

Rarefy: an R package for the calculation of taxonomic, functional and phylogenetic sample-based spatially explicit rarefaction curves

Elisa Thouverai^a, Sandrine Pavoine^b, Enrico Tordoni^{c,d}, Alessandro Chiarucci^d, Carlo Ricotta^e, Duccio Rocchini^{d,f}, Giovanni Bacaro^{a,*}

^a Department of Life Sciences, University of Trieste, Via L. Giorgeri 10, 34127 Trieste, Italy.

^b Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, 57 rue Cuvier, 75005 Paris, France.

^c Institute of Ecology and Earth Sciences, University of Tartu, J. Liivi 2, 50409 Tartu, Estonia.

^d Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum University of Bologna, Via Irnerio 42, 40126 Bologna, Italy

^e Department of Environmental Biology, University of Rome "La Sapienza", Piazzale Aldo Moro 4, 00185 Rome, Italy.

^f Czech University of Life Sciences Prague, Faculty of Environmental Sciences, Department of Applied Geoinformatics and Spatial Planning, Kamýcká 129, Praha–Suchbát 16500, Czech Republic

ARTICLE INFO

Keywords:

α -Diversity
 β -Diversity
 Directional accumulation curve
 Functional diversity
 Hill numbers
 Rao's quadratic entropy
 Species richness

ABSTRACT

Rarefaction methods are statistical techniques used to estimate diversity across increasing number of samples, generating rarefaction curves that depict diversity as a function of sample size. These methods are widely applied in ecological research to compare taxonomic, functional, and phylogenetic diversity across samples with varying collection efforts. However, incorporating spatially explicit rarefaction methods has become essential, as accounting for spatial autocorrelation substantially influences results and alters how biodiversity hotspots or conservation priorities are identified. This paper describes *Rarefy*, an R package that introduces novel functions capable of handling any diversity metric, allowing users to compute the expected values of taxonomic, functional, or phylogenetic indices for reduced sample sizes under spatially constrained, distance-based sampling unit arrangements. To showcase the package's functionalities, we estimated the functional diversity of plant communities along the northern Adriatic coastline. The spatially explicit rarefaction functions consistently produced lower diversity values compared to their non-spatial counterparts, reflecting the functional redundancy typical of spatially adjacent plant communities. These differences provide a practical diagnostic signal of spatial structure in the dataset. The *Rarefy* package offers ecologists a robust and straightforward tool to account for spatial constraints in rarefaction analyses, yielding more ecologically meaningful diversity estimates. Future developments will further integrate null-model procedures, extending the comparison between observed and randomized community diversity patterns illustrated in this study.

1. Introduction

Rarefaction and extrapolation are core tools in biodiversity research because ecological datasets are almost always incomplete and sampling effort differs across sites, periods, and methods (Hortal et al., 2015; Marchetto et al., 2024). By standardising diversity estimates to a comparable level of effort or sample coverage, rarefaction curves support fair comparisons in monitoring programs, meta-analyses, and applied conservation assessments. However, in many plot-based inventories and spatial surveys, sampling units are not independent: nearby samples tend to be more similar than distant ones because of spatial autocorrelation and environmental structure. In such cases, “standard” non-

spatial rarefaction can overstate the amount of independent information gained when additional plots are added and may lead to biased comparisons across areas or sampling designs (Bacaro et al., 2016). This motivates the need for tools that make spatial constraints explicit when constructing sample-based rarefaction curves, and that remain applicable to multiple facets of diversity (taxonomic, functional, phylogenetic).

Sample-based species accumulation curves have been widely used in ecology for comparing species richness in assemblages of various sizes after rarefaction to a common level of sampling effort or sample completeness (Chao and Jost, 2012; Colwell et al., 2004; Gotelli and Colwell, 2001), and for extrapolating species richness over larger

* Corresponding author.

E-mail address: gbacaro@units.it (G. Bacaro).

<https://doi.org/10.1016/j.ecoinf.2026.103803>

Received 5 February 2025; Received in revised form 28 April 2026; Accepted 29 April 2026

Available online 30 April 2026

1574-9541/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

sampling effort (Chao et al., 2014, 2015; Colwell et al., 2012). Usually, these curves are calculated as the mean of species richness obtained by randomly resampling, without replacement, the set of N sampling units and counting the number of species observed in 1, 2, ..., N plots (Bacaro et al., 2012). Since their introduction in the routine analysis of diversity, rarefaction techniques have also been proposed to rarefy functional (Ricotta et al., 2012) and phylogenetic (Chao et al., 2015) diversity metrics in relation to sampling size, but neither of these methods accounts for the autocorrelative structure in the distribution of data. More recently, Ricotta et al. (2019) offered new insights on this topic proposing a new way to summarize β -diversity along a gradient. In more detail, they proposed a metric able to summarize multiple-site dissimilarity as a result of a species turnover process, taking into account the spatial autocorrelation among plots along a pre-defined gradient. In summary, for a predefined gradient consisting of N sequentially ordered plots, a directional species accumulation curve can be generated by sequentially adding plots from one end of the gradient and calculating the cumulative species richness from plot 1 to N , acknowledging that diversity metrics are influenced by the plot order. Accounting for the spatial structure of the data is of primary importance for unbiasedly determining the structure and composition of species assemblages, as it directly affects the shape and values of the rarefaction curve (Bacaro et al., 2016; Chiarucci et al., 2009). Therefore, spatially-explicit rarefaction methods that account for spatial autocorrelation should be routinely used in diversity analyses to avoid spurious results (Kühn, 2007). Practically speaking, spatially-explicit rarefactions tend to show a decrease in species accumulation with spatial distance between samples, and the difference between the spatially-explicit curve and its non-spatially explicit counterpart depend on the degree of spatial autocorrelation in the data (Ricotta et al., 2019).

Over the past decade, the theoretical and computational foundations of rarefaction have expanded substantially, particularly following the development of unified interpolation–extrapolation frameworks. The introduction of Hill numbers as a coherent family of diversity measures (Chao et al., 2014) marked a key step forward, with subsequent extensions incorporating species abundances, sample completeness, phylogenetic structure, and β -diversity standardisation (Chao et al., 2015; Zou and Axmacher, 2020; Zou et al., 2025). These advances have been operationalised in widely used R packages, most notably *iNEXT* (Hsieh et al., 2016), which enabled rigorous comparisons of diversity across assemblages with unequal sampling effort, as well as in classic non-spatial implementations such as *specaccum* and *rarefy* in *vegan* (Oksanen et al., 2018) and the *collector.curve* function in the package *rtk* (Saary et al., 2017). Dedicated routines have further extended rarefactions to functional diversity (e.g., *rare.Rao* in *adiv*; Pavoine, 2020), while recent developments introduced parametric estimators of total expected richness using rarefaction-derived asymptotic models (Zou et al., 2023) and abundance-based solutions for α - and β -diversity standardisation under incomplete sampling (e.g. *rarestR*; Zou et al., 2025).

Despite this substantial methodological progress, current rarefaction frameworks remain mostly non-spatial. Most approaches treat sampling units as fully exchangeable, ignoring the intrinsic spatial structure and autocorrelation that characterize ecological assemblages. Even in cases where spatial heterogeneity is explicitly acknowledged—such as plot-aggregation rarefaction used to harmonise species–area relationships across forest inventories (Portier et al., 2022)—spatial information enters only indirectly as a statistical ordering rather than being embedded in the rarefaction process itself. A few tools allow limited spatial ordering of samples (e.g., proximity-based accumulation in *mobr*; McGlinn et al., 2019, 2021), but they do not provide a general framework for rarefying arbitrary taxonomic, functional, or phylogenetic diversity metrics under user-defined spatial constraints. Similarly, metagenomic implementations (e.g., *Metaxa2* Diversity Tools; Bengtsson-Palme et al., 2016) and recent computational advances, such as Bayesian estimators of rarefaction curves (Baek and Park, 2022), focus on estimation accuracy while still assuming spatial independence among

samples. This represents a critical limitation, as spatial autocorrelation is one of the dominant features of ecological data and profoundly affects both accumulation curves and estimates of sample completeness.

Aiming to provide a platform for both established and new rarefaction methods, we developed the *Rarefy* package, the first R package dedicated to gathering complementary functions for calculating rarefactions. *Rarefy* allows spatially constrained, sample-based rarefaction curves for taxonomic, functional, and phylogenetic diversity, and by allowing users to plug in custom diversity functions beyond the default indices. The current version of the *Rarefy* package displays an extensive collection of existing rarefaction approaches and proposes four new spatially-explicit methods to calculate rarefaction curves using various indices of taxonomic, functional and phylogenetic diversity.

Rarefy therefore fills a critical methodological gap by integrating spatial information directly into the rarefaction process and by generalizing rarefaction far beyond species richness. Through functions that operate on abundance matrices, functional and phylogenetic distance matrices, or user-defined diversity metrics, *Rarefy* provides a unified computational infrastructure for spatially explicit and distance-based rarefaction. This makes it possible to quantify how spatial autocorrelation shapes α - and β -diversity accumulation, to compare spatially constrained and unconstrained curves, and to extend rarefaction analysis to any dimension of biodiversity. As such, *Rarefy* complements and extends existing rarefaction frameworks, providing ecologists with a flexible, modern tool that reflects the current need for spatially aware, functionally integrative, and reproducible biodiversity analyses.

2. Generalization of spatially-explicit rarefaction to any plot-to-plot dissimilarity measure

One of the main issues in calculating a diversity index is the extent of the study area. A larger study area implies more environmental heterogeneity which increases proportionally to the distances among samples (Bacaro et al., 2012; Chiarucci et al., 2009; Nekola and White, 1999). Classically, a method to characterize heterogeneity in the diversity of ecological communities is represented by rarefaction curves, namely a descriptive curve providing a level of diversity (usually species richness) as a function of an increasing number of sampling units. Generally, this curve is calculated by randomly resampling the samples without replacement and by further averaging the accumulation curves derived by each randomization (sample-based rarefaction) or, as an alternative, by using an analytic formulation. For instance, an analytic formulation of the rarefaction curve using species richness was first derived by Shinozaki (1963), then independently rediscovered by several authors in the 1970s and again in the 2000s (see Chiarucci et al., 2008 for a literature review).

However, these methods do not consider the extent and environmental heterogeneity of the study area, an issue first proposed by Chiarucci et al. (2009) and later addressed by Bacaro et al. (2012). Indeed, species are almost never randomly distributed, either due to heterogeneity of environmental factors or to non-random dispersal mechanisms and, thus, autocorrelation in species composition (and other relevant diversity aspects) should be adequately evaluated using directional curves (sensu Ricotta et al., 2019). Spatial, temporal or environmental gradients can be taken into account when directional curves are calculated, however, for the sake of simplicity, we will mainly refer to spatial-explicit directional curves in the remains of the article.

The calculation of a spatially-explicit rarefaction cannot be derived theoretically and mathematically since each accumulation curve, on which is based the calculation of the spatially-explicit rarefaction curve, is case-specific and depends on the distribution of samples along a user-defined gradient (Ricotta et al., 2019). In *Rarefy*, the general approach for the calculation of taxonomic spatially-explicit rarefaction curve has thus been generalized to any measure of biodiversity (e.g., functional and phylogenetic diversities as provided by default in the *Rarefy* package) as follows (Fig. 1):

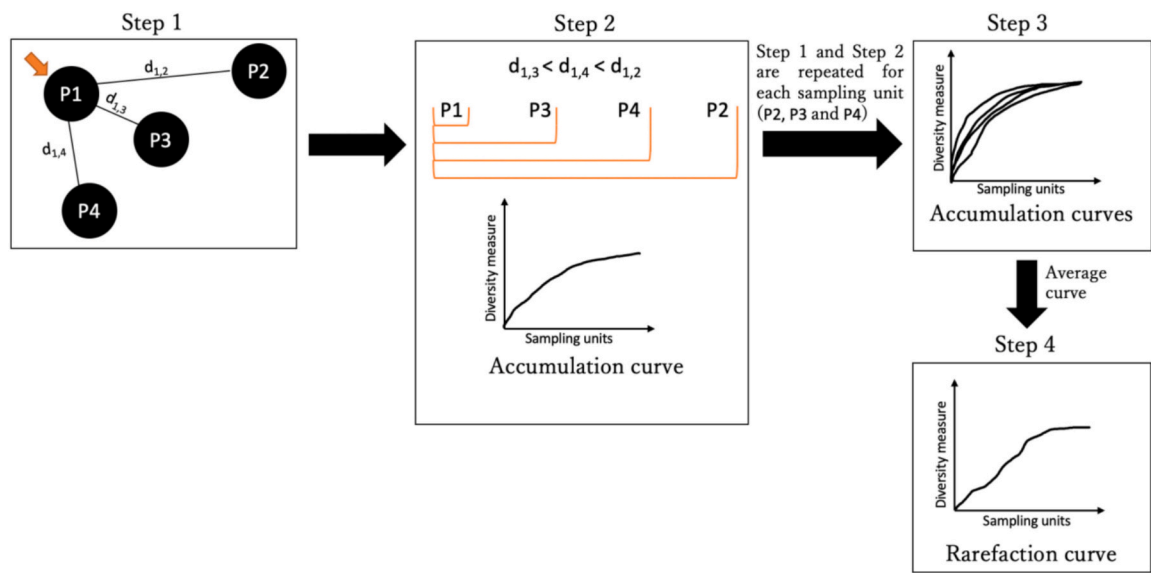


Fig. 1. Flowchart of the general framework implemented in *Rarefy* to calculate distance-based rarefaction curves. Four steps in the generation of distance-based rarefaction curves are defined: in step 1) any sample-based dissimilarity measure can be computed among the pairwise set of samples. Based on pairwise ordination on distance, an accumulation curve (Collector's curve) is calculated for each plot out of the N plots in step 2) where sampling units are aggregated based on their closeness. Finally, the spatially-explicit rarefaction is then calculated as the mean of the N accumulation curves obtained (steps 3 and 4).

1) for each sample in the pool of the N sampling units, the other samples from 2 to N are ordered based on increasing values of the selected metric of spatial distance;

2) the resulting biodiversity measure is then calculated for the set of 1, 2, 3, ..., N sampling units, giving origin to a “spatial-explicit accumulation curve” (that can be seen as a “Collector's curve” using a different notation as in [Gotelli and Colwell, 2001](#));

3) the process is repeated, starting the accumulation process from each sampling unit in the assemblage, and the rarefaction is then calculated as the mean of the N accumulation curves obtained.

The described approach can be generalized for any sample-based dissimilarity measure by choosing the preferred metric of dissimilarity in step 1 above such as those available in the *Rarefy* package (or imported by the user according to their needs): this gives origin to a general rarefaction curve in which the expected values of the selected diversity indices incorporate the autocorrelation pattern in the chosen dissimilarity metric. In addition, any type of diversity index can be used in step 2. In this way, a phylogenetic- or functional-explicit rarefaction curve (based on the diversity of the phylogenetic or functional dissimilarity matrix across species within sampling units) can be developed.

3. Generalization of rarefaction over any diversity metric

The *Rarefy* package enables users to rarefy any biodiversity metric using the *design_rare()* function. This function offers flexibility in selecting a custom metric, either by passing a user-defined or existing function (such as those from the *adiv* package; [Pavoine, 2020](#)) via the *fun_div* method or by providing the metric's equation as a string using the equation method. In the worked example, we demonstrate a simple analysis using the Shannon index with the equation approach and the functional [Pavoine et al. \(2017\)](#) R Entropy with the *fun_div* approach.

4. Description of the functions available in the *Rarefy* package

The whole list of diversity metrics proposed in *Rarefy* v2.0.0 is described in [Table 1](#) (further indices might be included in upcoming package releases) and in the related ‘*vignette*’. Indeed, *Rarefy* is designed to compile the most advanced rarefaction methods developed to date, positioning itself as a comprehensive repository for future functions

Table 1
Features of diversity components calculated by *Rarefy* functions, including a brief description of technical function requirements.

	Level	P/A vs. abund.*	Dis vs. tree†	Random or directional	Solution
<i>directionalSAC</i>	α & β	P/A	—	Both	Analytical
<i>rare_alpha</i>	α	Abund.	—	Both	Permutations
<i>rare_beta</i>	β	Both	—	Both	Permutations
<i>rareexpected_fun</i>	α	P/A	Dis	Random	Analytical
<i>rao_permuted</i>	α	Abund.	Dis	Random	Permutations
<i>ser_functional</i>	α	Abund.	Dis	Directional	Permutations
<i>rare_phylo</i>	α	Both	Tree	Random	Permutations
<i>ser_phylo</i>	α	Both	Tree	Directional	Permutations

* P/A = presence/absence; Abund = abundance data.
† functional or phylogenetic characteristics are entered in the function as a matrix of dissimilarity between species (“Dis”) or as a tree (a functional dendrogram or a phylogenetic tree, “Tree”).

related to this approach.
Specifically, brand-new functions (in bold) and already published functions not included in any package, are presented below while their main characteristics are detailed in the ‘*vignette*’ document enclosed in the *Rarefy* package:

- *directionalSAC* calculates directional and non-directional rarefaction curves of species richness (α -diversity) and species turnover (β -diversity) using the ratio of global [over all samples] species richness to local [within-sample] species richness over accumulated samples ([Ricotta et al., 2019](#)).
- *rare_alpha* calculates spatially and non-spatially-explicit rarefaction curves of species-based α diversity using Tsallis entropy ([Tsallis, 1988](#)) (also known as HCDT entropy), or the Hill numbers ([Hill, 1973](#)).
- *rare_beta* calculates spatially and non-spatially explicit rarefaction of species-based β -diversity, using Whittaker's species turnover, Jaccard dissimilarity coefficient, Bray-Curtis dissimilarity, or Cody index ([Koleff et al., 2003](#)). In case of Bray-Curtis and Jaccard dissimilarities, the accumulation curve is calculated as the mean of pairwise dissimilarities among sampling units.

- *rarexpected_fun* calculates an expected functional diversity rarefaction, by replacing the abundance of a given species with the probability that this species is found at least once in a set of n samples out of N (=relative contribution of the given species to the richness of the set of n samples) (Ricotta et al., 2010). As a functional diversity index, the function employs Rao's quadratic entropy (hereafter Rao's Q, Rao, 1982), which in the context of functional dissimilarities, measures the average functional distance between any two randomly selected species, weighted by their abundance.
- *ser_functional* calculates spatially explicit functional rarefaction using Rao's Q, or Chao et al. (2019) functional dissimilarity (FD) index. The function also allows for the calculation of classic (non-spatially explicit) functional rarefaction, as described by Ricotta et al. (2012).
- *rare_phylo* calculates non spatially-explicit phylogenetic rarefaction using one of the following indices of phylogenetic diversity: Faith's phylogenetic diversity (Faith, 1992), weighted phylogenetic diversity (Barker, 2002), Pavoine-Love-Bonsall index (Pavoine et al., 2009a), or feature richness indexes (Pavoine and Ricotta, 2019).
- *ser_phylo* calculates spatially-explicit phylogenetic rarefaction using the same indexes as *rare_phylo*.

5. Available dataset and worked example

The package also contains three datasets analyzed for the first time in Tordoni et al. (2018, 2019), who demonstrated that the plant assemblages of native species in coastal ecosystems are less functionally diverse than alien species, and that small-scale functional homogenization in the assemblage is driven by alien species. Data were collected in the northern Adriatic coastline using squared plots of 16 m² ($n = 128$) organized in 9 transects arranged along a sea–inland gradient. A set of 15 functional traits were also measured on more abundant species (for details about the functional traits in the dataset, see Tordoni et al., 2019).

The whole dataset, including plant community composition data, species functional traits and coordinates of the samples, is fully available in the *Rarefy* package and has been arranged in three list objects as follows:

1) a list, named *duneFVG*, of three independent objects of class *data.frame* (the first one for 128 plots $\times$ all 71 species; the second for 127 plots $\times$ 62 native species, the last one for 106 plots $\times$ 9 alien species) where entries represent species abundance per plot expressed as cover percentage.

2) a list, named *duneFVG.xy*, of three objects of class *data.frame* provides the spatial coordinates of plots for the three datasets described above.

3) a list, named *duneFVG.tr8*, of three objects of class *data.frame* reports the functional traits of the species (species as rows, traits as columns) for the total, native and alien groups.

6. Worked example

The package *Rarefy* can be loaded in the R console using the following commands:

```
install.packages("devtools")
```

```
library(devtools)
install_github("elith95/Rarefy")
library("Rarefy")
```

The 'Vignette' of the package can be obtained using:

```
browseVignettes("Rarefy")
```

Here, we provide an example of a typical R session using the function *ser_functional* to calculate a spatially-explicit functional rarefaction curve, proposed here for the first time. This curve represents an extension of the standard functional rarefaction for species abundance data proposed in Ricotta et al. (2012), where the spatial component (Euclidean distance among plots) is used to calculate the expected value of the functional Rao's Q as a function of the cumulated number of plots, taking into account the similarity in species composition due to the sampling unit proximity. The proposed example uses the data freely available in the package as indicated above.

As a first operation, the *duneFVG* data are loaded into the R workspace, along with the *ade4* and *adiv* packages:

```
require(ade4)
require(adiv)
data(duneFVG)
data(duneFVG.xy)
data(duneFVG.tr8)
```

A functional distance matrix is then calculated among species using species functional traits. Considering the different nature of these variables (ordinal, circular, qualitative and continuous) the modified Gower distance measure (Pavoine et al., 2009b) is calculated using the function *dist.ktab* available in the *ade4* package (Dray et al., 2007):

```
tr8_N <- duneFVG.tr8$traits.tot[c(1,3,4)]
tr8_D <- data.frame(duneFVG.tr8$traits.tot[,2])
tr8_Q <- duneFVG.tr8$traits.tot[,5:15]
tr8dist <- dist.ktab(ktab.list.df(list(tr8_N, tr8_D, tr8_Q)),
                    type = c('N', 'D', 'Q'))
```

A plot-to-plot Euclidean distance matrix is then calculated with the *dist* function (available in the basic *stats* package) and using the plot coordinates as input:

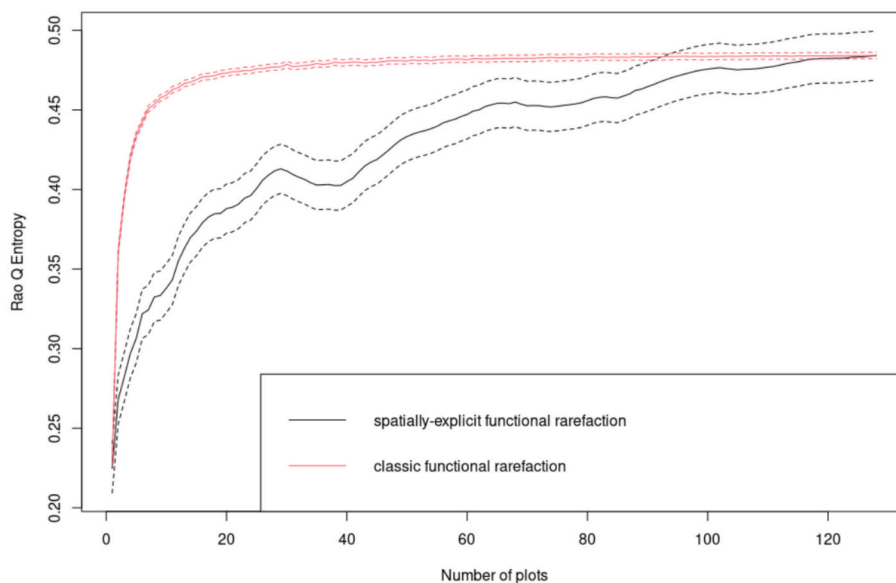
```
dist_sp <- dist(duneFVG.xy$tot.xy, method = "euclidean")
```

Now, we calculate the spatially-explicit functional rarefaction using Rao (1982) Q index:

```
rare_Q <- ser_functional(duneFVG$total, tr8dist, dist_sp, method='rao',
                        comparison=TRUE)
```

By setting the argument *comparison = TRUE*, a classic functional rarefaction curve is also calculated, and its values are provided in the

a)



b)

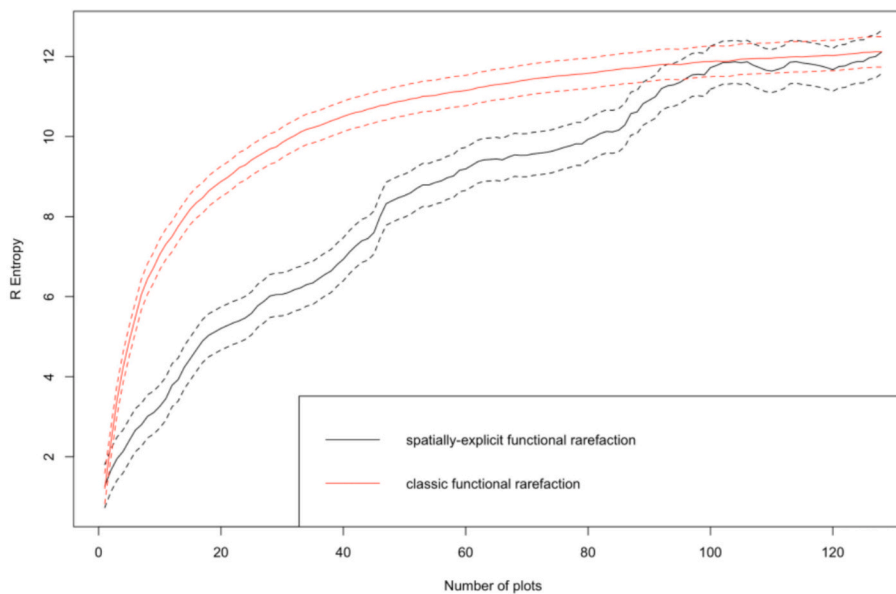


Fig. 2. Graphical comparison between spatially-explicit and classic functional rarefaction curves calculated using the *duneFVG* dataset available in the *Rarefy* package and two selected functional diversity indices: a) Rao (1982) Q index b) Pavoine et al. (2017) R Entropy. 95% confidence intervals are also included for each curve (black and red dashed lines). In both cases, spatially-explicit rarefactions and their non-spatially explicit counterparts are not coincident (with the spatially explicit curve having lower values for the same sampling effort). The degree of separation of the two curves depends on the degree of spatial autocorrelation in the data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

output for direct comparison with the spatial-explicit counterpart.

Lastly, we can build plots to visualize the computed rarefaction curves (Fig. 2a):

```
plot(rare_Q[,1], ylab = "Q Entropy", xlab = "Number of plots", type = "l", ylim=range(rare_Q, na.rm = TRUE))
lines(rare_Q[,2], lty = 2)
lines(rare_Q[,3], lty = 2)
lines(rare_Q[,4], col = 2)
lines(rare_Q[,5], lty = 2, col = 2)
lines(rare_Q[,6], lty = 2, col = 2)
legend("bottomright", legend=c("spatially-explicit functional rarefaction",
    "classic functional rarefaction"),
    lty = 1, col = 1:2)
```

Also, to familiarize the user to the “*fun_div*” method of the function *design_rare()*, we provided a direct comparison of the just computed spatially-explicit Rao's Q functional curve with the rarefaction of the closely related Pavoine et al. (2017) Entropy index, available in the

function *Rentropy* of the *adiv* package (Fig. 2b). First, a list is created to store the values required by the *Rentropy()* function. In this list, the community matrix should be replaced with NA, and the element names must match the argument names expected by *Rentropy()*. Next, the function name and the list of arguments are passed to *design_rare()* through the *fun* and *args* parameters, respectively:

```
#List of the values for Rentropy()
a <- list(NA, tr8dist)

#Assign to the list names matching the arguments of Rentropy()
names(a) <- c('comm','dis')

#Spatially-explicit and classic rarefaction curves
ser_R <- design_rare(duneFVG$total, dist_sp, method = "fun_div", args = a,
    fun = "Rentropy", spatial = TRUE)
rare_R <- design_rare(duneFVG$total, dist_sp, method = "fun_div", args = a,
    fun = "Rentropy", spatial = FALSE)

#Plot
plot(rare_R[,1], ylab = "R Entropy", xlab = "Number of plots", type = "l", ylim=range(rare_R, na.rm = TRUE))
lines(rare_R[,2], lty = 2)
lines(rare_R[,3], lty = 2)
lines(ser_R[,1], col = 2)
lines(ser_R[,2], lty = 2, col = 2)
lines(ser_R[,3], lty = 2, col = 2)
legend("bottomright", legend=c("spatially-explicit functional rarefaction",
    "classic functional rarefaction"), lty = 1, col = 1:2)
```

We provide below another example with the same data set to show how to extend rarefaction to any metric as decided by the user. Specifically, a Shannon-Wiener rarefaction curve (as provided, for instance, in [Li and Zao, 2015](#)) was calculated by using the “equation” method provided by the function `design_rare()`. The Shannon index is calculated as the negative sum, taken over all species in the community, of the proportional abundances p of each species multiplied by their natural logarithm:

```
#Formula

eq <- "-sum(p[p>0] * log(p[p>0]))"

#List of terms of the equation

#(p values are extracted from the community matrix)

a <- list(NA)

#Set the names of the list to match the formula terms

names(a) <- c('p')

#Rarefaction

rare_H <- design_rare(duneFVG$total, dist_sp, method = "equation", args = a,
                      formula = eq, spatial = TRUE, mean = TRUE)
```

7. Discussion

The package *Rarefy* implements spatially explicit rarefaction curves by constructing distance-constrained accumulation trajectories, thereby making spatial structure an explicit component of the standardisation process. In this way, *Rarefy* complements established non-spatial frameworks (e.g., those available in *iNEXT* and *vegan*) by addressing a common situation in ecological surveys: sampling units are spatially structured and therefore not exchangeable under random permutations. This is particularly relevant when rarefaction is used to compare sites or sampling designs across landscapes, where spatial autocorrelation and environmental heterogeneity strongly shape the rate at which diversity accumulates. *Rarefy* also allows both α - and β -diversity components to be analyzed. R packages such as the recent *rarestR* ([Zou et al., 2025](#)) have also focused on rarefaction-based estimation of diversity under incomplete sampling, including tools targeting α - and β -diversity comparability across uneven datasets. However, *Rarefy* differs from these approaches in its primary objective and workflow: rather than estimating diversity under incomplete sampling in general, it provides a flexible framework to compute spatially constrained, sample-based rarefaction for a wide range of diversity metrics, including functional and phylogenetic indices, under user-defined distance/order constraints. This flexibility is implemented through functions that allow custom diversity definitions, enabling users to extend the approach beyond the default indices provided in the package. From a practical perspective, *Rarefy* is intended to complement rather than replace established non-spatial frameworks. When sampling units can reasonably be treated as exchangeable, non-spatial rarefaction and extrapolation tools remain appropriate and often more efficient. By contrast, *Rarefy* is most informative when sampling units are spatially structured or clustered and users want rarefaction curves that explicitly respect distance-constrained accumulation. In this setting, *Rarefy* supports

taxonomic, functional, and phylogenetic diversity and can also accommodate user-defined diversity metrics, thereby extending spatially explicit rarefaction beyond species richness. *Rarefy* is under active development, and future package releases will incorporate additional functionality to further broaden its scope and facilitate interoperability with other common rarefaction workflows. In its current version, the most innovative (and flexible) characteristic of *Rarefy* lies in the user's ability to apply spatial/non-spatial rarefaction to any diversity measure

by computing the value of a chosen metric and then rarefying it in relation to the sampling effort, using the brand-new function `design_rare()`.

Spatial patterns are increasingly recognized as a fundamental component of biodiversity analyses, with consequences for both α - and β -diversity estimation ([Legendre, 1993](#); [Fortin and Dale, 2005](#)). When samples are spatially aggregated or environmentally structured, classical rarefaction tends to overestimate the rate at which diversity accumulates, because it implicitly assumes that sampling units are exchangeable ([Chiarucci et al., 2009](#)). By explicitly incorporating spatial distances, *Rarefy* makes these deviations directly observable. The comparison between classical and spatially constrained curves provides a practical diagnostic of spatial autocorrelation, consistent with the idea that species turnover, dispersal limitation and environmental filtering shape the spatial distribution of diversity ([Sandel, 2015](#); [McGill, 2019](#)). For example, in our worked example, spatially explicit rarefaction produced lower values than non-spatial rarefaction, which is consistent with the expectation that spatially close plots may share many species and traits (higher functional redundancy). However, differences between spatially constrained and non-spatial curves should be interpreted as a diagnostic signal rather than as a direct “measurement” of spatial autocorrelation. In practice, the spatial–non-spatial gap is conditional on three elements that should be reported and checked explicitly: (i) the distance metric used to order sampling units (e.g., Euclidean distance, cost-distance, or an environmental distance when appropriate), because different metrics capture different notions of “closeness”; (ii) landscape structure and habitat mosaics, which can generate non-linear turnover across space and therefore change the expected separation between curves; and (iii) sampling grain and plot size, because accumulation rates and the visibility of spatial structure depend on the scale at which samples are collected. For instance, in strongly patchy systems or along sharp steep gradients, spatial ordering may expose rapid turnover at short distances, potentially altering the relative behaviour of curves. We therefore recommend interpreting curve differences together with explicit information on the sampling design and landscape structure,

and, when relevant, complementary spatial analyses.

The extension of rarefaction to functional and phylogenetic dissimilarity matrices also allows *Rarefy* to capture multiple dimensions of diversity, building on established approaches for quantifying functional and phylogenetic similarity among communities (Pavoine and Ricotta, 2014; Ricotta et al., 2020). Functional and phylogenetic rarefaction curves have been used to investigate trait turnover along environmental gradients (Swenson, 2014), phylogenetic clustering in fragmented landscapes (Cadotte et al., 2010) and the scale-dependence of ecosystem functioning (Lepš et al., 2006). By integrating spatial constraints, *Rarefy* offers a way to evaluate whether these patterns stem primarily from ecological processes operating in space or from differences in sampling effort. This is particularly relevant in systems where environmental heterogeneity is strong or where functional redundancy may obscure spatial structuring.

Building on this framework, a promising avenue for future development involves integrating null models into rarefaction calculations, allowing ecologists to compare observed diversity patterns with those generated by randomized communities that retain the same taxonomic, functional, or phylogenetic characteristics for a given sampling effort. Null-model approaches have long been used to disentangle the effects of species richness, abundance structure and trait composition (Gotelli and Colwell, 2001; Chase et al., 2011). In the context of functional diversity, randomization procedures help distinguish deterministic processes such as environmental filtering from stochastic or neutral processes (Mason et al., 2011; Kraft et al., 2015). The *rao.permuted* function implements this logic within a rarefaction framework. Unlike *rarexpected.fun*, it enables comparisons between observed and expected functional diversity across sampling gradients. This approach complements developments applying rarefaction to trait-based analyses (Ricotta et al., 2012; Chiu and Chao, 2014) and aligns with the growing interest in simulation-based tools within ecological informatics.

This function was useful to test a null model of functional trait reallocation based on a random selection of a given number of s species out of S available species in a pool and to use a standardized community structure (composed by s species but with a specified abundance structure) to calculate the expected functional diversity for a given sampling effort using the Rao Q diversity index. A detailed example of function use and its application, applied for the first time in Tordoni et al. (2019), is described in the *Rarefy* ‘vignette’ and in Thouverai et al. (2026). In brief, the function works in three stages: 1) from a total set of S species sampled in M sampling units (i.e. regional species pool), a user-defined number of s species (with $s < S$) is randomly selected; 2) a functional dissimilarity matrix for the set of the s randomly selected species is calculated using their associated functional traits and then the rarefied Rao Q index is calculated for the set of M sampling units considering abundance of a reference plant community (the structure of the reference community can be passed to the function via the “*comm_str*” argument). 3) For each permutation, step 1 and 2 are repeated and the functional rarefaction curves are then averaged over the number of permutations. The functional rarefaction curve thus obtained can be compared to the functional rarefaction curve calculated for the reference community, and the deviation from the two curves (when positive or negative) can be evaluated as an effect of environmental filtering, niche differentiation, or biotic homogenization, alternatively.

Discussing the applicability and the limitations of *Rarefy*, we would like to recall that its applications are used to standardize diversity under explicit spatial constraints, but they are not designed to remove all sources of bias that affect biodiversity estimates. First, scale and plot-size effects can influence richness and other diversity metrics; when datasets combine inventories collected at different grains or with different plot areas, users should harmonise sampling units or adopt stratified analyses to ensure comparability (see also plot-size comparability issues discussed by Portier et al., 2022). Second, computational cost can increase rapidly with the number of sampling units because spatially constrained trajectories require repeated ordering and

resampling; for a very large number of plots N , users may need to reduce dimensionality (e.g., by aggregating nearby plots, limiting the range of distances, or running analyses by spatial blocks). Third, the output remains conditional on the input community matrix and therefore inherits uncertainties related to detectability and observation error, which may be particularly relevant in low-detectability settings.

In species-rich and heterogeneous ecosystems, where beta-diversity is high and community composition changes across short distances, spatially constrained rarefaction is informative but should be applied with attention to habitat stratification. Similarly, when strong seasonal or interannual turnover is expected, we recommend conducting separate analyses by season (or other relevant temporal strata) rather than pooling samples across time, so that rarefaction curves summarize spatial structure without confounding temporal dynamics. Finally, in extreme environments or datasets dominated by low detectability, *Rarefy* should be used as a standardisation tool within a broader analytical workflow, potentially combined with repeated sampling or other methods that explicitly address detection limits.

Future development of the package will include a spatially-explicit version of *rare.permuted* and, in general, opening new ways towards a combination of rarefaction methods with different types of null model simulations in community ecology. Several opportunities exist for extending this framework. Spatially explicit null models, which combine randomization with spatially structured permutations, could improve tests of dispersal limitation or spatial filtering (Wiegand et al., 2013). Phylogenetic randomizations could be used to explore the sensitivity of PD rarefaction to alternative evolutionary hypotheses (Molina-Venegas and Rodríguez, 2017). Moreover, integrating *Rarefy* with large-scale biodiversity monitoring schemes would enable the application of rarefaction methods to dynamic datasets, where spatial replication and repeated sampling are central elements. These extensions are consistent with current trends in ecological informatics, which increasingly emphasize reproducibility, integration across data types and explicit treatment of spatial structure.

Citation

Researchers using *Rarefy* in a published paper can cite this article and in addition can also cite the *Rarefy* package directly. Updated citation information can be obtained by typing *citation('Rarefy')* in R.

CRedit authorship contribution statement

Elisa Thouverai: Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis. **Sandrine Pavoine:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Enrico Tordoni:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Data curation. **Alessandro Chiarucci:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Carlo Ricotta:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Duccio Rocchini:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Giovanni Bacaro:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank all the contributors and testers who helped us develop the

methods and data sets used here. All data used in the manuscript are included in the R package published at the following address: <https://github.com/elith95/Rarefy>

Data availability

The data used are freely available within the *Rarefy* package. The code used is also provided in the main text and in the specific ‘vignette’ in the *Rarefy* package.

References

- Bacaro, G., Rocchini, D., Ghisla, A., Marcantonio, M., Neteler, M., Chiarucci, A., 2012. The spatial domain matters: spatially constrained species rarefaction in a free and open source environment. *Ecol. Complex.* 12, 63–69. <https://doi.org/10.1016/j.ecocom.2012.05.007>.
- Bacaro, G., Altobelli, A., Cameletti, M., Ciccarelli, D., Martellos, S., Palmer, M.W., Chiarucci, A., 2016. Incorporating spatial autocorrelation in rarefaction methods: implications for ecologists and conservation biologists. *Ecol. Indic.* 69, 233–238. <https://doi.org/10.1016/j.ecolind.2016.04.026>.
- Baek, S., Park, J., 2022. A computationally efficient approach to estimating species richness and rarefaction curve. *Comput. Stat.* 37, 1919–1941. <https://doi.org/10.1007/s00180-021-01185-1>.
- Barker, G.M., 2002. Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation. *Biol. J. Linn. Soc.* 76, 165–194. <https://doi.org/10.1111/j.1095-8312.2002.tb02081.x>.
- Bengtsson-Palme, J., Thorell, K., Wurzbacher, C., Sjöling, Å., Nilsson, R.H., 2016. Metaxa2 diversity tools: easing microbial community analysis with Metaxa2. *Ecol. Inform.* 33, 45–50. <https://doi.org/10.1016/j.ecoinf.2016.04.004>.
- Chao, A., Jost, L., 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93, 2533–2547. <https://doi.org/10.1890/11-1952.1>.
- Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., Ellison, A.M., 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* 84, 45–67. <https://doi.org/10.1890/13-0133.1>.
- Chao, A., Chiu, C.H., Hsieh, T.C., Davis, T., Nipperess, D.A., Faith, D.P., 2015. Rarefaction and extrapolation of phylogenetic diversity. *Methods Ecol. Evol.* 6, 380–388. <https://doi.org/10.1111/2041-210X.12247>.
- Chao, A., Chiu, C.-H., Villéger, S., Sun, I.-F., Thorn, S., Lin, Y.-C., Sherwin, W.B., 2019. An attribute-diversity approach to functional diversity, functional beta diversity, and related (dis)similarity measures. *Ecol. Monogr.* 89 (2), e01343. <https://doi.org/10.1002/ecm.1343>.
- Chase, J.M., Kraft, N.J.B., Smith, K.G., Vellend, M., Inouye, B.D., 2011. Using null models to disentangle variation in community dissimilarity from variation in α -diversity. *Ecosphere* 2 (2), 1–11. <https://doi.org/10.1890/ES10-00117.1>.
- Chiarucci, A., Bacaro, G., Rocchini, D., Fattorini, L., 2008. Discovering and rediscovering the sample-based rarefaction formula in the ecological literature. *Community Ecol.* 9, 121–123. <https://doi.org/10.1556/ComEc.9.2008.1.14>.
- Chiarucci, A., Bacaro, G., Rocchini, D., Ricotta, C., Palmer, M., Scheiner, S., 2009. Spatially constrained rarefaction: incorporating the autocorrelated structure of biological communities into sample-based rarefaction. *Community Ecol.* 10 (2), 209–214. <https://doi.org/10.1556/ComEc.10.2009.2.11>.
- Colwell, R.K., Mao, C.X., Chang, J., 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85, 2717–2727. <https://doi.org/10.1890/03-0557>.
- Chiu, C.-H., Chao, A., 2014. Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers. *PLOS One* 9 (7), e100014. <https://doi.org/10.1371/journal.pone.0100014>.
- Colwell, R.K., Chao, A., Gotelli, N.J., Lin, S.-Y., Mao, C.X., Chazdon, R.L., Longino, John T., 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *J. Plant Ecol.* 5 (1), 3–21. <https://doi.org/10.1093/jpe/rtr044>.
- Dray, S., Dufour, A., & Chessel, D. (2007) The ade4 package - II: two-table and K-table methods. *R News*, 7(2), 47–52. <https://cran.r-project.org/web/packages/ade4/index.html>.
- Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61, 1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3).
- Fortin, M.-J., Dale, M.R.T., 2005. *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge.
- Gotelli, N.J., Colwell, R.K., 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* 4, 379–391. <https://doi.org/10.1046/j.1461-0248.2001.00230.x>.
- Hill, M.O., 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427–432. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3).
- Hortal, J., De Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M., Ladle, R.J., 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* 46, 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- Hsieh, T.C., Ma, K.H., Chao, A., 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* 7, 1451–1456.
- Koleff, P., Gaston, K.J., Lennon, J.J., 2003. Measuring beta diversity for presence-absence data. *J. Anim. Ecol.* 72, 367–382. <https://doi.org/10.1046/j.1365-2656.2003.00710.x>.
- Kraft, N.J.B., Adler, P.B., Godoy, O., James, E.C., Fuller, S., Levine, J.M., 2015. Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* 29 (5), 592–599. <https://doi.org/10.1111/1365-2435.12345>.
- Kühn, I., 2007. Incorporating spatial autocorrelation may invert observed patterns. *Divers. Distrib.* 13, 66–69. <https://doi.org/10.1111/j.1472-4642.2006.00293.x>.
- Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74 (6), 1659–1673. <https://doi.org/10.2307/1939924>.
- Lepš, J., de Bello, F., Lavorel, S., Berman, S., 2006. Quantifying and interpreting functional diversity of natural communities: practical considerations matter. *Preslia* 78, 481–501.
- Li, L., Zao, X., 2015. Comparative analyses of fecal microbiota in Tibetan and Chinese Han living at low or high altitude by barcoded 454 pyrosequencing. *Sci. Rep.* 5, 14682. <https://doi.org/10.1038/srep14682>.
- Marchetto, E., Livornese, M., Sabatini, F.M., Tordoni, E., Da Re, D., Lenoir, J., Rocchini, D., 2024. Addressing multiple facets of bias and uncertainty in continental scale biodiversity databases. *Biodivers. Inform.* 18, 56–77. <https://doi.org/10.17161/bi.v18i.21810>.
- Mason, N.W.H., de Bello, F., Doležal, J., Lepš, J., 2011. Niche overlap reveals the effects of competition, disturbance and contrasting assembly processes in experimental grassland communities. *J. Ecol.* 99, 788–796. <https://doi.org/10.1111/j.1365-2745.2011.01801.x>.
- McGlinn, D.J., Engel, T., Blowes, S.A., Gotelli, N.J., Knight, T.M., McGill, B.J., Sanders, N.J., 2021. A multiscale framework for disentangling the roles of evenness, density, and aggregation on diversity gradients. *Ecology* 102, e03233.
- McGlinn, D.J., Xiao, X., May, F., Gotelli, N.J., Engel, T., Blowes, S.A., Knight, T.M., Purschke, O., Chase, J.M., 2019. Measurement of Biodiversity (MoB): a method to separate the scale-dependent effects of species abundance distribution, density, and aggregation on diversity change. *Methods Ecol. Evol.* 10, 258–269.
- Molina-Venegas, R., Rodríguez, M.A., 2017. Revisiting phylogenetic signal: strong or negligible impacts of polytomies and branch length information? *BMC Evol. Biol.* 17, 53. <https://doi.org/10.1186/s12862-017-0898-y>.
- Nekola, J.C., White, P.S., 1999. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* 26, 867–878. <https://doi.org/10.1046/j.1365-2699.1999.00305.x>.
- Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Wagner, H., 2018. *Vegan: Community Ecology Package*. R package version 2.5-3. URL: <https://CRAN.R-project.org/package=vegan>.
- Pavoine, S., 2020. Adiv: an R package to analyse biodiversity in ecology. *Methods Ecol. Evol.* 11, 1106–1112. <https://doi.org/10.1111/2041-210X.13430>.
- Pavoine, S., Love, M.S., Bonsall, M.B., 2009a. Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically-structured species assemblages: application to rockfish (genus: *Sebastes*) in the Southern California Bight. *Ecol. Lett.* 12, 898–908. <https://doi.org/10.1111/j.1461-0248.2009.01344.x>.
- Pavoine, S., Ricotta, C., 2014. Functional and phylogenetic similarity among communities. *Methods Ecol. Evol.* 5, 666–675. <https://doi.org/10.1111/2041-210X.12193>.
- Pavoine, S., Ricotta, C., 2019. A simple translation from indices of species diversity to indices of phylogenetic diversity. *Ecol. Indic.* 101, 552–561. <https://doi.org/10.1016/j.ecolind.2019.01.052>.
- Pavoine, S., Vallet, J., Dufour, A.B., Gachet, S., Daniel, H., 2009b. On the challenge of treating various types of variables: application for improving the measurement of functional diversity. *Oikos* 118 (3), 391–402. <https://doi.org/10.1111/j.1600-0706.2008.16668.x>.
- Pavoine, S., Bonsall, M.B., Dupaix, A., Jacob, U., Ricotta, C., 2017. From phylogenetic to functional originality: guide through indices and new developments. *Ecol. Indic.* 82, 196–205. <https://doi.org/10.1016/j.ecolind.2017.06.056>.
- Portier, J., Zellweger, F., Zell, J., et al., 2022. Plot size matters: toward comparable species richness estimates across plot-based inventories. *Ecol. Evol.* 12, e8965. <https://doi.org/10.1002/ece3.8965>.
- Rao, C.R., 1982. Diversity and dissimilarity coefficients: a unified approach. *Theor. Popul. Biol.* 21, 24–43. [https://doi.org/10.1016/0040-5809\(82\)90004-1](https://doi.org/10.1016/0040-5809(82)90004-1).
- Ricotta, C., Burrascano, S., Blasi, C., 2010. Incorporating functional dissimilarities into sample-based rarefaction curves: from taxon resampling to functional resampling. *J. Veg. Sci.* 21, 280–286. <https://doi.org/10.1111/j.1654-1103.2009.01142.x>.
- Ricotta, C., Pavoine, S., Bacaro, G., Acosta, A.T.R., 2012. Functional rarefaction for species abundance data. *Methods Ecol. Evol.* 3, 519–525. <https://doi.org/10.1111/j.2041-210X.2011.00178.x>.
- Ricotta, C., Acosta, T.R.A., Bacaro, G., Carboni, M., Chiarucci, A., Rocchini, D., Pavoine, S., 2019. Rarefaction of beta diversity. *Ecol. Indic.* 107, 105606. <https://doi.org/10.1016/j.ecolind.2019.105606>.
- Ricotta, C., Laroche, F., Széidl, L., Pavoine, S., 2020. From alpha to beta functional and phylogenetic redundancy. *Methods Ecol. Evol.* 11, 487–493. <https://doi.org/10.1111/2041-210X.13353>.
- Saary, P., Forslund, K., Bork, P., Hildebrand, F., 2017. RTK: efficient rarefaction analysis of large datasets. *Bioinformatics* 33 (16), 2594–2595. <https://doi.org/10.1093/bioinformatics/btx206>.
- Sandel, B., 2015. Towards a taxonomy of spatial scale-dependence. *Ecography* 38 (4), 358–369. <https://doi.org/10.1111/ecog.01034>.
- Shinozaki, K., 1963. Note on the species area curve. In: *Proceedings of the 10th Annual Meeting of the Ecological Society of Japan*, pp. 1–5 (in Japanese).
- Swenson, N.G., 2014. *Functional and Phylogenetic Ecology* in R. Springer, New York. <https://doi.org/10.1007/978-1-4614-9542-0>.

- Thouverai, E., Pavoine, S., Tordoni, E., Chiarucci, A., Ricotta, C., Rocchini, D., Bacaro, G., 2026. Rarefaction curves with the rarefy package. In: Rocchini, D. (Ed.), *R Coding for Ecology, Use R!* Springer Nature Switzerland, pp. 143–158. https://doi.org/10.1007/978-3-031-99665-8_6.
- Tordoni, E., Napolitano, R., Maccherini, S., Da Re, D., Bacaro, G., 2018. Ecological drivers of plant diversity patterns in remnants coastal sand dune ecosystems along the northern Adriatic coastline. *Ecol. Res.* 33 (6), 1157–1168. <https://doi.org/10.1007/s11284-018-1629-6>.
- Tordoni, E., Petruzzellis, F., Nardini, A., Savi, T., Bacaro, G., 2019. Make it simpler: alien species decrease functional diversity of coastal plant communities. *J. Veg. Sci.* 30, 498–509. <https://doi.org/10.1111/jvs.12734>.
- Tsallis, C., 1988. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* 52, 479–487. <https://doi.org/10.1007/BF01016429>.
- Wiegand, T., He, F., Hubbell, S.P., 2013. A systematic comparison of summary characteristics for quantifying point patterns in ecology. *Ecography* 36, 92–103. <https://doi-org.units.idm.oclc.org/10.1111/j.1600-0587.2012.07361.x>.
- Zou, Y., Axmacher, J.C., 2020. The Chord-Normalized Expected Species Shared (CNESS)-distance represents a superior measure of species turnover patterns. *Methods Ecol. Evol.* 11, 273–280.
- Zou, Y., Zhao, P., 2023. Estimating total species richness: fitting rarefaction by asymptotic approximation. *Ecosphere* 14, e4363.
- Zou, Y., Zhao, P., Wu, N., et al., 2025. rarestR: an R package using rarefaction metrics to estimate α - and β -diversity for incomplete samples. *Divers. Distrib.* 31, e13954.