables.

`sector` a factor with levels `E` `I` `M`

`site` a numeric vector

`Aaptos.sp` a numeric vector

`Acanthophora.spicifera` a numeric vector

`Acetabularia.crenulata` a numeric vector

`Aglaothamnion.sp` a numeric vector

`Amathia.sp` a numeric vector

`Amorphinopsis.atlantica` a numeric vector

`Amphimedon.erina` a numeric vector

`Anemonia.sargassensis` a numeric vector

`Aplidium.accarense` a numeric vector

`Aplysilla.glacialis` a numeric vector

`Ascidia.curvata` a numeric vector

`Ascidia.sp` a numeric vector

`Ascidia.sydneiensis` a numeric vector

`Balanus.sp` a numeric vector

`Bartholomea.annulata` a numeric vector

`Biemna.caribea` a numeric vector

`Bostrychia.tenella` a numeric vector

`Botrylloides.nigrum` a numeric vector

`Botrylloides.sp.1` a numeric vector

`Botrylloides.sp.2` a numeric vector

`Brachidontes.exustus` a numeric vector

`Branchiomma.conspersum` a numeric vector

`Branchiomma.nigromaculatum` a numeric vector

`Bryopsis.sp` a numeric vector

`Bugula.neritina` a numeric vector

`Bugula.sp` a numeric vector

`Calliactis.tricolor` a numeric vector

`Callyspongia..Callyspongia..pallida` a numeric vector

`Carijoa.riisei` a numeric vector

`Caulerpa.racemosa` a numeric vector

`Caulerpa.racemosa.var.peltata` a numeric vector

`Caulerpa.sertularioides` a numeric vector

`Caulerpa.verticillata` a numeric vector

`Caulibugula.sp` a numeric vector

`Celleporaria.sp` a numeric vector

`Ceramium.diaphanum` a numeric vector

`Chaetomorpha.sp.1` a numeric vector

`Chaetomorpha.sp.2` a numeric vector

`Chalinula.molitba` a numeric vector

`Chelonaplysilla.erecta` a numeric vector

`Chondrilla.nucula` a numeric vector

`Chthamalus.sp` a numeric vector

`Clathria..Clathria..microchela` a numeric vector

`Clathria.sp` a numeric vector

`Clavelina.oblonga` a numeric vector

`Clavelina.picta` a numeric vector

`Complejo.Cliona.celata` a numeric vector

`Crassostrea.rhizophorae` a numeric vector

`Dictyota.sp` a numeric vector

`Didemnum.cineraceum` a numeric vector

`Didemnum.perlucidum` a numeric vector

`Didemnum.sp` a numeric vector

`Diplosoma.listerianum` a numeric vector

`Distaplia.bermudensis` a numeric vector

`Distaplia.stylifera` a numeric vector

`Dynamena.sp` a numeric vector

`Dysidea.etheria` a numeric vector

`Dysidea.sp` a numeric vector

`Ecteinascidia.sp` a numeric vector

`Ecteinascidia.styeloides` a numeric vector

`Ecteinascidia.turbinata` a numeric vector

`Eudistoma.olivaceum` a numeric vector

`Eusynstyela.tincta` a numeric vector

`Exaiptasia.pallida` a numeric vector

`Ficopomatus.sp` a numeric vector

`Geodia.papyracea` a numeric vector

`Halichondria..Halichondria..magniconulosa` a numeric vector

`Halichondria..Halichondria..melanadocia` a numeric vector

`Haliclona..Halichoclona..magnifica` a numeric vector

`Haliclona..Reniera..implexiformis` a numeric vector

`Haliclona..Reniera..manglaris` a numeric vector

`Haliclona..Reniera..ruetzleri` a numeric vector

`Haliclona..Reniera..tubifera` a numeric vector

`Haliclona..Rhizoniera..curacaoensis` a numeric vector

`Haliclona..Soestella..caerulea` a numeric vector

`Haliclona..Soestella..smithae` a numeric vector

`Haliclona..Soestella..twincayensis` a numeric vector

`Halimeda.sp` a numeric vector

`Halisarca.sp` a numeric vector

`Halopteris.sp` a numeric vector

`Herdmania.pallida` a numeric vector

`Hippopodina.feegeensis` a numeric vector

`Hydroides.sp` a numeric vector

`Hyrtios.proteus` a numeric vector

`Iotrochota.birotulata` a numeric vector

`Ircinia.felix` a numeric vector

`Ircinia.sp` a numeric vector

`Isognomon.alatus` a numeric vector

`Kirchenpaueria.sp` a numeric vector

`Lissoclinum.sp` a numeric vector

`Lissodendoryx..Lissodendoryx..isodictyalis` a numeric vector

`Lithophyllum.pustulatum` a numeric vector

`Microcosmus.exasperatus` a numeric vector

`Molgula.occidentalis` a numeric vector

`Murrayella.periclados` a numeric vector

`Mycale..Aegogropila..carmigropila` a numeric vector

`Mycale..Aegogropila..citrina` a numeric vector

`Mycale..Carmia..magnirhaphidifera` a numeric vector

`Mycale..Carmia..microsigmatosa` a numeric vector

`Mycale..Mycale..laevis` a numeric vector

`Mycale..Zygomycale..angulosa` a numeric vector

`Mycale.sp` a numeric vector

`Nemalecium.sp` a numeric vector

`Notaulax.nudicollis` a numeric vector

`Obelia.sp` a numeric vector

`Oceanapia.nodosa` a numeric vector

`Padina.sp` a numeric vector

`Perna.viridis` a numeric vector

`Perophora.viridis` a numeric vector

`Phaeophyceae` a numeric vector

`Phallusia.nigra` a numeric vector

`Phyllangia.americana` a numeric vector

`Pinctada.imbricata` a numeric vector

`Plakortis.angulospiculatus` a numeric vector

`Polyclinum.constellatum` a numeric vector

`Polysiphonia.sp.1`	a numeric vector
`Polysiphonia.sp.3`	a numeric vector
`Polysiphonia.subtilissima`	a numeric vector
`Pteria.colymbus`	a numeric vector
`Pyura.sp..1`	a numeric vector
`Pyura.sp..2`	a numeric vector
`Pyura.vittata`	a numeric vector
`Rhizoclonium.sp`	a numeric vector
`Rhodosoma.turcicum`	a numeric vector
`Sabella.sp`	a numeric vector
`Sabellastarte.magnifica`	a numeric vector
`Schizoporella.pungens`	a numeric vector
`Scopalina.ruetzleri`	a numeric vector
`Scopalina.sp`	a numeric vector
`Scrupocellaria.sp`	a numeric vector
`Sphacelaria.rigidula`	a numeric vector
`Spongia..Spongia..pertusa`	a numeric vector
`Spongia..Spongia..tubulifera`	a numeric vector
`Sporolithon.episporum`	a numeric vector
`Spyridia.hypnoides`	a numeric vector
`Styela.canopus`	a numeric vector
`Styela.sp.1`	a numeric vector
`Styela.sp.2`	a numeric vector
`Suberites.aurantiacus`	a numeric vector
`Symplegma.brakenhielmi`	a numeric vector
`Symplegma.rubra`	a numeric vector
`Synnotum.circinatum`	a numeric vector
`Tedania..Tedania..ignis`	a numeric vector
`Terpios.manglaris`	a numeric vector
`Tethya.actinia`	a numeric vector
`Tethya.sp`	a numeric vector
`Trididemnum.orbiculatum`	a numeric vector
`Ulva.sp`	a numeric vector
`Viatrix.globulifera`	a numeric vector
`Zoobotryon.verticillatum`	a numeric vector

Details

Data consists of the coverage (by point-intercept) of 110 taxa identified in 240 mangrove roots, sampled under a hierarchically nested spatial design that included four random sites within each of three sectors of the lagoon system corresponding to a strong environmental gradient: external (E), intermediate (M), and internal (I). The abundance of epibenthic organisms of 8 roots were described within each site, producing a total of 32 roots in each sector. This spatial protocol was repeated five times over a period of 14 months. For demonstrative purpose, data from the 4th sampling period was randomly chosen as data for this package.

Source

https://doi.org/10.3354/meps11693

References

Guerra-Castro, E. J., J. E. Conde, and J. J. Cruz-Motta. (2016). Scales of spatial variation in tropical benthic assemblages and their ecological relevance: epibionts on Caribbean mangrove roots as a model system. Marine Ecology Progress Series 548:97-110.

Examples

```
data(epibionts)
str(epibionts)
```

`ioptimum` *Identification of the Optimal Sampling Effort*

Description

Estimates the sampling effort at which the improvement in precision (MultSE) per additional sampling unit becomes sub-optimal or redundant, based on predefined cut-off thresholds.

Usage

```
ioptimum(xx, multi.site = TRUE, c1 = 10, c2 = 5, c3 = 2.5)
```

Arguments

`xx`	A data frame generated by `summary_ssp`.
`multi.site`	Logical. Indicates whether multiple sites were simulated.
`c1`	First cut threshold. Default is 10% improvement over the highest MultSE.
`c2`	Second cut threshold. Default is 5% improvement over the highest MultSE.
`c3`	Third cut threshold. Default is 2.5% improvement over the highest MultSE.

Details

Sampling efforts between the minimum (e.g. 2 samples) and `c1` represent the necessary effort to achieve acceptable precision. Efforts between `c1` and `c2` reflect sub-optimal gains, and those between `c2` and `c3` are considered optimal. Beyond `c3`, any additional effort results in marginal improvements in MultSE and may be considered redundant. This classification helps support cost-benefit decisions in ecological survey design (see Underwood, 1990). If `c3` is not reached within the simulated range, the maximum available effort is returned with a warning.

Value

A vector or matrix indicating the sampling sizes corresponding to each cut-off point.

Note

The cut-off thresholds are arbitrary and should be adjusted based on the ecological question and resource availability. In some cases, `c3` may not be reached within the range of simulated sampling efforts.

References

Underwood, A. J. (1990). Experiments in ecology and management: Their logics, functions and interpretations. Australian Journal of Ecology, 15, 365–389.

Guerra-Castro, E.J., Cajas, J.C., Simões, N., Cruz-Motta, J.J., & Mascaró, M. (2021). SSP: an R package to estimate sampling effort in studies of ecological communities. Ecography 44(4), 561-573. doi: doi:10.1111/ecog.05284

See Also

`plot_ssp`, `summary_ssp`, `sampsd`

Examples

```
## Single site example
data(micromollusk)
par.mic <- assempar(data = micromollusk, type = "P/A", Sest.method = "average")
sim.mic <- simdata(par.mic, cases = 2, N = 20, sites = 1)
sam.mic <- sampsd(dat.sim = sim.mic,
                  Par = par.mic,
                  transformation = "P/A",
                  method = "jaccard",
                  n = 10,
                  m = 1,
                  k = 3)
summ.mic <- summary_ssp(results = sam.mic, multi.site = FALSE)
opt.mic <- ioptimum(xx = summ.mic, multi.site = FALSE)

## See the full multi-site workflow in the vignette
```

`micromollusk` *Micromollusks of marine shallow sandy bottoms around Cayo Nuevo, Gulf of Mexico, Mexico*

Description

Presence/absence of 68 species registered in six cores of 4 cm diameter and 10 cm depth taken in sandy bottoms around Cayo Nuevo, Gulf of Mexico, Mexico

Usage

```
data("micromollusk")
```

Format

A data frame with 6 observations on the following 69 variables.

`site` a numeric vector

`Leptochiton.sp.` a numeric vector

`Ischnochiton..Ischnochiton..erythronotus` a numeric vector

`Arcidae.sp.` a numeric vector

`Arca.imbricata` a numeric vector

`Barbatia.domingensis` a numeric vector

`Bentharca.sp.` a numeric vector

`Arcopsis.adamsi` a numeric vector

`Crenella.sp.` a numeric vector

`Anomia.sp..` a numeric vector

`Carditopsis.smithii` a numeric vector

`Lucinidae..` a numeric vector

`Chama.sinuosa` a numeric vector

`Chama.sp.` a numeric vector

`Galeommatidae.sp.` a numeric vector

`Chione.elevata` a numeric vector

`Semele.bellastriata` a numeric vector

`Gastropoda.sp..1..` a numeric vector

`Gastropoda.sp..2..` a numeric vector

`Gastropoda.sp..3..` a numeric vector

`Diodora.minuta` a numeric vector

`Diodora.sp...` a numeric vector

`Scissurella.redferni` a numeric vector

`Synaptocochlea.picta` a numeric vector

`Lodderena.ornata` a numeric vector

`Cerithium.sp...` a numeric vector

`Sansonia.tuberculata` a numeric vector

`Iniforis.turristhomae` a numeric vector

`Metaxia.rugulosa` a numeric vector

`Cerithiopsis.cf..iuxtafuniculata` a numeric vector

`Cerithiopsis.sp.` a numeric vector

`Vermetidae.incertae.sedis.irregularis` a numeric vector

`Dendropoma.corrodens` a numeric vector

`Vermetid.sp..C` a numeric vector

`Petaloconchus.mcgintyi` a numeric vector

`Thylacodes.sp.` a numeric vector

`Alvania.auberiana` a numeric vector

`Alvania.colombiana` a numeric vector

`Alvania.sp.` a numeric vector

`Simulamerelina.caribaea` a numeric vector

`Schwartziella.fischeri` a numeric vector

`Zebina.browniana` a numeric vector

`Zebina.sp.` a numeric vector

`Caecum.circumvolutum` a numeric vector

`Caecum.donmoorei` a numeric vector

`Caecum.floridanum` a numeric vector

`Caecum.johnsoni` a numeric vector

`Caecum.pulchellum` a numeric vector

`Caecum.textile` a numeric vector

`Caecum.sp..B` a numeric vector

`Meioceras.nitidum` a numeric vector

`Cochliolepis.striata` a numeric vector

`Parviturboides.interruptus` a numeric vector

`Vitrinella.sp.` a numeric vector

`Gibberula.lavalleeana` a numeric vector

`Prunum.apicinum` a numeric vector

`Volvarina.avena` a numeric vector

`Astyris.lunata` a numeric vector

`Phrontis.albus` a numeric vector

`Phrontis.sp.` a numeric vector

`Trachypollia.sp...`	a numeric vector
`Turridae.sp..1`	a numeric vector
`Turridae.sp..2..`	a numeric vector
`Turridae.sp..3..`	a numeric vector
`Ammonicera.lineofuscata`	a numeric vector
`Ammonicera.minortalis`	a numeric vector
`Rissoella.galba`	a numeric vector
`Pyramidellidae.sp.`	a numeric vector
`Pseudoscilla.babylonia`	a numeric vector

Details

Cayo Nuevo is a small reef cay located 240 km off the North-Western coast of Yucatan. Data correspond to a study about the biodiversity of marine benthic reef habitats off the Yucatan shelf (Ortigosa, Suarez-Mozo, Barrera et al. 2018).

Source

https://doi.org/10.3897/zookeys.779.24562

References

Ortigosa, D., Suarez-Mozo, N. Y., Barrera, N. C., & Simoes, N. (2018). First survey of Interstitial molluscs from Cayo Nuevo, Campeche Bank, Gulf of Mexico. Zookeys, 779. doi:10.3897/zookeys.779.24562

Examples

```
data(micromollusk)
```

`pilot` *Epibionts on Caribbean mangrove roots: pilot data*

Description

Data corresponds to a pilot study abput epibenthic organisms in mangrove roots from Laguna de La Restinga National Park, Venezuela (Guerra-Castro et al. 2011).

Usage

```
data("pilot")
```

Format

A data frame with 180 observations on the following 118 variables.

`Sector`	a factor with levels `E` `I` `M`
`Site`	a numeric vector
`sp1`	a numeric vector
`sp2`	a numeric vector
`sp3`	a numeric vector
`sp4`	a numeric vector
`sp5`	a numeric vector
`sp6`	a numeric vector
`sp7`	a numeric vector
`sp8`	a numeric vector
`sp9`	a numeric vector
`sp10`	a numeric vector
`sp11`	a numeric vector
`sp12`	a numeric vector
`sp13`	a numeric vector
`sp14`	a numeric vector
`sp15`	a numeric vector
`sp16`	a numeric vector
`sp17`	a numeric vector
`sp18`	a numeric vector
`sp19`	a numeric vector
`sp20`	a numeric vector
`sp21`	a numeric vector
`sp22`	a numeric vector
`sp23`	a numeric vector
`sp24`	a numeric vector
`sp25`	a numeric vector
`sp26`	a numeric vector
`sp27`	a numeric vector
`sp28`	a numeric vector
`sp29`	a numeric vector
`sp30`	a numeric vector
`sp31`	a numeric vector
`sp32`	a numeric vector
`sp33`	a numeric vector

`sp34`	a numeric vector
`sp35`	a numeric vector
`sp36`	a numeric vector
`sp37`	a numeric vector
`sp38`	a numeric vector
`sp39`	a numeric vector
`sp40`	a numeric vector
`sp41`	a numeric vector
`sp42`	a numeric vector
`sp43`	a numeric vector
`sp44`	a numeric vector
`sp45`	a numeric vector
`sp46`	a numeric vector
`sp47`	a numeric vector
`sp48`	a numeric vector
`sp49`	a numeric vector
`sp50`	a numeric vector
`sp51`	a numeric vector
`sp52`	a numeric vector
`sp53`	a numeric vector
`sp54`	a numeric vector
`sp55`	a numeric vector
`sp56`	a numeric vector
`sp57`	a numeric vector
`sp58`	a numeric vector
`sp59`	a numeric vector
`sp60`	a numeric vector
`sp61`	a numeric vector
`sp62`	a numeric vector
`sp63`	a numeric vector
`sp64`	a numeric vector
`sp65`	a numeric vector
`sp66`	a numeric vector
`sp67`	a numeric vector
`sp68`	a numeric vector
`sp69`	a numeric vector
`sp70`	a numeric vector

`sp71`	a numeric vector
`sp72`	a numeric vector
`sp73`	a numeric vector
`sp74`	a numeric vector
`sp75`	a numeric vector
`sp76`	a numeric vector
`sp77`	a numeric vector
`sp78`	a numeric vector
`sp79`	a numeric vector
`sp80`	a numeric vector
`sp81`	a numeric vector
`sp82`	a numeric vector
`sp83`	a numeric vector
`sp84`	a numeric vector
`sp85`	a numeric vector
`sp86`	a numeric vector
`sp87`	a numeric vector
`sp88`	a numeric vector
`sp89`	a numeric vector
`sp90`	a numeric vector
`sp91`	a numeric vector
`sp92`	a numeric vector
`sp93`	a numeric vector
`sp94`	a numeric vector
`sp95`	a numeric vector
`sp96`	a numeric vector
`sp97`	a numeric vector
`sp98`	a numeric vector
`sp99`	a numeric vector
`sp100`	a numeric vector
`sp101`	a numeric vector
`sp102`	a numeric vector
`sp103`	a numeric vector
`sp104`	a numeric vector
`sp105`	a numeric vector
`sp106`	a numeric vector
`sp107`	a numeric vector

`sp108`	a numeric vector
`sp109`	a numeric vector
`sp110`	a numeric vector
`sp111`	a numeric vector
`sp112`	a numeric vector
`sp113`	a numeric vector
`sp114`	a numeric vector
`sp115`	a numeric vector
`sp116`	a numeric vector

Details

Data consists of the coverage (by point-intercept) of 116 taxa identified in 180 mangrove roots, sampled under a hierarchically nested spatial design that included six random sites within each of three sectors of the lagoon system corresponding to a strong environmental gradient: external (E), intermediate (M), and internal (I). The abundance of epibenthic organisms of 10 roots were described within each site, producing a total of 60 roots in each sector. The analysis of these pilot data defined the sampling design used by Guerra-Castro et al. (2016).

Source

https://www.interciencia.net/wp-content/uploads/2018/01/923-GUERRA-8.pdf

References

Guerra-Castro, E., J. J. Cruz-Motta, and J. E. Conde. 2011. Cuantificación de la diversidad de especies incrustantes asociadas a las raíces de Rhizophora mangle L. en el Parque Nacional Laguna de La Restinga. Interciencia 36:923-930.

Guerra-Castro, E. J., J. E. Conde, and J. J. Cruz-Motta. (2016). Scales of spatial variation in tropical benthic assemblages and their ecological relevance: epibionts on Caribbean mangrove roots as a model system. Marine Ecology Progress Series 548:97-110.

Examples

```
data(pilot)
str(pilot)
```

`plot_ssp` *SSP Plot: Visualization of MultSE and Sampling Effort*

Description

Plots the relationship between MultSE and sampling effort using results from SSP simulations.

Usage

```
plot_ssp(xx, opt, multi.site)
```

Arguments

`xx`	A data frame generated by `summary_ssp`.
`opt`	A vector or data matrix generated by `ioptimum`.
`multi.site`	Logical. Indicates whether several sites were simulated.

Details

This function visualizes the behavior of MultSE (pseudo-multivariate standard error) as sampling effort increases. If simulations involve two sampling scales (e.g., sites and samples), separate graphs are generated. Two shaded bands highlight sub-optimal (light grey) and optimal (dark grey) improvements in precision. The graph also displays the relative gain in precision (as cumulative percentage) for each level of sampling effort, compared to the lowest.

This visualization helps identify when additional sampling effort results in diminishing returns. The plot is generated using `ggplot2` and can be further customized.

Value

A `ggplot2` object.

Note

This is an exploratory plot and can be edited or extended using standard `ggplot2` functions.

References

Guerra-Castro, E.J., Cajas, J.C., Simões, N., Cruz-Motta, J.J., & Mascaró, M. (2021). SSP: an R package to estimate sampling effort in studies of ecological communities. Ecography 44(4), 561-573. doi: doi:10.1111/ecog.05284

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer.

See Also

`ggplot2`

Examples

```
## Single site: micromollusk from Cayo Nuevo (Yucatan, Mexico)
data(micromollusk)
par.mic <- assempar(data = micromollusk, type = "P/A", Sest.method = "average")
sim.mic <- simdata(par.mic, cases = 2, N = 20, sites = 1)
sam.mic <- sampsd(dat.sim = sim.mic, Par = par.mic, transformation = "P/A",
                  method = "jaccard", n = 10, m = 1, k = 3)
summ.mic <- summary_ssp(results = sam.mic, multi.site = FALSE)
opt.mic <- ioptimum(xx = summ.mic, multi.site = FALSE)
plot_ssp(xx = summ.mic, opt = opt.mic, multi.site = FALSE)

## See the full multi-site workflow in the vignette
```

`sampsd` *Sampling Simulated Data and Estimation of Multivariate Standard Errors*

Description

For each simulated data set, this function performs repeated sampling across a range of effort levels and estimates the corresponding MultSE (pseudo-multivariate standard error) using dissimilarity-based methods.

Usage

```
sampsd(dat.sim, Par, transformation, method, n, m, k)
```

Arguments

`dat.sim`	A list of simulated data sets generated by `simdata`.
`Par`	A list of parameters estimated by `assempar`.
`transformation`	Mathematical transformation to reduce the influence of dominant species: one of "square root", "fourth root", "Log (X+1)", "P/A", or "none".
`method`	Dissimilarity metric to use, passed to `vegdist` (e.g., "bray", "jaccard", "gower").
`n`	Maximum number of sampling units per site (must be <= total units available).
`m`	Maximum number of sites to sample per data set (must be <= total number of sites).
`k`	Number of repetitions of each sampling configuration (samples × sites) for each data set.

Details

For multi-site simulations, the function selects subsets of sites (from 2 to `m`) and then draws `n` samples per site using a two-stage sampling method with inclusion probabilities (Tillé, 2006). For single-site simulations, repeated samples of size 2 to `n` are taken without replacement.

Each sample undergoes the selected transformation and a dissimilarity matrix is computed. MultSE is estimated using:

- *Single site:* pseudo-variance, with $MultSE = \sqrt{(}V/n)$
- *Multiple sites:* mean squares from a PERMANOVA model (residual and site effects)

This procedure is computationally intensive, especially with large `k`. Start with low values for exploration.

Value

A matrix containing the estimated MultSE values for each simulated data set, sampling effort combination, and repetition. This matrix is used by `summary_ssp`.

Note

For quick exploratory analysis, use small `k`. Once optimal sampling effort is explored, rerun with larger `k` (e.g. 100). Computation time will increase accordingly.

References

Anderson, M. J., & Santana-Garcon, J. (2015). Measures of precision for dissimilarity-based multivariate analysis of ecological communities. Ecology Letters, 18(1), 66–73.

Guerra-Castro, E. J., Cajas, J. C., Simoes, N., Cruz-Motta, J. J., & Mascaro, M. (2021). SSP: An R package to estimate sampling effort in studies of ecological communities. Ecography, 44(4), 561–573. doi:10.1111/ecog.05284

Tillé, Y. (2006). Sampling Algorithms. Springer, New York.

See Also

`assempar`, `simdata`, `summary_ssp`, `vegdist`

Examples

```
## Single site example
data(micromollusk)
par.mic <- assempar(data = micromollusk, type = "P/A", Sest.method = "average")
sim.mic <- simdata(par.mic, cases = 2, N = 20, sites = 1)
sam.mic <- sampsd(dat.sim = sim.mic, Par = par.mic, transformation = "P/A",
                  method = "jaccard", n = 10, m = 1, k = 3)

## For a multiple-site example, see the vignette
```

`simdata` *Simulation of Ecological Data Sets*

Description

Simulates multiple ecological data sets using parameters estimated from a pilot study. The output can be used in downstream SSP functions for quality evaluation and sampling effort estimation.

Usage

```
simdata(Par, cases, N, sites, jitter.base = 0)
```

Arguments

`Par`	A list of parameters estimated by `assempar`.
`cases`	Number of data sets to simulate.
`N`	Number of samples to simulate in each site.
`sites`	Number of sites to simulate in each data set.
`jitter.base`	Numeric scalar in $[0, 1)$. Standard deviation multiplier used to add Gaussian jitter to `fs` and `fw`. Defaults to 0.

Details

Presence/absence data are simulated using Bernoulli trials based on empirical frequencies of occurrence among sites (for site-level presence) and within sites (for local occurrence patterns). To better reflect realistic variability among nested sampling units (e.g., sites within regions), the simulation can apply controlled perturbations to the base parameters. This jittering introduces stochastic variation in occurrence probabilities across sites, while preserving the overall probabilistic structure of each species. As a result, simulated communities exhibit levels of multivariate dispersion closer to those observed in empirical data. These matrices are then converted into abundance matrices using values drawn from Poisson or negative binomial distributions (for count data), or from log-normal distributions (for continuous data like coverage or biomass), depending on the aggregation properties estimated in the pilot data.

This process is repeated `cases` times, producing a list of simulated data sets that reflect the statistical properties of the original assemblage, but without incorporating environmental constraints or species co-occurrence structures.

Value

A list of simulated community data sets, to be used by `datquality` and `sampsd`.

Note

This simulation assumes that differences in composition or abundance are due to spatial aggregation, as captured by the pilot data. It does not incorporate environmental gradients or species associations. For more advanced modeling of species associations, copula-based approaches as suggested by Anderson et al. (2019) may be integrated in future versions of SSP.

When using `jitter.base`, the simulation introduces stochastic perturbations to site-level parameters. For reproducibility, use `set.seed()` before calling `simdata()`.

References

Anderson, M. J., & Walsh, D. C. I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? Ecological Monographs, 83(4), 557–574.

Anderson, M. J., de Valpine, P., Punnett, A., & Miller, A. E. (2019). A pathway for multivariate analysis of ecological communities using copulas. Ecology and Evolution, 9, 3276–3294.

Guerra-Castro, E.J., Cajas, J.C., Simões, N., Cruz-Motta, J.J., & Mascaró, M. (2021). SSP: an R package to estimate sampling effort in studies of ecological communities. Ecography 44(4), 561-573. doi: doi:10.1111/ecog.05284

McArdle, B. H., & Anderson, M. J. (2004). Variance heterogeneity, transformations, and models of species abundance: a cautionary tale. Canadian Journal of Fisheries and Aquatic Sciences, 61, 1294–1302.

See Also

`sampsd`, `datquality`

Examples

```
## Single site simulation
data(micromollusk)
par.mic <- assempar(data = micromollusk, type = "P/A", Sest.method = "average")
sim.mic <- simdata(par.mic, cases = 2, N = 10, sites = 1)

## Multiple site simulation
data(sponges)
par.spo <- assempar(data = sponges, type = "counts", Sest.method = "average")
set.seed(42)
sim.spo <- simdata(par.spo, cases = 2, N = 20, sites = 10, jitter.base = 0.5)
```

`sponges` *Sponges in Alacranes Reef National Park (ARNP), Gulf of Mexico, Mexico*

Description

Counts of 41 species of sponges in 36 transects of 20 m * 1 m across 8 sites around ARNP

Usage

```
data("sponges")
```

Format

A data frame with 36 observations on the following 42 variables.

`site` Factor w/ 6 levels

`Agelas.clathrodes` a numeric vector

`Agelas.dispar` a numeric vector

`Agelas.tubulata` a numeric vector

`Agelas.wiedenmayeri` a numeric vector

`Aiolocroia.crassa` a numeric vector

`Amphimedon.copressa` a numeric vector

`Aplysina.archeri` a numeric vector

`Aplysina.cauliformis` a numeric vector

`Aplysina.fistularis` a numeric vector

`Aplysina.fulva` a numeric vector

`Aplysina.insularis` a numeric vector

`Aplysina.lacunosa` a numeric vector

`Callyspongia.plicifera` a numeric vector

`Callyspongia.vaginalis` a numeric vector

`Callispongia.fallax` a numeric vector

`Callispongia.armigera` a numeric vector

`Cliona.delitrix` a numeric vector

`Cliona.varians` a numeric vector

`Cribochalina.vascolum` a numeric vector

`Dragmacidon.sp.` a numeric vector

`Dysidea.variabilis` a numeric vector

`Ectyoplasia.ferox` a numeric vector

`Geodia.neptuni` a numeric vector

`Hymeniacidon.caerulea` a numeric vector

`Iotrochota.birotulata` a numeric vector

`Igernella.notabilis` a numeric vector

`Ircinia.felix` a numeric vector

`Ircinia.strobilina` a numeric vector

`Monanchora.arbuscula` a numeric vector

`Mycale.laxissima` a numeric vector

`Mycale.laevis` a numeric vector

`Nipahtes.amorpha` a numeric vector

`Niphates.erecta` a numeric vector

`Niphathes.digitalis` a numeric vector

`Phorbas.amaranthus` a numeric vector

`Scopalina.rutzleri` a numeric vector

`Svenezea.flava` a numeric vector

`Spirastrella.coccinea` a numeric vector

`Verongula.reswigui` a numeric vector

`Verongula.rigida` a numeric vector

`Xestospongia.muta` a numeric vector

Details

This data corresponds to a pilot study about sponge biodiversity in reef habitats in the Yucatán shelf (Ugalde et al., 2015)

Source

https://biotaxa.org/Zootaxa/article/view/zootaxa.3911.2.1

References

Ugalde, D., Gomez, P., & Simoes, N. (2015). Marine sponges (Porifera: Demospongiae) from the Gulf of Mexico, new records and redescription of Erylus trisphaerus (de Laubenfels, 1953). Zootaxa, 3911(2), 151-183.

Examples

```
data(sponges)
str(sponges)
```

`summary_ssp` *Summary of MultSE for Each Sampling Effort in Simulated Data Sets*

Description

Computes the average MultSE (pseudo-multivariate standard error) for each sampling effort across simulated datasets, and estimates associated variation and rate of change.

Usage

```
summary_ssp(results, multi.site)
```

Arguments

`results`	A matrix generated by `sampsd` containing MultSE values for each simulation and sampling configuration.
`multi.site`	Logical. Indicates whether multiple sites were simulated.

Details

For each sampling effort in each simulated data set, the average MultSE is computed (Anderson & Santana-Garcon, 2015). The function then calculates the overall mean and associated lower and upper quantiles of these averages. To evaluate how precision improves with effort, the average MultSE values are relativized to the maximum (typically at the lowest effort), and a numerical forward finite difference derivative is calculated to approximate the rate of change.

This output is used to support the identification of optimal and redundant sampling efforts based on precision gain.

Value

A data frame summarizing MultSE for each sampling effort, including the mean, quantiles, relativized values, and estimated derivative.

Note

This data frame can be used to plot MultSE versus sampling effort and to apply cutoff rules using `ioptimum`.

References

Anderson, M. J., & Santana-Garcon, J. (2015). Measures of precision for dissimilarity-based multivariate analysis of ecological communities. Ecology Letters, 18(1), 66–73.

Guerra-Castro, E.J., Cajas, J.C., Simões, N., Cruz-Motta, J.J., & Mascaró, M. (2021). SSP: an R package to estimate sampling effort in studies of ecological communities. Ecography 44(4), 561-573. doi: doi:10.1111/ecog.05284

See Also

`sampsd`, `ioptimum`

Examples

```
## Single site example
data(micromollusk)
par.mic <- assempar(data = micromollusk, type = "P/A", Sest.method = "average")
sim.mic <- simdata(par.mic, cases = 2, N = 10, sites = 1)
sam.mic <- sampsd(dat.sim = sim.mic, Par = par.mic, transformation = "P/A",
                  method = "jaccard", n = 10, m = 1, k = 3)
summ.mic <- summary_ssp(results = sam.mic, multi.site = FALSE)

## See the full multi-site workflow in the vignette
```

Index