

## Introduction to Online Program SpadeR (Species-richness Prediction And Diversity Estimation in R)

by

Anne Chao, K. H. Ma, T. C. Hsieh and Chun-Huo Chiu

Institute of Statistics

National Tsing Hua University, TAIWAN 30043

The program SpadeR is the R-based online version of SPADE available via the link [http://chao.stat.nthu.edu.tw/wordpress/software\\_download/](http://chao.stat.nthu.edu.tw/wordpress/software_download/) or <https://chao.shinyapps.io/SpadeR/>. Clicking these links, you will be directed to the online interface window. **Users do not need to learn/understand R to run SpadeR.** The interactive web application was built using the Shiny (a web application framework). SpadeR includes nearly all of the important features from the original program SPADE while also having the advantages of expanded output displays and simplified data input formats. Further, owing to the power of the R language, SpadeR now offers high-resolution plots/figures that the original SPADE lacks. Note, however, that some features in SPADE have been expanded to become an independent online program iNEXT; see below.

Like the original SPADE, the program SpadeR computes various biodiversity indices based on two major types of sample data (abundance data and replicated incidence data) taken from one or multiple communities. A detailed SpadeR User's Guide [http://chao.stat.nthu.edu.tw/wordpress/wp-content/uploads/software/SpadeR\\_UserGuide.pdf](http://chao.stat.nthu.edu.tw/wordpress/wp-content/uploads/software/SpadeR_UserGuide.pdf) illustrates how to run this program in an easily accessible way through numerical examples with proper interpretations of portions of the output. SpadeR is divided into six parts:

- Part I: Species (estimating species richness for one community).
- Part II: Diversity Profile Estimation (estimating a continuous diversity profile and various diversity indices in one community). This expanded part also features plots of empirical and estimated continuous diversity profiles.
- Part III: Shared Species (estimating the number of shared species between two communities).
- Part IV: Two-Community (Similarity) Measures (estimating various similarity indices between two assemblages). Both richness- and abundance-based two-community similarity indices are included.
- Part V: Multiple-Community (Similarity) Measures (estimating various  $N$ -community similarity indices). Both richness- and abundance-based  $N$ -community similarity indices are included.
- Part VI: Genetics (Differentiation) Measures (applying Part V to estimate allelic dissimilarity/differentiation among sub-populations based on multiple-population genetics data).

NOTE: Part III of the original SPADE (Prediction) has been expanded to become an independent program known as iNEXT (iNterpolation and EXTrapolation) which provides diversity estimates for rarefied and extrapolated samples up to a maximum sample size or sample completeness specified by the user. The program iNEXT also features seamless plots of sample-size- and coverage-based rarefaction and extrapolation sampling curves. Users can download iNEXT from the same links given above.

There are several data input formats options; see User's Guide for details. **If species identification names/codes or site/community labels are recorded in your original data, they must be removed to conform to the SpadeR data formats.** These identification information or labels are irrelevant in our analysis. The running procedures are summarized as follows.

- Step 1. Select an analysis part from the top menu of SpadeR window.
- Step 2. Select your data input format from the Data Setting in the left panel.  
(For Parts I and II, five data input formats are supported; for Parts III, IV and V, three data input formats are supported; for Part VI, only one format is supported.)
- Step 3. Check the  Demo data radio button to load demo data (you can load your own data by checking  Upload data).  
(To load your own incidence-raw data, you must specify the number of sampling units in each community in the left panel of SpadeR window.)
- Step 4. For Part V, choose an order  $q$  (0, 1 or 2) and the comparison target (relative abundances or absolute abundances) you would like to use to compute the (dis)similarity for any pair of samples. For Part VI, only an order  $q$  (0, 1 or 2) needs to be specified.
- Step 5. Press the  button to get the output.

(We use either an analytic method or a bootstrap resampling method to compute s.e. and confidence interval of an estimator. For the latter, the default number of bootstrap replications is 100. You may specify a larger number to obtain more accurate results, but it will take a longer time to get the output. Also, the bootstrap resampling procedures vary with trial, meaning that two different runs for the same data may result in different s.e. estimates and different confidence intervals.)

Along the second row (output) menu, there are four output selection tabs:

- In the “**Estimation**” tab panel, basic data information and various estimates are shown for the demo/uploaded data. You can click “download as txt file” at the bottom of the output to download these estimates as a txt file.
- The functionality of the “**Visualization**” tab panel varies by part. For Species and Shared Species parts, this tab panel shows various estimates and their confidence intervals to facilitate clear comparisons, while for Diversity Profile Estimation Part, this tab panel shows the plots of empirical diversity profile and the estimated profile. These figures can be downloaded by clicking “download as PNG file” at the bottom of the displayed figure.
- The function of the “**Data Viewer**” tab panel is to display the first ten records of the demo/uploaded data. The entire demo/uploaded data can be downloaded by clicking “download as txt file” at the bottom of the displayed data.

- In the “**Introduction**” panel, users can view a brief introduction to SpadeR and a summary of the running procedures.
- In the “**User Guide**” panel, a link will direct users to SpadeR User’s Guide.

To gain familiarity with the program, we suggest that users first run the demo data sets included in SpadeR and check the output with that given in the SpadeR User’s Guide (link given above). Part of the output for each example is also interpreted in the guide to help users understand the numerical results. The formulas for estimators featured in SpadeR with relevant references are provided in the Appendix of SpadeR User’s Guide.

*Please do not use SpadeR in any commercial form or distribute it to other people. Instead, potential users should access the program directly through the SpadeR webpage (see above). If you publish your work based on results from SpadeR, please make references to the relevant papers mentioned in each section of SpadeR User’s Guide and also use the following reference to cite SpadeR:*

Chao, A., Ma, K. H., Hsieh, T. C. and Chiu, C. H. (2015) Online Program SpadeR (Species-richness Prediction And Diversity Estimation in R). Program and User’s Guide published at [http://chao.stat.nthu.edu.tw/wordpress/software\\_download/](http://chao.stat.nthu.edu.tw/wordpress/software_download/).

---

We recommend the following recent papers for pertinent background on biodiversity measures and statistical analyses. These papers can be directly downloaded from Anne Chao’s website.

Chao, A., and Chiu, C. H. (2012). Estimation of species richness and shared species richness. In N. Balakrishnan (ed). *Methods and Applications of Statistics in the Atmospheric and Earth Sciences*. p.76–111, Wiley, New York. ([Background on species richness and shared species richness estimation](#))

Chao, A., and Chiu, C. H. (2016). Nonparametric estimation and comparison of species richness. *Wiley Online Reference in the Life Science*. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0026329. ([Background on comparing species richness across communities](#))

Chao, A., and Chiu, C. H. (2016). Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods in Ecology and Evolution*, 7, 919–928. ([A unified theoretical framework on similarity/differentiation measures](#))

Chao, A., Chiu, C. H. and Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity/differentiation measures through Hill numbers. *Annual Reviews of Ecology, Evolution, and Systematics* 45: 297–324. ([A unified theoretical framework on diversity measures](#))

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K. and Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and

estimation in species diversity studies. *Ecological Monographs* 84:45–67. ([Background on comparing diversity measures across communities](#))

Chao, A. and Jost, L. (2015). Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution*, **6**, 873–882. ([A unified approach to estimating diversity in a community based on incomplete samples](#))

Chao, A., Wang, Y. T. and Jost, L. (2013). Entropy and the species accumulation curve: a novel entropy estimator via discovery rates of new species. *Methods in Ecology and Evolution*, **4**, 1091–1100. ([A nearly optimal estimator of Shannon entropy/diversity based on incomplete samples](#))