

# ECOGRAPHY

## Software notes

### comspat: an R package to analyze within-community spatial organization using species combinations

James L. Tsakalos, Stefano Chelli, Giandiego Campetella, Roberto Canullo, Enrico Simonetti and Sandor Bartha

J. L. Tsakalos (<https://orcid.org/0000-0001-5067-196X>) ✉ ([jamestsakalos@gmail.com](mailto:jamestsakalos@gmail.com)), S. Chelli (<https://orcid.org/0000-0001-7184-8242>), G. Campetella (<https://orcid.org/0000-0001-6126-522X>), R. Canullo (<https://orcid.org/0000-0002-9913-6981>) and E. Simonetti, School of Biosciences and Veterinary Medicine, Plant Diversity and Ecosystems Management Unit, Univ. of Camerino, Camerino, MC, Italy. JLT also at: Harry Butler Inst., Murdoch Univ., Murdoch, Perth, WA, Australia. – S. Bartha (<https://orcid.org/0000-0001-6331-7521>), Inst. of Ecology and Botany, Centre for Ecological Research, Vácrátót, Hungary.

#### Ecography

2022: e06216

doi: 10.1111/ecog.06216

Subject Editor:

Michael Krabbe Borregaard

Editor-in-Chief: Miguel Araújo

Accepted 7 March 2022



The diversity of species combinations observable in sampling units reflects a species' uneven distribution and preference for specific abiotic and biotic conditions – a phenomenon most commonly expressed in terms of ecological assembly rules of plant communities and other sessile organisms (e.g. subtidal algae, invertebrates and coral reefs). We present `comspat`, a new R package that uses grid or transect data sets to measure the number of realized (observed) species combinations (NRC) and the Shannon diversity of realized species combinations (compositional diversity; CD) as a function of spatial scale. NRC and CD represent two measures from a model family developed by Pál Juhász-Nagy based on information theory. Classical Shannon diversity measures biodiversity based on the number and relative abundance of species, whereas the specific version of Shannon diversity presented here characterizes biodiversity and provides information on species coexistence relationships; both measures operate at fine-scale within the sampling unit or within the community. `comspat` offers two commonly applied null models, complete spatial randomness and random shift, to disentangle the textural, intraspecific and interspecific effects on the observed spatial patterns. Combined, these models assist users in detecting and interpreting spatial associations and inferring assembly mechanisms. Our open-sourced package provides a vignette that describes the method and reproduces the figures from this paper to help users contextualize and apply functions to their data.

Keywords: beta diversity, community assembly rules, information theory, multi-species co-occurrence, Shannon diversity, spatial scale

#### Background

Here we introduce a bioinformatic version of Shannon diversity to analyze within-community spatial organization using species combinations. The classical Shannon diversity uses both the number and the relative abundance of species to characterize the



[www.ecography.org](http://www.ecography.org)

© 2022 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

biodiversity of a sample or a community (Magurran 2004). Whereas the bioinformatic version of Shannon's diversity of species combinations (compositional diversity; CD) uses the number and relative abundance of realized (observed) species combinations (NRC) to characterize the biodiversity and provide information on species coexistence relationships at fine-scale within the sampling unit or within the community. The variability of species combinations are important because they reflect a species' uneven distribution and preference for specific abiotic and biotic conditions (Wilson 1999). Communities are usually characterized by the overall species composition and mean species abundances, yet relatively little is known about how the fine-scale patterns contribute to these general patterns (Chase and Knight 2013, Sandel 2015, Ben-Hur and Kadmon 2020).

Let us imagine a plant community with ten species. Several patterns may emerge when we sample this community with many sampling units (plots) of a given size. In the simplest theoretical case, all sampling units can have no species or be monodominant (i.e. in both cases, NRC, is one, please see the worked examples in the package vignette). In the opposite case, all sampling units can have different species combinations ( $NRC = 2^s = 2^{10} = 1024$ ). Different NRCs (i.e. ranging from 1 to 1024) appear when different spatial patterns emerge; classical Shannon diversity does not reflect differences in the fine-scale species combinations (we describe these calculations in the Supporting information). Due to this variation, the number and diversity of within-community species combinations are sensitive indicators of within-community species coexistence (Juhász-Nagy and Podani 1983, Bartha et al. 1998).

Theoretically, we could determine the optimum size of plots for detecting species combinations based on the size of individuals (or ramets) and the magnitude of scales relevant for biotic interactions, dispersal distances and environmental heterogeneity. However, these scales cannot be set as constant attributes in a community as they may vary by species, ontogenetic stages and many other factors. Instead, ecologists should apply a series of increasing plots sizes (Dengler et al. 2018) to deal with the scale dependence of species combinations (Chase and Knight 2013; e.g. the size of the sampling unit affects diversity estimates, Wilson et al. 2012). By plotting the NRC against the increasing plot sizes, we can visualize unimodal curves with the maxima emerging at specific plot sizes. These observed maximum values and corresponding scales express the optimum size of plots for capturing the species' local behavior and overall relationship (i.e. see characteristic maximum scales in Juhász-Nagy 1967b, Juhász-Nagy and Podani 1983).

To analyze the within-community patterns of the number and diversity of species combinations, Pál Juhász-Nagy (Juhász-Nagy 1967a, 1976, 1984a, b, 1993, Juhász-Nagy and Podani 1983) developed several diversity measures including information about the spatial structure of the data based on information theory. Juhász-Nagy Pál's models (JNP models) provide a suite of diversity measures designed for representing the spatial dependence of populations within a community

and have been thoroughly tested and verified in the past (Podani et al. 1993, Bartha et al. 1998). In the absence of rigorous sampling and experimentation, the detailed analyses of the patterns of within-community species combinations compared with null models can potentially reveal the rules and history of the local assembly (Juhász-Nagy and Podani 1983, Bartha et al. 1998). For example, they have been implicated in the hunt for assembly rules (Bartha et al. 1998), applied and led to the improvement of restoration practices after open cut mining (Bartha 1992) and detecting compositional changes along ecological gradients (Bartha et al. 2011, 2020).

The R environment is rich in contributed packages (ca 20 000 current packages, <[www.r-project.org](http://www.r-project.org)>), and several exist which analyze biodiversity with information theory. `entropy` (Hausser and Strimmer 2009), `entropart` (Marcon and Hérault 2015) and `EntropyEstimation` (Cao and Grabchak 2015) are perhaps most widely known, while recent contributions of `rasterdiv` (Rocchini et al. 2021) and `SpatEntropy` (Altieri et al. 2021) will undoubtedly gain popularity (see the total downloads in the Supporting information); all calculate the Shannon's diversity index (Shannon 1948). Although colloquially described as a measure of diversity, the Shannon index is not a measure of diversity; instead, it is an entropy estimation, giving uncertainty in the outcome of a sampling process (Jost 2006). `entropy` focuses on providing estimation-bias corrections for small samples. `entropart` focuses on partitioning entropy into alpha and beta components and providing phylogenetic and functional entropy measures and Simpson's evenness index (Simpson 1949). `EntropyEstimation` works in similar spaces to `entropy` and `entropart` but provides variants on Rényi's entropy (Rényi 1961). The latter two packages contribute by incorporating entropy while considering the spatial location of the data. `rasterdiv` is optimized to calculate entropy using remotely sensed environmental raster data and provides the functionality for users to supply a single spatial sampling scale. `SpatEntropy` computes a large set of spatial entropy measures applied to biodiversity and environmental data sets. Despite the comprehensive literature on spatial entropy measures and available R packages, to our knowledge, none estimate entropy using species combinations (an inherent property of JNP's models, cf. individual species), include null models or accept a series of user-defined spatial scales.

We present our new R package `comspat`, which is short for 'within-community spatial organization'. The main goal of `comspat` is to provide users with open access to the tools required to analyze and interpret within-community organization of vegetation across a range of spatial scales based on the spatial pattern of species combinations. Specifically, our package calculates two commonly used JNP models from grid or transect data at different spatial scales supplied by the user. The package offers null models and custom plotting to disentangle textural (i.e. the number of species combinations and their relative abundance) intraspecific and interspecific

effects on the observed spatial patterns. The package features a vignette showcasing an expanded explanation with simulated and real transect data analysis.

## Methods and features

We introduce `comspat`, which calculates the number of realized species combinations (NRC) and diversity of species combinations (called compositional diversity; CD) as a function of the sampling scale. `comspat` provides two main functions: one to calculate results (`'comspat'`) and one to assist in their interpretation (`'comspat_plot'`). The package includes six data sets to demonstrate its operation; we used three in-text (described below) and present the remaining in the vignette. The main functions provided by the package are as follows:

1. `comspat(data, params, dim_max, type, measures, randomization_type, iterations)` calculates the JNP measures (supplied as `'NRC'` and `'CD'`) over a range of secondary sample scales provided by the `params` argument. If a `randomization_type` is specified, `comspat` generates null model distributions (using the number of supplied iterations). The function returns a list containing raw data and summary statistics for each secondary sampling scale.
2. `comspat_plot(data, params, type, measure, stats_output, su_size, ...)` is a function designed to interpret `comspat` results. The function accepts a list of `comspat` data results (or null model results) and returns base R plot objects. The `stats_output` argument returns 95% confidence intervals.

The following section guides users through the Information Theory models (i.e. NRC and CD). Specifically, we provide notes on randomization and a guide showing how to use `comspat` to detect within-community spatial organization.

## Information theory models

Our measures of beta-diversity are the number of realized (observed) species combinations (NRC) and the Shannon diversity of realized species combinations (compositional diversity; CD) at a given scale. NRC and CD are the two commonly used community-level models defined by Juhász-Nagy (for a review, Bartha et al. 1998). CD is a descriptor of the frequency distribution of realized species combinations within a community expressed as the probability of the species combinations in a sample:

$$H_j(A, B, \dots, Y) = -\sum_{k=1}^{2^s} p_{kj} \log(p_{kj})$$

where  $A, B, \dots, Y$  are the  $s$  species of the community;  $p_{kj}$  is the probability of finding a particular species combination (i.e. the  $k$ th combination of species) within the sample at a

specific sampling unit size  $j$ ;  $k$  ranges from 1 to  $2^s$ . Due to different assembly constraints, the observed number of species combinations (NRC) in natural communities is usually smaller than the theoretical maxima (Bartha 1992).

These measures reflect compositional variability within-communities, i.e. by definition, they belong to the families of beta-diversity measures (Anderson et al. 2011). Other beta diversity measures express compositional variability using compositional dissimilarity or species diversity differences between scales (Podani 2000). NRC and CD represent a unique approach by distinguishing and counting local species combinations.

## Notes on randomization

Null models are fundamental tools in spatial pattern analyses used to disentangle the influences of different factors (species richness, species abundance distribution, spatial aggregations and associations and effects of sampling design) on the detected characteristics of patterns (Podani 1984a, Dale and Fortin 2014). The package offers complete spatial randomness (CSR) and random shifts (RS) null models. CSR randomly re-allocates the individuals of each species from the original data into a new spatial sample; this maintains all textural (i.e. related to species abundance distribution) and removes the structural (i.e. associated with the spatial distribution of species) components of association. CSR null models help to show the combined effects of individual species level spatial aggregations and interspecific associations on observed (realized) coexistence relationships. RS maintains the spatial patterns of individual species but uses reflections, rotations and shifts of each species to randomize interspecific patterns. RS null models help to show the effects of interspecific associations after removing the influence of intraspecific aggregations on observed (realized) coexistence relationships. The separation of textural, intraspecific and interspecific effects assists in detecting and interpreting the spatial associations and inferring assembly mechanisms. We urge caution with such generality because to unambiguously assume the action of specific assembly mechanisms from examining a single spatial pattern alone is risky (Rejmánek and Lepš 1996, Blanchet et al. 2020). To better interpret community assembly mechanism – spatial pattern relationships replicates in a ‘space as a surrogate’ approach combined with experimental manipulation (where practicable) should be considered (McIntire and Fajardo 2009). The `randomization_type` argument can call both randomizations by supplying `'CSR'` or `'RS'`. By default, the package generates 999 independent randomizations to assess the distributions of each randomized model.

## Example

### Step 1. Import the grid data set provided by the package

`comspat` requires field data collected by standardized sampling design involving fine-scale grid or transect sampling

(Juhász-Nagy and Podani 1983, Bartha et al. 2020). The raw data ( $x$ ) accepted by all functions in the package is a matrix ( $x$ ) coding the presence of  $m$  species in  $n$  sampling units as rows. In addition, sampling units collected as a grid must contain columns (named 'X' and 'Y') codifying the sampling unit location using cartesian coordinates. A contiguous number reported in the 'X' column must identify sampling units collected from a transect.

The grid data sets ( $64 \times 64$  sampling units) feature one simulation (`grid_patchy_associated`) and two simulation derivatives (`grid_random` and `grid_patchy_no_isc`) produced by randomization (Fig. 1).

We used a spatially explicit individual-based simulation model (PATPRO) to create the `grid_patchy_associated` data set; this simulation represents a realistic grid community pattern featuring interspecific species associations and environmental responses (Czaran 1993, Bartha et al. 1998). RS and CSR randomizations of the original simulation (`grid_patchy_associated`) produced the `grid_n_isc` and `grid_random` data sets. The `grid_n_isc` data set maintains species-level patterns; however, the interspecific relationships are randomized. The `grid_random` data set shows random species distribution, no intraspecific autocorrelations and

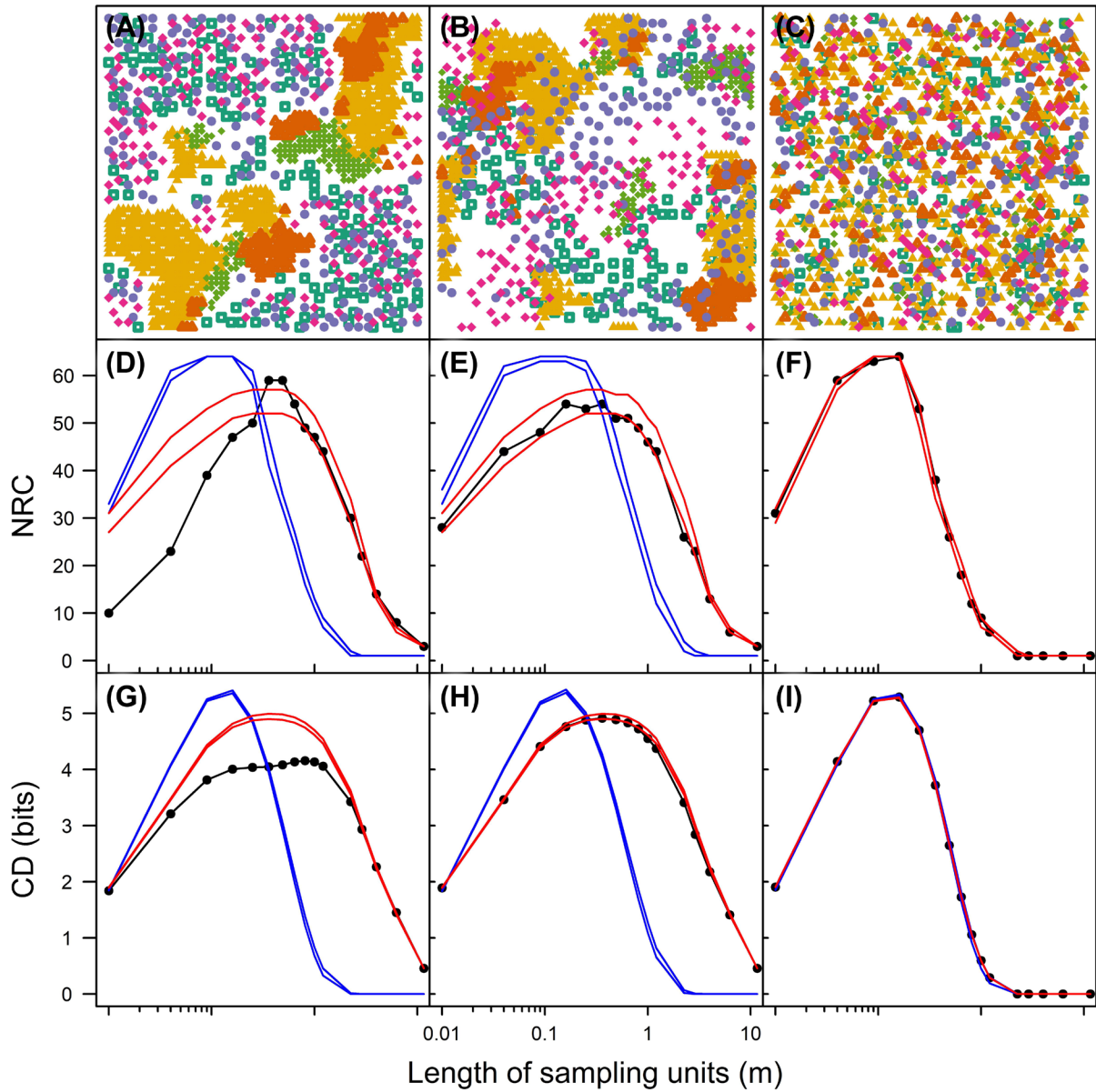


Figure 1. Mapped patterns (simulation, A and two simulation derivatives, B and C) and the corresponding obtained results in terms of the number of realized combinations of species (NRC; D, E, F) and compositional diversity (CD; G, H, I) in grid data sets. The black lines with dots represent the observed NRC and CD values, while 95% confidence intervals (999 randomizations) generated by the complete spatial randomization and random shift null models are represented by blue and red lines, respectively.



no interspecific spatial associations. For additional details, please see the package vignette.

```
# Installing the package
library(remotes)
install_github("jamestsakalos/comspat", build_vignettes=TRUE)
library(comspat)
# Viewing the vignette
vignette(package="comspat")
# Import the grid data set
data(grid_patchy_associated)
```

## Step 2. Import data set controlling spatial scaling

The package provides example parameter data sets (`params_grid` and `params_tran`) for the grid and transect demonstration data. The parameter data sets control the number of scaling steps (`'Steps_of_scaling'`) and the size (the number of primary sampling units to create the length and height of the secondary sampling unit, captured in the `'Length_of_plots'` and `'Height_of_plots'` columns) of the secondary sampling units. In our examples, the parameter data sets use 16 steps of scaling and perform complete resampling of all sample units.

```
# Import the parameter data
data(param_grid)
```

## Step 3. Analyze and plot the data set

We can now run `comspat` to calculate NRC and CD and produce null models using Monte Carlo randomizations. The `comspat` function outputs the raw and statistical report data sets. For example, with nine iterations, the list length equals 10 (representing nine randomization iterations and the original observed pattern). Each list contains two components; one presents NRC results and one for CD. The statistical report shows the results of the observed data on the first row, and the remaining rows show the null model (randomizations); each column represents a spatial scale.

```
# Calculate NRC and CD
assoc_csr <- comspat(data=grid_patchy_associated, params=
  param_grid, dim_max=64, type="Grid",
  randomization_type=
  "CSR", iterations=10)
assoc_rs <- comspat(data=grid_patchy_associated, params=
  param_grid, dim_max=64, type="Grid",
  randomization_type =
  "RS", iterations=10)
# Plot the data and null model 95% CI
Data <- list("CSR"=assoc_csr[[2]],
  "Random Shift"=assoc_rs[[2]])
p_col <- list("blue", "red")
comspat_plot(data=Data, params=
  param_grid, type="Grid", measure=
  "NRC", su_size=0.01, ymin=0, ymax=65,
  p_col=p_col, stats_output=TRUE) # try
  changing NRC to CD.
```

Proper estimation and interpretation of NRC and CD requires large sample sizes and applying null models (defaulting to 999 randomizations); understandably, increased sample size (involving the complete resampling of all measured sampling units) and the use of null model randomizations increases the computational time. Therefore, we have included parallel computing when randomizations are enabled to improve the efficiency of our code. Specifically, our package opens sequential processing using the `future` R package (Bengtsson 2021). Using parallel computing and the example grid data set, the processing time for 999 randomizations was ca. 1 hr and 10 min versus ca. 3 hrs and 30 min without multicore computing using a 1.5 GHz Intel i7 with 8 GB RAM. Users can enhance the timing by considering only the patterns of dominant or abundant species; the concept of spatial pattern has little meaning for rare species (Tóthmérész and Erdei 1992).

Testing was a vital part of the development of this package. We corroborated our results against unit tests (we have incorporated the manual calculations of CD and NRC from the vignette with the `testthat` package; Wickham 2011) and outputs generated from existing programs (Podani 1984b, Bartha et al. 1994, Horváth 1998). In this manner, we promote the validity of our calculations within the package and the commensurability between historical analyses programs.

Here we present an example with grid data comparing three artificial communities. These communities have the same species richness and abundance distribution of species (see details in the package vignette) and contrasting spatial patterns (Fig. 1). In the first community (simulated by a spatially explicit individual-based model), species have aggregated patterns and are also associated with each other (Fig. 1A). NRC and CD calculated from this data (black line with dots; Fig. 1D and G) significantly differ from the null models. The other two communities have been derived from the first community by different randomizations. The interspecific associations were removed by applying random shifts (null model with RS), and the species level aggregated patterns (intraspecific spatial dependence) were maintained (Fig. 1B). The third pattern represents complete spatial randomness (null model with CSR) where both interspecific and intraspecific spatial dependence was removed (Fig. 1C). Across three different spatially organized communities NRC and CD are variable in their local optima, characteristic scales and how they deviate from null models (Fig. 1). Classical Shannon diversity was the same ( $H' = 1.58$ ) for all three communities. This variability presented by the bioinformatic version of Shannon diversity reflects the differences in spatial community organization and can be used to disentangle the textural, intraspecific and interspecific effects on the observed spatial patterns.

## Discussion

In this manuscript, we present the new R package `comspat`. This R package offers a set of Information Theory models developed by Juhász-Nagy (Juhász-Nagy 1967a,

1976, 1984b, 1993, Juhász-Nagy and Podani 1983). These models provide a suite of diversity measures designed for representing the fine-scale spatial organization in communities and have been thoroughly verified in the past (Podani et al. 1993, Bartha et al. 1998). Not only in assemblages of plant communities (Bartha 1992, Bartha et al. 2011, 2020), but the method is also amenable to the spatial pattern analysis of other sessile organisms (coral reefs, Price et al. 2021, e.g. subtidal algae and invertebrates, Terlizzi et al. 2007), where limited dispersal, local interactions and fine-scale environmental heterogeneity shape the assemblage.

On the topic of spatial pattern analysis, it would be remiss if discourse connecting JNP's family of models with the well-established point pattern analyses were not considered (Baddeley et al. 2015). Point pattern analysis involves sampling a cloud of points mapping the location of every individual in a given sampling universe (Wiegand and Moloney 2013, Dale and Fortin 2014). In contrast, our quadrat-based sampling records the presence of a species in a series of adjacent sampling units of a given size (configured as a grid or transect = sampling universe). We use quadrat-based JNP models in grasslands or forest understory instead of the point pattern analyses because widespread clonality makes it difficult (often impossible) to identify the location of individuals using points. Most studies involving point pattern analysis are from systems where individuals can be mapped, for example, Panamá's famous Barro Colorado Island tree data (Wiegand et al. 2013). Podani and Czárán (1997) have further investigated the links between JNP models and classic point pattern analyses; herein, they emphasized that the advantage of the JNP model was in its ability to represent multi-species patterns in a relatively simple way. Nonetheless, we could add a feature for resampling spatial point patterns with newly generated interest in the package.

Not only does this package contribute to the detection and understanding of assembly rules in the context of the observed plant species combinations (i.e. taxonomic units per se.), but the method is compatible with other entities. For example, suppose plant traits (Kattge et al. 2020) or groups of species sharing similar traits replace species entities (i.e. 'plant functional types', see: Lavorel et al. 1997, Tsakalos et al. 2019). In this manner, `comspat` can explore the patterns and drivers of taxonomic or functional diversity (Podani et al. 2013). `comspat` serves as a platform for fusing traditional floristic-sociological and plant functional trait approaches in searching for vegetation patterns and understanding the processes underpinning those patterns.

While our package focuses on the observed species combinations, absent species combinations with the potential to occur within the species pool might provide additional information in detecting and understanding assembly rules (i.e. dark diversity, Pärtel et al. 2011). Recent methodological developments have used missing species combinations to calculate probabilistic dark diversity estimates from nested

sampling designs (Carmona and Pärtel 2021). Future developments of Information Theory models may provide a crucial link allowing the exact calculation of scale-dependent dark diversity.

To cite `comspat` or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 1.0':

Tsakalos, J. L. et al. 2022. `comspat`: an R package to analyze within-community spatial organization using species combinations. – *Ecography* 2022: XX–XX (ver. 1.0).

*Acknowledgements* – We thank Russell Gray for creating, managing and providing sound advice within the Ecology in R (and the very helpful coders therein) public Facebook group. We thank the reviewers who have provided excellent advice on improving the structure and quality of our R package.

*Funding* – JLT, DC and RC were supported by the LIFE MODERN (NEC) project (LIFE20 GIE/IT/000091). SB was supported by the NKFI K 129068 research grant.

## Author contributions

**James L. Tsakalos:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Software (equal); Validation (equal); Writing – original draft (lead); Writing – review and editing (lead). **Stefano Chelli:** Conceptualization (equal); Writing – review and editing (equal). **Giandiego Campetella:** Conceptualization (equal); Supervision (equal); Writing – review and editing (equal). **Roberto Canullo:** Conceptualization (equal); Project administration (equal); Supervision (equal); Writing – review and editing (equal). **Enrico Simonetti:** Conceptualization (equal); Software (equal); Writing – review and editing (equal). **Sandor Bartha:** Conceptualization (equal); Methodology (equal); Project administration (equal); Software (equal); Validation (equal); Writing – original draft (equal); Writing – review and editing (equal).

## Transparent peer review

The peer review history for this article is available at <<https://publons.com/publon/10.1111/ecog.06216>>.

## Data availability statement

Code, data and package presented in this manuscript are available from the Dryad Digital Repository: <<https://doi.org/10.5061/dryad.7h44j0zw6>> (Tsakalos et al. 2022a) and the Zenodo Digital Repository, <<https://doi.org/10.5281/zenodo.6338037>> (Tsalakos et al. 2022b). Development versions of the package can be accessed from Github <<https://github.com/jamestsakalos/comspat>>.

## Supporting information

The supporting information associated with this article is available from the online version.

## References

- Altieri, L. et al. 2021. Spatial entropy for biodiversity and environmental data: the R-package *SpatEntropy*. – *Environ. Model. Softw.* 144: 105149.
- Anderson, M. J. et al. 2011. Navigating the multiple meanings of  $\beta$  diversity: a roadmap for the practicing ecologist. – *Ecol. Lett.* 14: 19–28.
- Baddeley, A. et al. 2015. *Spatial point patterns: methodology and applications with R*. – Chapman and Hall/CRC Press.
- Bartha, S. 1992. Preliminary scaling for multi-species coalitions in primary succession. – *Abstr. Bot.* 16: 31–41.
- Bartha, S. et al. 1994. JNP-models 1.0: program package for detecting coexistential patterns in multi-species assemblages. – In: Peregovits, L. and Lőkös, L. (eds), *Abstracts of the 3rd Hungarian congress of ecology*. Hungarian Ecological Society, pp. 4–6.
- Bartha, S. et al. 1998. Exploring plant community dynamics in abstract coenostate spaces. – *Abstr. Bot.* 22: 49–66.
- Bartha, S. et al. 2011. Beta diversity and community differentiation in dry perennial sand grasslands. – *Ann. Bot.* 1: 9–18.
- Bartha, S. et al. 2020. Unimodal relationships of understory alpha and beta diversity along chronosequence in coppiced and unmanaged beech forests. – *Diversity* 12: 1–18.
- Bengtsson, H. 2021. A unifying framework for parallel and distributed processing in R using *Futures*. – *R J.* 13: 273–291.
- Ben-Hur, E. and Kadmon, R. 2020. Heterogeneity–diversity relationships in sessile organisms: a unified framework. – *Ecol. Lett.* 23: 193–207.
- Blanchet, F. G. et al. 2020. Co-occurrence is not evidence of ecological interactions. – *Ecol. Lett.* 23: 1050–1063.
- Cao, L. and Grabchak, M. 2015. *EntropyEstimation*: estimation of entropy and related quantities. – R package ver. 1.2. <<https://CRAN.R-project.org/package=EntropyEstimation>>.
- Carmona, C. P. and Pärtel, M. 2021. Estimating probabilistic site-specific species pools and dark diversity from co-occurrence data. – *Global Ecol. Biogeogr.* 30: 316–326.
- Chase, J. M. and Knight, T. M. 2013. Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. – *Ecol. Lett.* 16: 17–26.
- Czaran, T. 1993. PATPRO: a Monte-Carlo simulation program for multispecies neighborhood competition. – *Abstr. Bot.* 17: 275–281.
- Dale, M. R. T. and Fortin, M.-J. (eds) 2014. *Spatial analysis: a guide for ecologists*. – Cambridge Univ. Press.
- Dengler, J. et al. 2018. *GrassPlot* – a database of multi-scale plant diversity in Palaeartic grasslands. – *Phytocoenologia* 48: 331–347.
- Hausser, J. and Strimmer, K. 2009. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. – *J. Mach. Learn. Res.* 10: 1469–1484.
- Horváth, A. 1998. Infoterm Program: new possibilities of spatial series analysis based on information theory methods. – *Tiscia* 31: 71–84.
- Jost, L. 2006. Entropy and diversity. – *Oikos* 113: 363–375.
- Juhász-Nagy, P. 1967a. On association among plant populations I. – *Acta Biol. Debr.* 5: 43–56.
- Juhász-Nagy, P. 1967b. Some ‘characteristic areas’ of plant community stands. – In: Rényi, A. (ed.), *Proceedings of the colloquium on information theory*. Bolyai Mathematical Society, pp. 269–282.
- Juhász-Nagy, P. 1976. Spatial dependence of plant populations. Part I. Equivalence analysis (an outline for a new model). – *Acta Bot. Acad. Sci. Hung.* 22: 61–78.
- Juhász-Nagy, P. 1984a. Notes on diversity. Part I. Introduction. – *Abstr. Bot.* 8: 43–55.
- Juhász-Nagy, P. 1984b. Spatial dependence of plant populations. Part 2. A family of new models. – *Acta Bot. Hung.* 30: 363–402.
- Juhász-Nagy, P. 1993. Notes on compositional diversity. – *Hydrobiologia* 249: 173–182.
- Juhász-Nagy, P. and Podani, J. 1983. Information theory methods for the study of spatial processes and succession. – *Vegetatio* 51: 129–140.
- Kattge, J. et al. 2020. TRY plant trait database – enhanced coverage and open access. – *Global Change Biol.* 26: 119–188.
- Lavorel, S. et al. 1997. Plant functional classifications: from general groups to specific groups based on response to disturbance. – *Trends Ecol. Evol.* 12: 474–478.
- Magurran, A. E. (ed.) 2004. *Measuring biological diversity*. – Blackwell Publishing.
- Marcon, E. and Hérault, B. 2015. *entropart*: an R package to measure and partition diversity. – *J. Stat. Softw.* 67: 1–26.
- McIntire, E. J. B. and Fajardo, A. 2009. Beyond description: the active and effective way to infer processes from spatial patterns. – *Ecology* 90: 46–56.
- Pärtel, M. et al. 2011. Dark diversity: shedding light on absent species. – *Trends Ecol. Evol.* 26: 124–128.
- Podani, J. 1984a. Analysis of mapped and simulated vegetation patterns by means of computerized sampling techniques. – *Acta Bot. Hung.* 30: 403–425.
- Podani, J. 1984b. SYN-TAX II. Computer programs for data analysis in ecology and systematics. – *Abstr. Bot.* 8: 73–94.
- Podani, J. (ed.) 2000. *Introduction to the exploration of multivariate biological data*. – Backhuys Publishers.
- Podani, J. and Czárán, T. 1997. Individual-centered analysis of mapped point patterns representing multi-species assemblages. – *J. Veg. Sci.* 8: 259–270.
- Podani, J. et al. 1993. Pattern, area and diversity: the importance of spatial scale in species assemblages. – *Abstr. Bot.* 17: 37–51.
- Podani, J. et al. 2013. Combinatorial functional diversity: an information theoretical approach. – *Community Ecol.* 14: 180–188.
- Price, D. M. et al. 2021. Fine-scale heterogeneity of a cold-water coral reef and its influence on the distribution of associated taxa. – *Front. Mar. Sci.* 8: e556313.
- Rejmánek, M. and Lepš, J. 1996. Negative associations can reveal interspecific competition and reversal of competitive hierarchies during succession. – *Oikos* 76: 161–168.
- Rényi, A. 1961. On measures of entropy and information. – In: Neymand, J. (ed.), *Proceedings of the 4th Berkeley symposium on mathematical statistics and probability*. Univ. of California Press, pp. 547–561.
- Rocchini, D. et al. 2021. *rasterdiv* – an information theory tailored R package for measuring ecosystem heterogeneity from space: to the origin and back. – *Methods Ecol. Evol.* 12: 1093–1102.
- Sandel, B. 2015. Towards a taxonomy of spatial scale-dependence. – *Ecography* 38: 358–369.
- Shannon, C. E. 1948. A mathematical theory of communication. – *Bell Syst. Techn. J.* 27: 379–423.
- Simpson, E. H. 1949. Measurement of diversity. – *Nature* 163: 688.

- Terlizzi, A. et al. 2007. Scales of spatial variation in Mediterranean subtidal sessile assemblages at different depths. – *Mar. Ecol. Progr. Ser.* 332: 25–39.
- Tóthmérész, B. and Erdei, Z. 1992. The effect of species dominance on information theory characteristics of plant communities. – *Abstr. Bot.* 16: 43–47.
- Tsakalos, J. L. et al. 2019. Trait-based formal definition of plant functional types and functional communities in the multi-species and multi-traits context. – *Ecol. Complex.* 40: e100787.
- Tsakalos, J. L. et al. 2022a. Data from: comspat: an R package to analyze within-community spatial organization using species combinations. – Dryad Digital Repository, <<https://doi.org/10.5061/dryad.7h44j0zw6>>.
- Tsakalos, J. L. et al. 2022b. Code from: comspat: an R package to analyze within-community spatial organization using species combinations. – Zenodo Digital Repository, <<https://doi.org/10.5281/zenodo.6338037>>.
- Wickham, H. 2011. testthat: get started with testing. – *R J.* 3: 5–10.
- Wiegand, T. and Moloney, K. A. 2013. Handbook of spatial point-pattern analysis in ecology. – CRC Press.
- Wiegand, T. et al. 2013. A systematic comparison of summary characteristics for quantifying point patterns in ecology. – *Ecography* 36: 92–103.
- Wilson, J. B. 1999. Assembly rules in plant communities. – In: Weiher, E. and Keddy, P. (eds), *Ecological assembly rules: perspectives, advances, retreats.* – Cambridge Univ. Press, pp. 130–164.
- Wilson, J. B. et al. 2012. Plant species richness: the world records. – *J. Veg. Sci.* 23: 796–802.