

# shar: An R package to analyze species-habitat associations using point pattern analysis

Maximilian H. K. Hesselbarth<sup>\*1, 2</sup>

<sup>1</sup> Department of Ecosystem Modelling, University of Goettingen, Buesgenweg 4, 37077, Goettingen

<sup>2</sup> Department of Ecology and Evolutionary Biology, University of Michigan, 1105 N University Ave, Ann Arbor, Michigan 48109, USA

DOI: [10.21105/joss.03811](https://doi.org/10.21105/joss.03811)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: Hugo Gruson ↗

## Reviewers:

- [@lionel68](#)
- [@tretherington](#)

Submitted: 01 October 2021

Published: 06 December 2021

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

Studying species-habitat associations is one approach to reveal the importance of abiotic processes in shaping the spatial distribution of species. Even though the R programming language offers many packages for spatial point pattern analysis, currently there is no comprehensive package specifically designed to analyze species-habitat associations. The *shar* package builds on widely used R packages for spatial analyses and provides an easy and straightforward way to uncover species-habitat associations for discrete environmental data.

## Statement of need

Species-habitat associations are a result of certain species being specialized to certain environmental conditions (Tilman & Pacala, 1993) and typically result in a clustering of individuals at suitable habitats (Comita et al., 2007; Harms et al., 2001). Thus, analyzing species-habitat associations can help to understand the importance of abiotic processes shaping the spatial distribution of species (Garzon-Lopez et al., 2014). However, since biotic processes (e.g., competition, limited dispersal) can also lead to a clustering of individuals, analyses of species-habitat associations need to control for potential biotic processes because they result in a violation of the independence assumption of similar statistical tests, such as the  $\chi^2$  test, ordination methods, or canonical correspondence analysis (Harms et al., 2001; Plotkin et al., 2000). Spatial autocorrelation of the environmental conditions could also violate the independence assumption of the previously mentioned statistical tests (Harms et al., 2001). For example, previous research has shown that violating the independence assumptions of the  $\chi^2$  test resulted in more (i.e., possible false-positive) species-habitat associations than the more conservative methods that are provided by the *shar* package (Harms et al., 2001; Plotkin et al., 2000).

Ecologists use the spatial distribution of ecological objects to infer the underlying processes that shaped their distribution because spatial patterns can act as a “memory” of the processes that shaped them (Velázquez et al., 2016). A spatial point pattern contains the fully mapped locations (in terms of x,y coordinates) of all individual objects in a normally two-dimensional study area and assumes that the object locations can be described by discrete points (Velázquez et al., 2016; Wiegand & Moloney, 2014). For example, many studies use individual tree locations to infer the processes that determined their distribution (Velázquez et al., 2016), and further examples include gopher mounds (Klaas et al., 2000), gorilla nests (Funwi-Gabga & Mateu, 2012), “fairy circles” (Getzin et al., 2015), or bacteria on leaves (Esser et al., 2015).

<sup>\*</sup>corresponding author

The `spatstat` package (Baddeley et al., 2015) allows ecologists to access many methods of spatial point pattern analysis, such as summary functions and null model simulations, and to simulate heterogeneous point process models that can be used to show the importance of abiotic processes for continuous environmental data (Getzin et al., 2008). However, even though many ecological studies on species-habitat associations using discrete environmental data can be found in the literature (Du et al., 2017; Furniss et al., 2017; Garzon-Lopez et al., 2014; Guo et al., 2016; John et al., 2007; Yang et al., 2016), `spatstat` cannot be used to reveal such associations without larger programming efforts by the users. The `inlabru` package provides an approach to analyze the importance of abiotic processes, mostly for continuous environmental data, using Bayesian spatial modelling (Bachl et al., 2019). The `fgeo` package (Lepore et al., 2019) allows to visualize and analyze forest diversity, including species-habitat associations. But, the `fgeo` was designed to specifically handle ForestGEO data (<https://forestgeo.si.edu>) and furthermore includes only a subset of methods available to analyze species-habitat associations.

Thus, the `shar` package was developed to provide a comprehensive tool set to analyze species-habitat associations of spatial point patterns. All methods in the `shar` package are designed for discrete environmental data and have the advantage of very few assumptions about the data characteristics. In order to make the `shar` package as accessible for as many people as possible, it builds on two of the most commonly used R packages for spatial data, namely the `spatstat` and `raster` packages (Hijmans, 2019).

## Methodological background

To analyze species-habitat associations, potential interdependence between the object locations and the environmental conditions needs to be broken by randomizing the data as a null model. Within the field of spatial point pattern analysis, there are two related approaches to break potential dependencies (Harms et al., 2001; Plotkin et al., 2000). Both require the spatial location of all objects, as well as discrete raster data for environmental conditions.

The first approach to simulate null model data is to randomize the environmental data, while keeping the object locations stable. This can be achieved by shifting the raster data around a torus (“torus translation test”) or using a random walk algorithm (“randomized-habitats procedure”) (Harms et al., 2001). The second approach is to randomize the object locations, while keeping the environmental data stable. This can be achieved by fitting point process models to the object locations (“gamma test”) (Plotkin et al., 2000) or using a simulated annealing approach (“pattern reconstruction”) (Kirkpatrick et al., 1983).

The two approaches differ in how they randomize the null model data, but both control for potential biotic processes by preserving the spatial structure of the data (Plotkin et al., 2000; Wiegand & Moloney, 2014) and result in similar results. Finally, species-habitat associations are present if species are found in certain habitats in the data more often than expected compared to the randomized null model data (Harms et al., 2001; Plotkin et al., 2000). Given the characteristics of the method, a positive association to one habitat inevitably leads to a negative association to at least one of the other habitats (and vice versa) (Yamada et al., 2006).

## How to use the package

Analyzing species-habitat associations is straightforward with the `shar` package. Only two objects are needed to quantify species-habitat associations, namely a `spatstat` object that includes all object locations within the study area and a `raster` object with discrete habitat classes.

However, all methods require “fully mapped data” in the sense that NA cells of the environmental data are allowed only if simultaneously these areas cannot accommodate any locations of the point pattern (e.g., a water body within a forest area). This needs to be reflected in the observation window of the point pattern. For the torus translation method, no NA values are allowed at all.

To randomize the environmental data, either `translate_raster()` or `randomize_raster()` can be used. For the former, the number of randomizations of the null model data is automatically determined by the number of rows and columns of the raster object. For the latter, the number of randomizations must be specified using the `n_random` argument.

```
torus_trans <- translate_raster(raster = landscape_discrete)

random_walk <- randomize_raster(raster = landscape_discrete, n_random = 39)
```

Alternatively, to randomize the object locations, either `fit_point_process()`, `reconstruct_pattern()`, or `reconstruct_pattern_marks()` can be used. In all cases, the number of randomization must be specified using the `n_random` argument. In order to preserve the spatial structure of the input as detailed as possible, several options are present to acknowledge for example if the input object locations are clustered or heterogeneously distributed in the study area.

```
gamma_test <- fit_point_process(pattern = species_a, n_random = 39,
                                process = "cluster")

reconstruction <- reconstruct_pattern(pattern = species_a, n_random = 39,
                                       method = "cluster")
```

Lastly, the input data and the randomized null model data are used to test if significant species-habitat associations are present. The `results_habitat_association()` function automatically detects which part of the data was randomized and can be used identically with either of the used randomization approach.

```
results_habitat_association(pattern = species_a, raster = random_walk,
                             significance_level = 0.01)
```

```
> Input: randomized raster
> Quantile thresholds: negative < 0.005 || positive > 0.995
  habitat count lo hi significance
1         1   35 10 35          n.s.
2         2   44 19 53          n.s.
3         3   36 15 49          n.s.
4         4    4 15 58       negative
5         5   73 48 90          n.s.
```

The `shar` package also provides several utility and plotting functions such as a generic `plot()` function to plot the null model data, `calculate_energy()` to calculate the difference between the input object locations and the randomized null model data object locations, or `classify_habitats()` to classify continuous environmental data into discrete habitats. For all functions, please see the “Functions” article on the `shar` homepage (<https://r-spatialecology.github.io/shar>).

## Parallelization

One major drawback of the `shar` package is the computation time related to some of the randomization methods for null model data. This is the case especially for pattern reconstruction, even though most point pattern analysis studies use less than 1000 null model randomizations (Velázquez et al., 2016). However, since the randomizations of the null model data are independent of each other, this could be parallelized using available frameworks, such as the future (Bengtsson, 2021) or parallel (R Core Team, 2021) package. The `shar` package does not allow to run code in parallel internally to not limit users to a specific parallelization framework. For a short example how to simulate null model data in parallel using the future package, please see the “Parallelization” article on the `shar` homepage (<https://r-spatialecology.github.io/shar>). However, the presented approach could be used with any other parallelization framework as well.

## Acknowledgements

Support was provided by the German Research Association (DFG) Research Training Group 1644 “Scaling Problems in Statistics,” grant number 152112243. M.H.K.H. is thankful to Sebastian Hanss und Marco Sciaini for their help during the development of the `shar` package and Katrina Munsterman, Bridget Shayka and Samantha Iliff for comments on earlier drafts of the manuscript. Thomas Etherington and Lionel Hertzog provided valuable feedback during the review process that improved the manuscript and the `shar` package.

## References

- Bachl, F. E., Lindgren, F., Borchers, D. L., & Illian, J. B. (2019). Inlabru: An r package for bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, 10(6), 760–766. <https://doi.org/10.1111/2041-210X.13168>
- Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: Methodology and applications with r*. Chapman; Hall/CRC Press. ISBN: 978-1-4822-1020-0
- Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r using futures. *The R Journal*. <https://doi.org/10.32614/RJ-2021-048>
- Comita, L. S., Condit, R., & Hubbell, S. P. (2007). Developmental changes in habitat associations of tropical trees. *Journal of Ecology*, 95(3), 482–492. <https://doi.org/10.1111/j.1365-2745.2007.01229.x>
- Du, H., Hu, F., Zeng, F., Wang, K., Peng, W., Zhang, H., Zeng, Z., Zhang, F., & Song, T. (2017). Spatial distribution of tree species in evergreen-deciduous broadleaf karst forests in southwest china. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-15789-5>
- Esser, D. S., Leveau, J. H. J., Meyer, K. M., & Wiegand, K. (2015). Spatial scales of interactions among bacteria and between bacteria and the leaf surface. *FEMS Microbiology Ecology*, 91(3), 1–13. <https://doi.org/10.1093/femsec/fiu034>
- Funwi-Gabga, N., & Mateu, J. (2012). Understanding the nesting spatial behaviour of gorillas in the kagwe sanctuary, cameroon. *Stochastic Environmental Research and Risk Assessment*, 26(6), 793–811. <https://doi.org/10.1007/s00477-011-0541-1>
- Furniss, T. J., Larson, A. J., & Lutz, J. A. (2017). Reconciling niches and neutrality in a subalpine temperate forest. *Ecosphere*, 8(6), e01847. <https://doi.org/10.1002/ecs2.1847>

- Garzon-Lopez, C. X., Jansen, P. A., Bohlman, S. A., Ordonez, A., & Olff, H. (2014). Effects of sampling scale on patterns of habitat association in tropical trees. *Journal of Vegetation Science*, 25(2), 349–362. <https://doi.org/10.1111/jvs.12090>
- Getzin, S., Wiegand, K., Wiegand, T., Yizhaq, H., von Hardenberg, J., & Meron, E. (2015). Adopting a spatially explicit perspective to study the mysterious fairy circles of namibia. *Ecography*, 38(1), 1–11. <https://doi.org/10.1111/ecog.00911>
- Getzin, S., Wiegand, T., Wiegand, K., & He, F. (2008). Heterogeneity influences spatial patterns and demographics in forest stands. *Journal of Ecology*, 96(4), 807–820. <https://doi.org/10.1111/j.1365-2745.2008.01377.x>
- Guo, Y., Wang, B., Mallik, A. U., Huang, F., Xiang, W., Ding, T., Wen, S., Lu, S., Li, D., He, Y., & Li, X. (2016). Topographic species-habitat associations of tree species in a heterogeneous tropical karst seasonal rain forest, china. *Journal of Plant Ecology*, 10(3), 450–460. <https://doi.org/10.1093/jpe/rtw057>
- Harms, K. E., Condit, R., Hubbell, S. P., & Foster, R. B. (2001). Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *Journal of Ecology*, 89(6), 947–959. <https://doi.org/10.1111/j.1365-2745.2001.00615.x>
- Hijmans, R. J. (2019). *Raster: Geographic data analysis and modeling*. R package version 2.9-5. <<https://cran.r-project.org/package=raster>>.
- John, R., Dalling, J. W., Harms, K. E., Yavitt, J. B., Stallard, R. F., Mirabello, M., Hubbell, S. P., Valencia, R., Navarrete, H., Vallejo, M., & Foster, R. B. (2007). Soil nutrients influence spatial distributions of tropical tree species. *Proceedings of the National Academy of Sciences of the United States of America*, 104(3), 864–869. <https://doi.org/10.1073/pnas.0604666104>
- Kirkpatrick, S., Gelatt, C. D. Jr., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671–680. <https://doi.org/10.1126/science.220.4598.671>
- Klaas, B. A., Moloney, K. A., & Danielson, B. J. (2000). The tempo and mode of gopher mound production in a tallgrass prairie remnant. *Ecography*, 23, 246–256. <https://doi.org/10.1111/j.1600-0587.2000.tb00280.x>
- Lepore, M., Arellano, G., Condit, R., Davies, S., Detto, M., Gonzalez-Akre, E., Hall, P., Harms, K., Herrmann, V., Kenfack, D., Lao, S., McMahon, S., Russo, S., Anderson-Teixeira, K., Zemunik, G., & Zuleta, D. (2019). *Fgeo: Analyze forest diversity and dynamics*. R package version 1.1.4. <<https://CRAN.R-project.org/package=fgeo>>.
- Plotkin, J. B., Potts, M. D., Leslie, N., Manokaran, N., LaFrankie, J. V., & Ashton, P. S. (2000). Species-area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology*, 207(1), 81–99. <https://doi.org/10.1006/jtbi.2000.2158>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Tilman, D., & Pacala, S. W. (1993). The maintenance of species richness in plant communities. In R. E. Ricklefs & D. Schluter (Eds.), *Species diversity in ecological communities* (pp. 13–25). University of Chicago Press.
- Velázquez, E., Martínez, I., Getzin, S., Moloney, K. A., & Wiegand, T. (2016). An evaluation of the state of spatial point pattern analysis in ecology. *Ecography*, 39, 1–14. <https://doi.org/10.1111/ecog.01579>
- Wiegand, T., & Moloney, K. A. (2014). *Handbook of spatial point-pattern analysis in ecology*. Chapman; Hall/CRC Press. ISBN: 978-1-4200-8254-8
- Yamada, T., Tomita, A., Itoh, A., Yamakura, T., Ohkubo, T., Kanzaki, M., Tan, S., & Ashton, P. S. (2006). Habitat associations of sterculiaceae trees in a bornean rain forest plot.

*Journal of Vegetation Science*, 17(5), 559–566. <https://doi.org/10.1111/j.1654-1103.2006.tb02479.x>

Yang, Q.-S., Shen, G.-C., Liu, H.-M., Wang, Z.-H., Ma, Z.-P., Fang, X.-F., Zhang, J., & Wang, X.-H. (2016). Detangling the effects of environmental filtering and dispersal limitation on aggregated distributions of tree and shrub species: Life stage matters. *PLoS ONE*, 11(5), 1–16. <https://doi.org/10.1371/journal.pone.0156326>