

Package ‘EcoIndR’

November 8, 2023

Version 2.0

Title Ecological Indicators

Author Castor Guisande Gonzalez

Maintainer Castor Guisande Gonzalez <castor@uvigo.es>

Description Calculates several indices, such as of diversity, fluctuation, etc., and they are used to estimate ecological indicators.

License GPL (>= 2)

Encoding UTF-8

Depends R (>= 3.1.1), picante, Rarity, vegan, splancs, FD

Suggests shape, plotrix, sp, geoR, mgcv, raster, alphahull, sf, methods

Repository CRAN

NeedsCompilation no

Date/Publication 2023-11-08 00:30:09 UTC

R topics documented:

| | |
|--------------------------|-----------|
| DER | 2 |
| EnVarIP | 13 |
| Rotifers | 14 |
| SINENVAP | 14 |
| VirtualSpecies | 24 |
| Index | 25 |

DER

DIFFERENTIATING SAMPLES USING RARITY, HETEROGENEITY, EVENNESS, TAXONOMIC, PHYLOGENETIC AND FUNCTIONAL DIVERSITY INDICES

Description

An algorithm for differentiating samples on the basis of the rarity, heterogeneity, evenness, taxonomic/phylogenetic and functional diversity indices that better reflect the differences among assemblages.

Usage

```
DER(data, Samples, Species, Taxon, TaxonFunc=NULL, TaxonPhyl=NULL,
pos=NULL, varSize="Richness", varColor="Rarity.G", Index=NULL, corr="sqrt",
palette= "heat.colors", size=c(1,5),digitsS=1, digitsC=2, ncolor=100,
transparency=1, references=TRUE, a=1.5, q=2.5, ResetPAR=TRUE, PAR=NULL,
dbFD=NULL, LEGENDS=NULL, TEXT=NULL, COLOR=c("#EEC591FF", "black", "grey50"),
file1="Diversity indices.csv", file2="Polar coordinates.csv",
file3="Indices and area of the polygon.csv", na="NA", dec=",", row.names=FALSE, save=TRUE)
```

Arguments

data Data file with the taxonomy, abundance of the species and functional traits (optional). The format of the file must be: an optional column with the position of labels' samples in the DER plot (blue column) in the same order as the variables with the species' abundance in the samples (red columns), the columns with the taxonomy of the species (as many as needed, green columns), the columns with the abundance of the species in each sample (red columns) and optionally the columns with the functional traits of the species. Each row is a unique species, genus, family, etc.

| Pos | Class | Subclass | Superorder | Order | Family | Genus | Species | J1.1 | M.1 | G.1 | F2.1 |
|-----|-------------|-------------|--------------|----------------|---------------|-----------------|----------------|------|-----|-----|------|
| 2 | Eurotatoria | Monogononta | Pseudotrocha | Plolima | Asplanchnidae | Asplanchnopus | multiceps | | | | |
| 4 | Eurotatoria | Monogononta | Pseudotrocha | Plolima | Euchlanidae | Beauchampieella | eudactylota | | | | |
| 3 | Eurotatoria | Monogononta | Pseudotrocha | Plolima | Brachionidae | Brachionus | quadridentatus | | | | |
| 2 | Eurotatoria | Monogononta | Pseudotrocha | Plolima | Notammatidae | Cephalodella | gibba | 1 | | | |
| 3 | Eurotatoria | Monogononta | Gnesiatrocha | Flosculariacea | Canochilidae | Canochilus | dossuarius | | | | 16 |
| 4 | Eurotatoria | Monogononta | Pseudotrocha | Plolima | Euchlanidae | Dipleuchlanis | propatula | | | | |
| 4 | Eurotatoria | Monogononta | Pseudotrocha | Plolima | Euchlanidae | Euchlanis | depidula | 4 | 1 | 23 | 6 |
| 4 | Eurotatoria | Monogononta | Pseudotrocha | Plolima | Euchlanidae | Euchlanis | sp1 | | 3 | | |
| 4 | Eurotatoria | Monogononta | Pseudotrocha | Plolima | Brachionidae | Keratella | trapias | 2 | 10 | | 5 |

Samples Variables with the abundance of the species in each sample: sampling sites, dates, etc.

Species Variable with the name of the species (without including the genus). It may be other node of the phylogenetic tree, such as the genus, family, etc., for genus level phylogenies, family level phylogenies, etc.

Taxon Variables with the taxonomy of the species (taxonomic diversity), as many levels as needed but without including the variable with the node of the argument

| | |
|--------------|---|
| | <i>Species.</i> |
| TaxonFunc | Optionally variables with the functional traits (functional diversity). |
| TaxonPhyl | Optionally the name of the RData file of the class phylo with the phylogeny. The file must be in the working directory. |
| pos | Optionally it is possible to indicate a column with the position of labels' samples in the DER plot. It must be as many as the number of samples and in the same order than the variables described in the argument <i>Samples</i> . Values of 1, 2, 3 and 4, respectively indicate positions below, to the left of, above and to the right of the specified coordinates. |
| varSize | This variable defines the size of the bubble in the DER plot. |
| varColor | This variable defines the color gradient of the bubbles in the DER plot. |
| Index | The four/five indices used in the DER algorithm. If it is NULL the algorithm select one, index of rarity, one of heterogeneity, one of evenness one of taxonomy and one of the functional group (if functional traits are provided in the argument <i>TaxonFunc</i>) that achieve a higher dispersion among samples in a polar coordinates system. |
| corr | Character string specifying the correction method to use, in the function <code>dbFD</code> , when the species-by-species distance matrix cannot be represented in a Euclidean space. Options are "sqrt" (default), "cailliez", "lingoes" or "none". |
| palette | The color gradient of the bubbles may be one of these palettes: "heat.colors", "terrain.colors", "gray.colors", "topo.colors" or "cm.colors", or any other option defined by the user. |
| size | Range of size of the bubbles. Two values: minimum and maximum size. |
| digitsS | Number of digits of the bubble size legend. |
| digitsC | Number of digits of the color legend. |
| ncolor | Gradient color of the color legend. |
| transparency | Transparency of the color gradient, from 0 to 1. |
| references | If it is TRUE the reference points are depicted on the DER plot. |
| a | Scale of Renyi diversity. |
| q | Scale of Tsallis diversity. |
| ResetPAR | If it is FALSE, the default condition of the function PAR of the package StatR is not placed and maintained those defined by the user in previous graphics. |
| PAR | It accesses the function PAR of the package StatR that allows to modify many different aspects of the graph. |
| dbFD | It accesses the function <code>dbFD</code> which allows to specify the arguments that calculates the functional diversity indices. |
| LEGENDS | It allows to modify the legend of the bubble size. |
| TEXT | It allows to modify the text of the labels in the bubbles. |
| COLOR | A vector with three values: color of the ellipse, color of the points in the legend of the size of the bubbles and color of the references points in the ellipse, respectively. |

| | |
|-----------|--|
| file1 | CSV FILES. Filename with values of total abundance, richness and the rarity, heterogeneity, evenness, taxonomic, phylogenetic and functional diversity indices of each sample. |
| file2 | CSV FILES. Filename with the polar coordinates of all samples considering the four/five selected indices. |
| file3 | CSV FILES. Filename with the area of the convex hull (alpha=6) and Euclidean distance obtained in the polar coordinates system for all combinations of the indices. |
| na | CSV FILE. Text that is used in the cells without data. |
| dec | CSV FILE. It defines if the comma "," is used as decimal separator or the dot ".". |
| row.names | CSV FILE. Logical value that defines if identifiers are put in rows or a vector with a text for each of the rows. |
| save | If it is TRUE, the CSV files are saved. |

Details

DER algorithm

The steps of DER algorithm are described below:

1. The function DER calculates the most often used indices (see below): a total of 39 indices that includes 2 of rarity, 14 of heterogeneity, 7 of evenness, 2 of taxonomic diversity, 8 of phylogenetic diversity and 6 of functional diversity. It is important to mention that the indices included in the groups of phylogenetic diversity and functional diversity, each explores a different facet of phylogenetic diversity (Kembel et al., 2010) and functional diversity (Laliberte et al., 2010), respectively.

Rarity indices

In the following equations S is the number of species (species richness) in the sample, s is the number of samples, p_{ij} is the presence or absence (0 or 1) of the species i in the sample j , r_{ij} is the number of records or abundance of the species i in the sample j , R_i is the total number of records or total abundance of the species i considering all samples, Q_i is the occurrence of species i , Q_{min} and Q_{max} are respectively the minimum and maximum occurrences in the species pool, r is the chosen rarity cutoff point (as a percentage of occurrence), w_i is the weight of the i th species in the assemblage, w_{min} and w_{max} the minimum and maximum weights respectively.

Leroy (Leroy et al., 2012; 2013)

$$w_i = \frac{1}{e^{(\frac{Q_i - Q_{min}}{rQ_{max}Q_{min}} * 0.97 + 1.05)^2}}$$

$$I_{RR} = \frac{\sum_{i=1}^S w_i}{w_{max} - w_{min}} - w_{min}$$

Geographical rarity This index is a novel contribution of this package.

$$GR = 1 - \frac{\sum_{i=1}^S \frac{\sum_{j=1}^s p_{ij}}{s}}{S}$$

Occurrential rarity This index is a novel contribution of this package.

$$OR = 1 - \frac{\sum_{i=1}^S \frac{\sum_{j=1}^s r_{ij}}{R_i}}{S}$$

Heterogeneity indices

In the following equations S is the number of species (species richness) in the sample, p_i is the abundance's proportion of species i , N is the total number of individuals in the sample, n_i is the number of individuals of the species i , n_{max} is the number of individuals of the most abundant species, a and q are the orders of Renyi and Tsallis indices respectively and, finally, in Fisher's alpha the index is the α parameter.

log Shannon-Wiener (S.W.LOG2) and *ln Shannon-Wiener (S.W)* (Wiener, 1939; 1948; 1949; Shannon, 1948; Shannon & Weaver, 1949). See Spellerger & Fedor (2013) for an explanation of the dual use of the terms Shannon-Wiener and Shannon-Weaver to refer to this diversity index.

$$H = - \sum_{i=1}^S p_i \log_2 p_i$$

$$H' = - \sum_{i=1}^S p_i \ln p_i$$

Fisher's alpha (Fisher et al., 1943)

$$\alpha x, \frac{\alpha x^2}{2}, \frac{\alpha x^3}{3}, \dots, \frac{\alpha x^n}{n}$$

Simpson (Simpson, 1949)

$$D_1 = 1 - \sum_{i=1}^S p_i^2$$

Inverse Simpson (InvSimpson) Williams (1964)

$$D_2 = \frac{1}{\sum_{i=1}^S p_i^2}$$

Brillouin (Brillouin, 1956)

$$H_B = \frac{\ln N! - \sum_{i=1}^S \ln n_i!}{N}$$

Margalef (Margalef, 1959)

$$D_{Mg} = \frac{(S - 1)}{\ln N}$$

Renyi entropy (Renyi, 1961)

$$H_a = \frac{1}{(1-a)} \log \sum_{i=1}^S p_i^a$$

Menhinick (Menhinick, 1964)

$$D_{Mn} = \frac{S}{\sqrt{N}}$$

McIntosh (McIntosh, 1967)

$$D_{Mc} = \frac{N - \sqrt{\sum_{i=1}^S n_i^2}}{N - \sqrt{N}}$$

Inverse Berger-Parker (InvB.P) (Berger & Parker, 1970)

$$D_{BP} = \frac{1}{\frac{n_{max}}{N}}$$

Hill numbers (Hill, 1973)

Hill-Renyi

$$N_a = e^{H_a}$$

Hill-Tsallis

$$N_q = (1 - (q-1)H_q)^{\frac{1}{(1-q)}}$$

where

a or q = 0 is species richness

a or q = 1 is Shannon's index (H')

a or q = 2 is Inverse Simpson's index (D_2)

a or q = Inf is Inverse Berger-Parker index (D_{BP})

Tsallis entropy (Patil & Taillie, 1982; Tsallis, 1988)

$$H_q = \frac{1}{(q-1)} \left(1 - \sum_{i=1}^S p_i^q\right)$$

Evenness indices

Annotations of the equations as mentioned for heterogeneity indices.

Simpson evenness (SimpsonE) (Simpson, 1949)

$$E_{D1} = \frac{D_1}{S}$$

Pielou (PielouE) (Pielou, 1966)

$$J' = \frac{H'}{\ln S}$$

McIntosh evenness (McIntoshE) (McIntosh, 1967)

$$E_{Mc} = \frac{N - \sqrt{\sum_{i=1}^S n_i^2}}{N - \frac{N}{\sqrt{S}}}$$

Hill evenness (HillE) (Hill, 1973)

It is used Hill-Renyi numbers where in N_2 the value of $a = 2$ and in N_1 the value of $a = 1$

$$E_{2,1} = \frac{N_2}{N_1}$$

Heip evenness (HeipE) (Heip, 1974)

$$E_{Heip} = \frac{e^{H'} - 1}{S - 1}$$

Camargo (CamargoE) (Camargo, 1992)

$$E_{Camargo} = 1 - \sum_{i=1}^S \sum_{j=i'+1}^S \frac{|p_i - p_j|}{s}$$

Smith and Wilson's Index (Evar) (Smith and Wilson, 1996)

$$E_{var} = 1 - \left(\frac{2}{\pi} \right) \left(\arctan \left(\frac{\sum_{i=1}^S \left(\ln n_i - \frac{\sum_{j=1}^S n_j}{S} \right)^2}{S} \right) \right)$$

Taxonomic diversity indices

In the following equations summation goes over species i and j , ω are the taxonomic distances among taxa, x are species abundances, and n is the total abundance for a site.

Taxonomic diversity (D) (Clarke, 1995; 1998; 2001):

$$\Delta = \frac{\sum \sum_{i < j} \omega_{ij} x_i x_j}{n(n-1)/2}$$

Taxonomic distinctness (Dstar) (Clarke and Warwick, 1998):

$$\Delta^* = \frac{\sum \sum_{i < j} \omega_{ij} x_i x_j}{\sum \sum_{i < j} x_i x_j}$$

Phylogenetic diversity indices

| | |
|-------------------------------------|----------------------|
| Faith's phylogenetic diversity | Faith (1992) |
| Mean pairwise phylogenetic distance | Webb et al. (2008) |
| Mean nearest taxon distance | Webb et al. (2008) |
| Phylogenetic species richness | Helmus et al. (2007) |
| Phylogenetic species variability | Helmus et al. (2007) |
| Phylogenetic species evenness | Helmus et al. (2007) |
| Phylogenetic species clustering | Helmus et al. (2007) |
| Quadratic entropy | Rao (1982) |

Functional diversity indices

In the following equations S is the number of species, where d_{ij} is the difference between the i -th and j -th species, p_i is the abundance's proportion of the species i , p_j is the abundance's proportion of the species j , EW is weighted evenness, $\text{dist}(i, j)$ is the Euclidean distance between species i and j , the species involved is branch l , w_i is the relative abundance of the species i , PEW_l is the partial weighted evenness, g_k are the coordinates of the center of gravity of the V species forming the vertices of the convex hull, x_{ik} is the coordinate of species i on trait k , dG_i is the Euclidean distance of the center of gravity, \bar{dG} is the mean distance of the S species to the center of gravity, Δd is the sum of abundance-weighted deviances and $\Delta|d|$ absolute abundance-weighted deviances.

Rao's quadratic entropy (Rao, 1982; Botta-Dukat, 2005):

$$Q = \sum_{i=1}^{S-1} \sum_{j=i+1}^{S-1} d_{ij} p_i p_j$$

$$d_{ij} = \sum_{k=1}^n \sum_{l=1}^n w_{kl} (X_{ik} - X_{jk})(X_{il} - X_{jl})$$

Functional group richness (FGR) Petchey and Gaston (2006)

Functional richness (FRic) is measured as the amount of functional space (convex hull volume) filled by the community (Villegger et al., 2008).

Functional evenness (FEve) (Villegger et al., 2008):

$$FEve = \frac{\sum_{l=1}^{S-1} \min(PEW_l, \frac{1}{S-1}) - \frac{1}{S-1}}{1 - \frac{1}{S-1}}$$

$$PEW_l = \frac{EW_l}{\sum_{l=1}^S EW_l} \quad EW_l = \frac{\text{dis}(i,j)}{w_i + w_j}$$

Functional divergence (FDiv) (Villegger et al., 2008):

$$FDiv = \frac{\Delta d + \bar{dG}}{\Delta|d| + \bar{dG}}$$

$$\Delta|d| = \sum_{i=1}^S w_i |dG_i - \bar{dG}| \quad \Delta d = \sum_{i=1}^S w_i (dG_i - \bar{dG})$$

$$\bar{dG} = \frac{1}{S} \sum_{i=1}^S dG_i \quad dG_i = \sqrt{\sum_{k=1}^T (x_{ik} - g_k)^2} \quad g_k = \frac{1}{V} \sum_{i=1}^V x_{ik}$$

Functional dispersion (FDis) (Laliberte and Legendre, 2010):

$$FDisp = \frac{\sum a_j z_j}{\sum a_j} \quad c = \frac{\sum a_j x_{ij}}{\sum a_j}$$

2. Each index is transformed to a scale range between 0 and 1 for all samples with the following equation:

$$\frac{(\text{index of the sample} - \text{min})}{(\text{max} - \text{min})}$$

where *min* and *max* are the minimum and maximum values of the index considering all samples, respectively.

3. With the standardized values of the indices, the algorithm calculates the polar coordinates of all samples with all possible combinations among all groups of indices. Therefore, in each combination an index of each group of rarity, heterogeneity (species richness is included in this group), evenness, taxonomic/ phylogenetic diversity and functional diversity (if it is included functional traits in the analysis) is used for calculating the polar coordinates of all samples. In the group of taxonomic/phylogenetic diversity the user must use either taxonomy or a phylogenetic tree, so either taxonomic diversity or phylogenetic diversity indices are used in the algorithm. The X and Y polar coordinates for each sample are estimated using the following equations:

$$X = \sum_{i=1}^4 |z_j| \cos(\alpha) \quad Y = \sum_{i=1}^4 |z_j| \sin(\alpha)$$

where *z* is the standardized value of the index *j* of the four groups considered.

Each index is assigned an angle (α). Degrees to radians angle conversion is carried out assuming that 1 degree = 0.0174532925 radians.

4. With the polar coordinates of the samples obtained for each combination, it is calculated the convex hull ($\alpha = 6$) and the mean Euclidean distance, and the values are saved in a file.

5. The algorithm selects the combination of indices with the highest value of the mean between convex hull and mean Euclidean distance among samples, therefore priority is given to maximize dispersion among samples (see Fig. 1). The polar coordinates of the selected combination are depicted on a diagram, where it is possible to see the differences in rarity, heterogeneity, evenness and taxonomic/phylogenetic diversity and/or functional diversity (if it is included) among assemblages.

6. Finally, DER function allows the user to select the four/five indices to be used in the diagram,

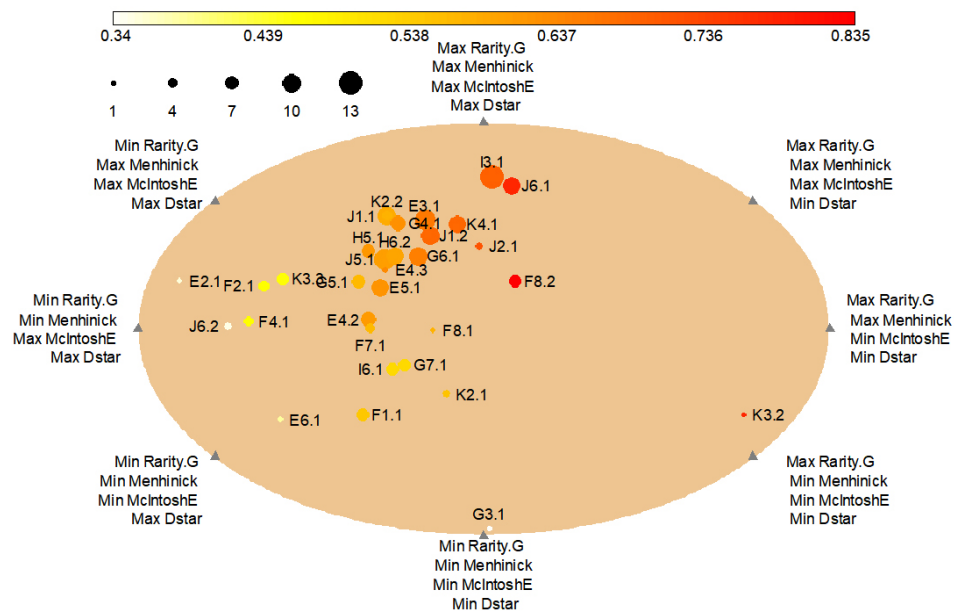
so the algorithm of selecting the combination with the maximum dispersion among samples is not applied.

FUNCTIONS

The index Fisher alpha was estimated with the function `fisher.alpha`, the index Renyi with the function `renyi`, the index Tsallis with the function `tsallis`, the taxonomic diversity and taxonomic distinctness with the functions `taxa2dist` and `taxondive`, all of them of the package `vegan` (Oksanen et al., 2016). The ellipse is depicted with the function `plotellipse` of the package `shape` (Soetaert, 2016). The convex hull (alpha=6) was calculated with the function `areapl` of the package `splancs` (Bivand et al., 2016). The color legend of DER plot was depicted with the function `color.legend` of the package `plotrix` (Lemon et al., 2016). The rarity index of Leroy was calculated with the functions `rWeights` and `lrr`, both of the package `Rarity` (Leroy et al., 2012; 2103; Leroy, 2016). The functional diversity indices were calculated with the function `dbFD` of the package `FD` (Laliberté et al., 2015). The phylogenetic indices were calculated with the functions `psv`, `psr`, `pse`, `psc`, `raoD`, `mntd`, `mpd` and `pd` of the package `picante` (Kembel et al. 2010 2016)

EXAMPLE

The example without functional diversity is a dataset with the abundance of rotifers species in ponds (see table 1 in Mazuelos et al., 1993). In the argument `Index` were selected `Rarity`, `Menhinick`, `McIntoshE` and `Dstar`, which are the indices selected by the algorithm when `Index=NULL` (default option). The sample G3.1 had the lowest values of the indices of rarity, heterogeneity, evenness and taxonomic diversity and the pond I3.1 the highest values for all indices.



Value

It is depicted a plot of polar coordinates estimated with the rarity, heterogeneity, evenness, taxonomic/phylogenetic and functional diversity indices, CSV files are saved with all the indices, the polar coordinates estimated with the indices specified in the argument `Index` or estimated by the algorithm, and the area of the convex hull and mean Euclidean distance obtained in the polar coor-

dinates system for all combinations of the indices.

References

- Berger, W.H., Parker, F.L. (1970) Diversity of planktonic Foramenifera in deep sea sediments. *Science*, 168: 1345-1347.
- Bivand, R., Rowlingson, B., Diggle, P., Petris, G., Eglén, S. (2016) Spatial and Space-Time Point Pattern Analysis. R Package Version 2.01-39. <https://CRAN.R-project.org/package=splancs>
- Botta-Dukat, Z. (2005) Rao's quadratic entropy as a measure of functional diversity based on multiple traits. *Journal of Vegetation Science*, 16: 533-540.
- Brillouin, L. (1956) *Science and information theory*. New York: Academic Press.
- Camargo, J.A. (1992) New diversity index for assessing structural alterations in aquatic communities. *Bulletin of Environmental Contamination and Toxicology*, 48, 428-434.
- Clarke, K.R. & Warwick, R.M. (1998). A taxonomic distinctness index and its statistical properties. *Journal of Applied Ecology*, 35: 523-531.
- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61: 1-10.
- Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of animal population. *Journal of Animal Ecology*, 12: 42-58.
- Heip, C. 1974. A new index measuring evenness. *Journal of the Marine Biological Association of the United Kingdom*, 54: 555-557.
- Helmus, M.R., Bland, T.J., Williams, C.J., Ives, A.R. (2007) Phylogenetic measures of biodiversity. *The American Naturalist*, 169: E68-E83.
- Hill, M.O. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54: 427-432.
- Hurlbert, S.H. 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, 52: 577-586.
- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26: 1463-1464.
- Kembel, S.W., Ackerly, D.D., Blomberg, S.P., Cornwell, W.K., Cowan, P.D., Helmus, M.R., Morlon, H. & Webb, C.O. (2016) R tools for integrating phylogenies and ecology. R Package Version 1.6-2. <https://CRAN.R-project.org/package=picante>
- Laliberte, E. & Legendre, P. (2010) A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, 91: 299-305.
- Laliberte, E., Legendre, P. & Shipley, B. (2015) Measuring functional diversity (FD) from multiple traits, and other tools for functional ecology. R package version 1.0-12. Available at: <https://CRAN.R-project.org/package=FD>.
- Lemon, J., Bolker, B., Oom, S., Klein, E., Rowlingson, B., Wickham, H., Tyagi, A., Etteradossi, O., Grothendieck, G., Toews, M., Kane, J., Turner, R., Witthoft, C., Stander, J., Petzoldt, T., Duursma, R., Biancotto, E., Levy, O., Dutang, C., Solymos, P., Engelmann, R., Hecker, M., Steinbeck, F., Borchers, H., Singmann, H., Toal, T. & Ogle, D. (2016) Various plotting functions. R package version 3.6-3. Available at: <https://CRAN.R-project.org/package=plotrix>.

- Leroy, B., Petillon, J., Gallon, R., Canard, A. & Ysnel, F. (2012) Improving occurrence-based rarity metrics in conservation studies by including multiple rarity cut-off points. *Insect Conservation and Diversity*, 5, 159-168.
- Leroy, B., Canard, A. & Ysnel, F. (2013) Integrating multiple scales in rarity assessments of invertebrate taxa. *Diversity and Distributions*, 19, 794-803.
- Leroy, B. (2016) Calculation of Rarity Indices for Species and Assemblages of Species. R package version 1.3-4. Available at: <https://CRAN.R-project.org/package=Rarity>.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H. & Wagner, H. (2016) Community Ecology Package. R package version 2.4-0. Available at: <https://CRAN.R-project.org/package=vegan>.
- Patil, G.P. & Taillie, C. (1982) Diversity as a concept and its measurement. *Journal of the Acoustical Society of America*, 77: 548-561.
- Pielou, E.C. (1966) The measurement of diversity in different types of biological collections. *Journal Theoretical Biology*, 13: 131-144.
- Petchey, O.L. & Gaston, K.J. (2002) Functional diversity (FD), species richness and community composition. *Ecology Letters*, 5: 402-411.
- Margalef, R. (1958) Information theory in ecology. *General Systems*, 3: 36-71.
- Mazuelos, N., Toja, J. & Guisande, C. (1993) Rotifers in ephemeral ponds of Doñana National Park. *Hydrobiologia*, 255/256: 429-434.
- McIntosh, R.P. (1967) An index of diversity and the relation of certain concepts to diversity. *Ecology*, 48: 392-404.
- Menhinick, E.P. (1964) A comparison of some species-individual diversity indices applies to samples of field insects. *Ecology*, 45: 859-861.
- Renyi, A. (1961) On measures of information and entropy. *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability Vol. 1*: 547-561.
- Rao, C.R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*, 21: 24-43.
- Shannon, C.E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, 27 (4): 379-423, 623-656.
- Shannon, C.E. & Weaver, W. (1949) *The Mathematical Theory of Communication*. The University of Illinois Press, Illinois.
- Simpson, E.H. (1949) Measurement of diversity. *Nature*, 163: 688.
- Smith, B. & J.B. Wilson (1996) A consumer's guide to evenness indices. *Oikos*, 76: 70-82.
- Soetaert, K. 2016. Functions for plotting graphical shapes, colors. R package version 1.4.2. Available at: <https://CRAN.R-project.org/package=shape>.
- Spellerger, I.F. & Fedor P.J. (2013) A tribute to Claude Shannon (1916-2001) and a plea for more rigorous use of species richness, species diversity and the "Shannon-Wiener" Index. *Global Ecology and Biogeography*, 12: 177-179.
- Tsallis, C. (1988) Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52: 479-487.
- Villegger, S., Mason, N.W.H. & Mouillot, D. (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89: 2290-2301.

Webb, C.O., Ackerly, D.D. & Kembel, W. (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24: 2098-2100.

Wiener, N. (1939) The ergodic theorem. *Duke Mathematical Journal*, 5: 1-18.

Wiener, N. (1948) *Cybernetics*. Wiley, New York.

Wiener, N. (1949) *The interpolation, extrapolation, and smoothing of stationary time series*. Wiley, New York.

Examples

```
#An example without functional diversity
```

```
data(Rotifers)
```

```
DER(data=Rotifers, Samples=c("J1.1", "K4.1", "G3.1", "F2.1", "K2.2", "F8.2", "F8.1",
"F1.1", "F4.1", "J2.1", "E5.1", "H5.1", "K3.2", "E4.2", "I6.1", "K2.1", "J5.1", "I3.1",
"K3.3", "G5.1", "E6.1", "J1.2", "J6.1", "G7.1", "G6.1", "G4.1", "E3.1", "E4.3", "E2.1",
"H6.2", "F7.1", "J6.2"), Species="Species", Taxon=c("Class", "Subclass",
"Superorder", "Order", "Family", "Genus"), pos="Pos", Index=c("Rarity.G", "Menhinick",
"McIntoshE", "Dstar"), save=FALSE)
```

EnVarIP

ENVIRONMENTAL VARIABLES IN THE IBERIAN PENINSULA

Description

Geographic coordinates with values of temperature and precipitation in the Iberian Peninsula.

Usage

```
data(EnVarIP)
```

Format

Data frame with longitude, latitude and values of temperature, rainfall and altitude, and the geographic coordinates of the polygons of Spain and Portugal in the Iberian Peninsula.

Rotifers

*ABUNDANCE OF ROTIFERS IN PONDS***Description**

Abundance of species of rotifers in ponds of Doñana National Park (Spain), which were obtained from Table 1 of Mazuelos et al. (1993).

Usage

```
data(Rotifers)
```

Format

An data frame with 40 columns: position of the label of the pond in the DER plot, 7 columns with the taxonomy of the species and the rest of columns are the abundance of the species of rotifers in the ponds.

References

Mazuelos, N., Toja, J., & Guisande, C. (1993) Rotifers in ephemeral ponds of Doñana National Park. *Hidrobiologia*, 255/256: 429-434.

SINENVAP

*SPATIAL INTERPOLATION OF VARIABLES WITHIN POLYGONS***Description**

An algorithm for finding an optimal spatial interpolation model for variables within polygons using kriging.

Usage

```
SINENVAP(data=NULL, var=NULL, dataLat=NULL, dataLon=NULL, polyLat=NULL,
polyLon=NULL, zonedata=NULL, zonepoly=NULL, convex=FALSE, alpha=0.07, ASC=NULL,
shape=NULL, shapenames=NULL, Area=NULL, validation=30, type.krige="OK",
trend.d="cte", trend.l="cte", model="AUTO", minimisation="optim", weights="npairs",
maxdist=NULL, nugget=NULL, sill=NULL, range=NULL, kappa=NULL, beta=NULL,
jitter="jitter", maxjitter=0.00001, direction=c(0,45,90,135), inside=TRUE,
error=FALSE, ResetPAR=TRUE, PAR=NULL, BOXPLOT=NULL, OUTLINE=FALSE, XLABP=NULL,
YLABP=NULL, XLAB=NULL, YLAB=NULL, XLABB="Model", YLABB="Accuracy measures",
MAIN="", XLIM=NULL, YLIM=NULL, ZLIM=NULL, COLOR="rev(heat.colors(100))",
COLORC="black", COLORB=NULL, COLORM="transparent", NLEVELS=10, LABCEX=0.6,
contour=TRUE, breaks=10, ndigits=0, xl=0, xr=0, pro=TRUE, cell=NULL,
file1="Predictions data.csv", file2="Predictions polygon.csv",
file3="Accuracy measures.csv", file4="Semivariogram.csv",
file5="Standard errors.csv", file6="Model selected.txt", na="NA",
dec=",", row.names=FALSE)
```

Arguments

| | |
|------------|---|
| data | Data file (CSV, RData or Excel) with the latitudes and longitudes and the values of the environmental variables. This file may also include the latitudes and longitudes of the polygons. Each polygon must be separated by a blank row. |
| var | Variable with the values of the environmental variable. |
| dataLat | Variable with the latitudes of the environmental variable. |
| dataLon | Variable with the longitudes of the environmental variable. |
| polyLat | If the geographic coordinates of the polygons are in the same file than the data of the variable, here it is indicated the variable with the latitudes of the polygons. |
| polyLon | If the geographic coordinates of the polygons are in the same file than the data of the variable, here it is indicated the variable with the longitudes of the polygons. |
| zonedata | If the latitude and longitude of the data are in UTM, it is necessary to specify the variable with the zone of each pair of coordinates in this argument. |
| zonepoly | If the latitude and longitude of the polygons are in UTM, it is necessary to specify the variable with the zone of each pair of coordinates in this argument. |
| convex | If it is TRUE, it is considered as polygon the alpha shape of the distribution of the variable. This option is useful if the variable is for instance the abundance of a species, so the spatial interpolation is performed considering the limits of the distribution of the species. |
| alpha | Alpha value of the alpha shape. |
| ASC | ASC file with the values of the variable. It is not necessary to specify the latitude and longitude of the variable, but it is mandatory to specify the polygon (in the argument <i>data</i> , <i>shape</i> or <i>Area</i>). |
| shape | It is possible to use a shape file for importing the coordinates of the polygons. In this case, it is not necessary to specify the latitude and longitude of the polygons in the arguments <i>polyLat</i> and <i>polyLon</i> . |
| shapenames | Variable in the shapefile with the names of the polygons. |
| Area | Only if using RWizard. It is also possible to use the polygons available in RWizard of administrative areas and river basins. A character with the name of the administrative area or a vector with several administrative areas (countries, regions, etc.) or river basins. In this case, it is not necessary to specify the latitude and longitude of the polygons in the arguments <i>polyLat</i> and <i>polyLon</i> . If it is "World" the entire world is plotted. |
| validation | Percentage of cases used from original data for validation. These data are not used for the estimation of the model and they are just used for evaluating the accuracy of the model (see details). If it is zero, all data are used for estimating the accuracy measures. If there are many data, a way for shortening the running time is to increase the number of data for validation, so reducing the number of data used for estimating the models. |
| type.krige | Type of kriging to be performed. Options are simple "SK" or ordinary kriging "OK". Kriging with external trend and universal kriging can be defined setting <code>type.krige="OK"</code> and specifying the trend model using the arguments <i>trend.d</i> and <i>trend.l</i> . |

| | |
|-------------------|---|
| trend.d | It specifies the trend (covariate) values at the data locations (see function krige.conv of the package <code>geoR</code>). |
| trend.l | It specifies the trend (covariate) values at prediction locations. It must be of the same type as for trend.d. Only used if prediction locations are provided in the argument locations (see function krige.conv of the package <code>geoR</code>). |
| model | If it is "AUTO", the algorithm tries to find the model with the highest accuracy measures (see details). It is also possible to select one or several of the following models: "exponential", "matern", "gaussian", "spherical", "circular", "cubic", "wave", "power", "linear", "cauchy", "gneiting", "powered.exponential", and/or "pure.nugget". |
| minimisation | Minimization function used to estimate the parameters of the model fitted to the semivariogram. The options are "optim", "nlm" or "nls" (see function variofit of the package <code>geoR</code>). |
| weights | Type weights used in the loss function when fitting the model to the semivariogram. The options are "npairs", "cressie" or "equal" (see the function variofit of the package <code>geoR</code>). |
| maxdist | Maximum distance in the semivariogram. If it is NULL, it is half of the maximum distance of the semivariogram. |
| nugget,sill,range | The value of the nugget variance parameter τ^2 , fixed value of the sill parameter (σ^2) and fixed value of the range parameter (ϕ). If they are NULL (default) the algorithm tries to find the optimal values for each of the models specified in the argument <i>model</i> . |
| kappa | One numerical value required for the following models: "matern", "cauchy", "gneiting.matern" and "powered.exponential", and two values for the model "gencauchy". If they are NULL (default) the algorithm tries to find the optimal values for each of the models specified in the argument <i>model</i> . |
| beta | Numerical value of the mean (vector) parameter. Only used if type.krige="SK". If it is NULL, it is automatically estimated by the algorithm. |
| jitter | It may be one of these three options: "NO" means no action, "jitter" means that jitters duplicated coordinates of the environmental variable, and "mean" means that the mean of the environmental variable is estimated for those duplicated coordinates. |
| maxjitter | Maximum jittering distance in decimal degrees. |
| direction | A vector with values of 4 angles, indicating the directions for which the variograms will be computed. Default corresponds to c(0,45,90,135 (degrees). |
| inside | If it is TRUE only those geographic coordinates of the environmental variables inside the polygons are considered for the estimation of the model. |
| error | If it is TRUE, a contour map with the standard errors is depicted. |
| ResetPAR | If it is FALSE, the default condition of the function PAR is not placed and maintained those defined by the user in previous graphics. |
| PAR | It accesses the function PAR that allows to modify many different aspects of the graph. |
| BOXPLOT | It allows to specify the characteristics of the function boxplot . |

| | |
|------------------|--|
| OUTLINE | If it is TRUE, the outliers are shown in the boxplot. |
| XLABP, YLABP | Legends of X and Y axes of the plot with the relationship between observed and predicted values of the model. |
| XLAB, YLAB | Legends of X and Y axes of the contour plot with the spatial interpolation. |
| XLABB, YLABB | Legends of X and Y axes of the boxplot . |
| MAIN | Main title of the contour plot with the spatial interpolation. |
| XLIM, YLIM, ZLIM | Limits of the contour plot. |
| COLOR | Palette of colours or a vector with the colours of the contour plot. |
| COLORC | Colour of the lines in the contour plot. |
| COLORB | Vector with the colours of the models or just one colour for all models of the boxplot. |
| COLORM | Colour of the administrative areas and river basins, if any area has been specified in the argument <i>Area</i> and <i>convexhull=TRUE</i> . |
| NLEVELS | Numeric vector of levels at which to draw contour lines. |
| LABCEX | Size of the text in the contour lines. |
| contour | If it is TRUE, the contour lines are depicted in the contour plot. |
| breaks | Number of breakpoints of the colour legend in the contour plot. |
| ndigits | Number of decimals in the legend of the colour scale in the contour plot. |
| x1, xr | The left and right limits of the colour legend considering the X axis of the contour plot. |
| pro | If it is TRUE, an automatic calculation is made in order to correct the aspect ratio y/x along latitude. |
| cell | Cell size in decimal degrees of the grid inside the polygons with the predictions of the model. If it is NULL, it is automatically estimated according to the limits of the polygons. To select an appropriate cell size according to the polygon size is important for shortening the running time. |
| file1 | CSV FILES. Filename with the predictions of the models. |
| file2 | CSV FILES. Filename with the predictions inside the polygons. |
| file3 | CSV FILES. Filename with accuracy measures of the models. |
| file4 | CSV FILES. Filename with values of the semivariogram. |
| file5 | CSV FILES. Filename with standard errors of the predictions. |
| file6 | TXT FILE. Model selected with indication of the accuracy measures. |
| na | CSV FILE. Text that is used in the cells without data. |
| dec | CSV FILE. It defines if the comma "," is used as decimal separator or the dot ".". |
| row.names | CSV FILE. Logical value that defines if identifiers are put in rows or a vector with a text for each of the rows. |

Details

SINENVAP algorithm

The aim of this algorithm is to select a model, from a set of different models, with the nugget variance parameter τ^2 , fixed value of the sill parameter (σ^2) and fixed value of the range parameter (ϕ), as close as possible to those values that generates an optimal spatial interpolation, and to validate the predictions obtained. The model and parameters selected by the algorithm may be utilized by users as references, or to make modifications for spatial interpolation prediction improvement.

The algorithm uses the package *geoR* (Ribeiro and Diggle, 2001; 2018) to estimate simple, ordinary, and universal kriging. The corresponding algorithm is detailed below.

1. Data with variable and polygon coordinates.

The algorithm must be supplied with variable values (argument *var*), latitude and longitude for each datum (arguments *dataLat* and *dataLon*), and polygon coordinates (arguments *polyLat* and *polyLon*), for the estimation of spatial interpolation.

Variable and polygon latitudes and longitudes may be in either decimal or UTM form. If they are in UTM form, a column with the zone of each coordinate in the *zonedata* and *zonepoly* arguments must be added for variable and polygon coordinates, respectively. Polygon variables and coordinates are not required to be in the same units. Therefore, variables may be in decimal form, and polygons in UTM form, vice versa, or both may be in the same units.

Variable and polygon information data may be in CSV, EXCEL, or RData files. ASC files (argument *ASC*) may also be used for the variable, but in this case, it is not necessary to provide latitude or longitude information for each datum in the *dataLat* and *dataLon* arguments. Polygon coordinates may be in the same file as variable data in CSV, EXCEL, or RData files.

If RWizard is used, any of the administrative areas available in RWizard may be chosen as polygons in the *Area* argument: countries, departments, provinces, etc. River basins may also be used as polygons, in the database available in RWizard (Gonzalez-Vilas et al., 2015).

Shape files may be used to import polygon coordinates with the *shape* argument. If a shape file or the polygons available in RWizard are used, coordinate specification is unnecessary in *polyLat* and *polyLon* arguments.

Finally, if the argument *convex=TRUE*, the alpha shape distribution is considered to be a polygon. This last option is useful when the variable is, for instance, the abundance of a species, such that spatial interpolation is performed considering the limits of species distribution. If *convex=TRUE*, the specification of information in *polyLat* and *polyLon* arguments is unnecessary.

2. Algorithm design.

Algorithm steps are described as follows:

1. If the argument *inside=TRUE* (default option), based on all variable data available, only those data inside the polygons are used to estimate spatial interpolation. If the argument *inside=FALSE*, all available variable data are used.
2. Duplicated coordinates may be treated in two ways: with the application of a jitter function (default) or with estimation of the mean value of the variable for duplicated coordinates.
3. Spatial interpolation estimates the values on a grid. Therefore, it is necessary to first create a grid with a fixed cell size, in which the spatial interpolation will be estimated. If the argument *cell=NULL* (default), the algorithm estimates the optimal grid cell size, in accordance with polygon

size. The output TXT provides information about the cell size chosen. The user may specify different cell sizes. Appropriate cell size selection, in accordance with polygon size, shortens running time.

4. Only those points in polygons are selected from the grid created, for the spatial interpolation prediction.

5. By the default $validation=30$, which means that 30% of variable data are not used in the spatial interpolation model. However, they are employed to test the model. If $validation=0$, all data are used in the model, and model validation is performed with all data. If validation is higher than zero, because validation data are randomly selected, spatial interpolation values may vary each time the script is implemented.

6. In order to test the model, the grid coordinates nearest to data coordinates are selected. Cross validation is performed by comparing variable data reserved for validation (see step 5) to spatial interpolation predictions on the grid, so as to verify which coordinates are nearest to the data reserved for validation.

7. Next, the semivariogram is depicted (Fig. 1). If the argument $maxdist=NULL$ (default), it is considered to be half of the maximum semivariogram distance for performing the models.

8. Any of the following models may be implemented: cauchy, circular, cubic, exponential, gaussian, gneiting, linear, matern, power, powered.exponential, pure.nugget, spherical, or wave. If $model="AUTO"$, all models are tested, with the exception of matern. The *variofit* function in the geoR package is used to find the nugget, range, and sill for each model. The data reserved for validation are compared to model predictions, and seven accuracy measures, described in the following section, are used to decide which model made the best prediction. The model that made the best prediction is chosen by the algorithm to depict variable spatial interpolation in the selected polygons.

3. Accuracy measures.

The following accuracy measures are used in the algorithm, so as to compare and evaluate model predictions, where n is the number of observations, and P and O are the predicted and observed values for each i datum, respectively.

The measures catalogued as "normalized" are adapted to reflect values of 1 when the model is most efficient (i.e., predictions are the same as observed values). Thus, in all accuracy measures used in the algorithm, the maximum value is 1. This indicates a model with maximum predictive power. Various measures were utilized to obtain improved evaluation framework (i.e. consideration of a group of skill scores that show different result characteristics) (see Li and Head, 2011).

3.1. *r-squared* (r^2). This is the square of the Pearson correlation coefficient (r) between the predictions of the interpolation model and the observed values. It ranges from 0 to 1.

$$r = \frac{\sum_{i=1}^n [(O_i - \bar{O})(P_i - \bar{P})]}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}}$$

3.2. *Normalized mean absolute error (NMAE)*. This is a measure of the average error between

predictions and observed values. It ranges from $-\infty$ to 1.

$$NMAE = 1 - \frac{\sum_{i=1}^n |P_i - O_i|}{\sum_{i=1}^n O_i}$$

3.3. *Normalized root mean square error (NRMSE)*. This measure shows the distribution error variability. It ranges from $-\infty$ to 1.

$$NRMSE = 1 - \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n O_i}}$$

3.4. *Nash-Sutcliffe coefficient (E)*. Nash-Sutcliffe efficiencies may range from $-\infty$ to 1 (Nash & Sutcliffe, 1970). An efficiency of 1 ($E = 1$) corresponds to a perfect match between model and observations. An efficiency of 0 indicates that the model predictions are as accurate as the mean of the observed data, whereas an efficiency less than zero ($-\infty < E < 0$) occurs when the observed mean is a better predictor than the model.

$$E = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$$

3.5. *Index of agreement (d)*. It was developed by Willmott (1981) as a standardized measure of the degree of model prediction error and varies between 0 and 1.

$$d = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2}$$

3.6. *Normalized relative mean absolute error (NRMAE)*. This is a modification of a measure developed by Li and Head (2011), whose maximum value is 1. According to the authors, this measure removes the effect of unit/scale and is not sensitive to changes in unit/scale. It ranges from $-\infty$ to 1.

$$NRMAE = 1 - \frac{\sum_{i=1}^n \frac{|P_i - O_i|}{O_i} * 100}{\sum_{i=1}^n O_i}$$

3.7. *Normalized relative root mean square error (NRRMSE)*. This is a modification of a measure

developed by Li and Head (2011), whose maximum value is 1. It ranges from 0 to 1.

$$NRRMSE = 1 - \left[\frac{\sum_{i=1}^n \left(\frac{[P_i - O_i]}{O_i} \right)^2}{\sum_{i=1}^n O_i} \right]^{\frac{1}{2}} * 100$$

FUNCTIONS

Spatial interpolation, using simple, ordinary, and universal kriging, is performed with `as.geodata`, `variofit`, `krige.conv` and `krige.control` functions, as well as the jitter of points with the `jitterDupCoords` function, all of which are from the `geoR` package (Ribeiro and Diggle, 2001; 2018).

The ASC files are loaded with the raster function from the raster package (Hijmans et al., 2018).

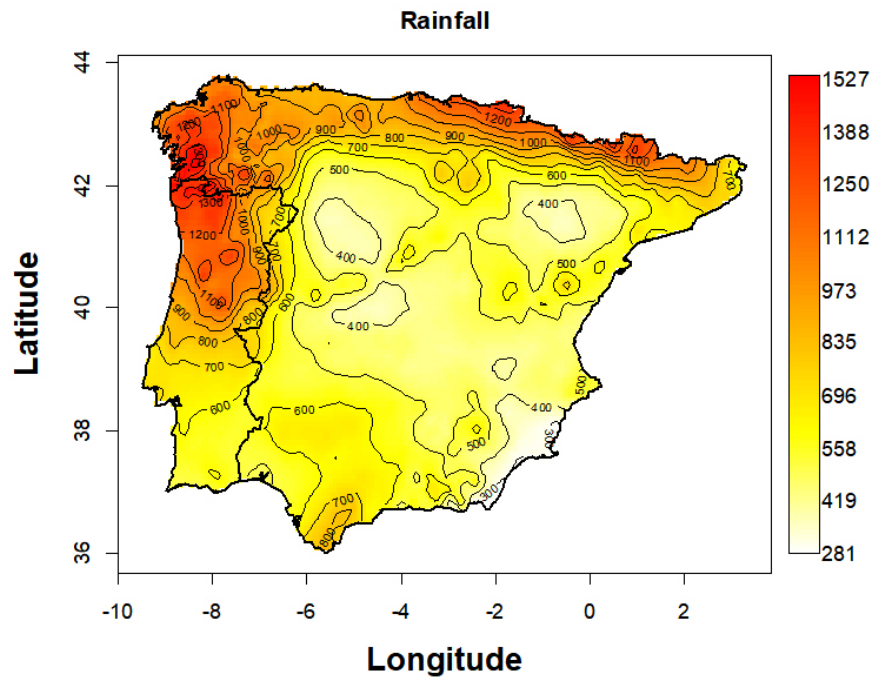
Points inside polygons are estimated with the `in.out` function from the `mgcv` package (Wood, 2018).

The `sp` package (Pebesma and Bivand, 2005; Pebesma et al., 2018) is used to process shape files.

The BreuschPagan test is performed with the `bptest` function from the `lmtest` package (Zeileis and Hothorn, 2002; Hothorn et al., 2018).

The color scale is depicted with the `color.legend` function from the `plotrix` package (Lemon, 2006; Lemon et al., 2018).

EXAMPLE The figure shows the contour map of the spatial interpolation of the example 1, the rainfall in the Iberian Peninsula.



Value

It is obtained:

1. A CSV file, called "Predictions data.CSV" by default, contains model data and predictions. If *validation=0*, the observed data are from the original dataset. Predicted values are those points inside the polygon, which are, spatially, the nearest neighbors to the observed data.
2. A CSV file, "Predictions polygon.CSV" by default, contains predictions for inside polygons.
3. A CSV file, called "Accuracy measures.CSV" by default, contains the values of the seven accuracy measures shown above, for all models.
4. A CSV file, called "Semivariogram.CSV" by default, contains semivariogram values.
5. A CSV file, called "Standard errors.CSV" by default, contains the standard prediction errors.
6. A TXT file called "Model selected.TXT" by default, contains the full details of the model selected by the algorithm.
7. A plot of the semivariogram, with the values used in the models, is depicted in green. Application of the maximum distance specified in the *maxdist* argument, and the points in red, yield a semivariogram without distance limitations.
8. The directional variogram in four directions.
9. A plot with the relationship between observed and predicted values. As mentioned above, the observed values may be either randomly selected values or those from the original dataset. If *validation=0*, the observed data are from the original dataset. Predicted values are those points inside the polygon, which are, spatially, the nearest neighbors to the observed data.
10. If the argument *model* is "AUTO", or there is more than one model, a boxplot is depicted with the median value of the seven accuracy measures from each model.
11. The contour plot with the spatial interpolation predictions of the selected model, i.e., the model with the highest accuracy measures mean.
12. If the argument *error=TRUE*, the contour plot is depicted with the selected model's standard errors.

References

- Hijmans, R.J., van Etten, J., Cheng, J., Sumner, M., Mattiuzzi, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Bivand, R., Busetto, L., Canty, M., Forrest, D., Ghosh, A., Golicher, D., Gray, J., Hiemstra, P., Karney, C., Mosher, S., Nowosad, J., Pebesma, E., Racine, E.B., Rowlingson, B., Shortridge, A., Venables, B., Wueest, R. (2018) Geographic Data Analysis and Modeling. R package version 2.8-4. Available at: <https://CRAN.R-project.org/package=raster>.
- Gonzalez-Vilas, L., Guisande, C., Vari, R. P., Pelayo-Villamil, P., Manjarres-Hernandez, A., Garcia-Rosello, E., Gonzalez-Dacosta, J., Heine, J., Perez-Costas, E., Granado-Lorencio, C., Palau-Ibars, A., Lobo, J. M. (2016) Geospatial data of freshwater habitats for macroecological studies: an example with freshwater fishes. *Journal of Geographical Information Science*, 30: 126-141.
- Lemon, J. (2006) Plotrix: a package in the red light district of R. *R-News*, 6: 8-12.
- Lemon, J., Bolker, B., Oom, S., Klein, E., Rowlingson, B., Wickham, H., Tyagi, A., Etteradossi, O., Grothendieck, G., Toews, M., Kane, J., Turner, R., Witthoft, C., Stander, J., Petzoldt, T., Duursma, R., Biancotto, E., Levy, O., Dutang, C., Solymos, P., Engelmann, R., Hecker, M., Steinbeck, F., Borchers, H., Singmann, H., Toal, T. & Ogle, D. (2018). Various plotting functions. R package version 3.7-4. Available at: <https://CRAN.R-project.org/package=plotrix>.

Li, J. & Heap, A.D. (2011) A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, 6: 248-251.

Nash, J.E. & Sutcliffe, J.V. (1970) River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, 10, 282-290.

Pebesma, E., Bivand, R.S., Rowlingson, B., Gomez-Rubio, V., Hijmans, R., Sumner, M., MacQueen, D., Lemon, J. & O'Brien, J. (2018) Classes and Methods for Spatial Data. R package version 1.3-1. Available at: <https://CRAN.R-project.org/package=sp>.

Pebesma, E.J. & Bivand, R.S. (2005) Classes and methods for spatial data in R. *R News*, 5: 9-13.

Ribeiro, P.A. & Diggle, P.J. (2001) geoR: a package for geostatistical analysis. *R-News* 1, 14-18.

Ribeiro, P.J. & Diggle, P.J. (2018) Analysis of Geostatistical Data. R package version 1.7-5.2.1. Available at: <https://CRAN.R-project.org/package=geoR>.

Willmott, C.J. (1981) On the validation of models. *Physical Geography*, 2, 184-194.

Wood, S. (2018) Mixed GAM Computation Vehicle with Automatic Smoothness. R package version 1.8-26. Available at: <https://CRAN.R-project.org/package=mgcv>.

Examples

```
## Not run:
```

```
#Example 1.
```

```
#An example with the geographic coordinates of the polygons
#in the same file than the environmental variable
```

```
data(EnVarIP)
```

```
SINENVAP(data=EnVarIP, var="Rainfall", dataLat="dataLat",
dataLon="dataLon", polyLat="polyLat", polyLon="polyLon",
model=c("cubic", "spherical"), MAIN="Rainfall", dec=".")
```

```
#Example 2. Only to be used with RWizard
```

```
#An example using the administrative areas available in RWizard.
```

```
data(EnVarIP)
```

```
@_Build_AdWorld_
```

```
SINENVAP(data=EnVarIP, var="Temperature", dataLat="dataLat", dataLon="dataLon",
Area = c("Galicia>A Coruña", "Galicia>Lugo", "Galicia>Ourense",
"Galicia>Pontevedra"), model=c("spherical"), MAIN="Temperature", ndigits=1, dec=".")
```

```
#Example 3. Only to be used with RWizard
```

```
#An example with a virtual species using as polygon the alpha
#shape of the species (argument convex=TRUE).
```

```
data(VirtualSpecies)
```

```
@_Build_AdWorld_
```

```
SINENVAP(data=VirtualSpecies, var="Probability", dataLat="Lat",
dataLon="Lon", model=c("circular", "exponential"), convex=TRUE,
Area=c("France"), validation=90, COLORM="#DEDE64", ndigits=2, dec=".")
```

```
## End(Not run)
```

VirtualSpecies

PROBABILITY OF A VIRTUAL SPECIES

Description

Probability of a virtual species in France.

Usage

```
data(VirtualSpecies)
```

Format

The 3 bioclimatic variables of WorldClim at a resolution of five arc-minutes (Hijmans et al., 2005) BIO8 (Mean Temperature of Wettest Quarter), BIO12 (Annual Precipitation) and BIO15 (Precipitation Seasonality, Coefficient of Variation) were used to generate the distribution of the virtual species by using the virtualspecies R package (Leroy et al. 2016; 2018). The limits were from -5° to 6° in latitude, and from 44° to 49°. Subsequently, 50% of the data were randomly selected, in order to simulate a random sampling of the species. The final continuous suitability maps obtained in this way were scaled between 0 and 1.

References

- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25: 1965-1978.
- Leroy, B., Meynard, C. N., Bellard, C., Courchamp, F. (2016) Virtualspecies, an R package to generate virtual species distributions. *Ecography*, 39: 599-607.
- Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F., Delsol, R. & Gaul, W. (2018) Generation of Virtual Species Distributions. R package version 1.4-4. Available at: <https://CRAN.R-project.org/package=virtualspecies>.

Index

- * **DER**
 - DER, [2](#)
- * **EnVarIP**
 - EnVarIP, [13](#)
- * **Rotifers**
 - Rotifers, [14](#)
- * **SINENVAP**
 - SINENVAP, [14](#)
- * **VirtualSpecies**
 - VirtualSpecies, [24](#)

areapl, [10](#)

boxplot, [16](#)

color.legend, [10](#)

dbFD, [3](#), [10](#)
DER, [2](#)

EnVarIP, [13](#)

fisher.alpha, [10](#)

Irr, [10](#)

krige.conv, [16](#)

plotellipse, [10](#)

renyi, [10](#)
Rotifers, [14](#)
rWeights, [10](#)

SINENVAP, [14](#)

taxa2dist, [10](#)
taxondive, [10](#)
tsallis, [10](#)

variofit, [16](#)
VirtualSpecies, [24](#)