

# **A Brief Introduction to iNEXT.4steps Online: software for Four Steps of interpolation and extrapolation of Biodiversity**

**Anne Chao, K. H. Hu**

*Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043*

## **Overview**

iNEXT.4steps (four steps for iNterpolation and EXTrapolation) Online is the R-based interactive online version of iNEXT.4steps available via the link [https://chao.shinyapps.io/iNEXT\\_4steps/](https://chao.shinyapps.io/iNEXT_4steps/) or [http://chao.stat.nthu.edu.tw/wordpress/software\\_download/](http://chao.stat.nthu.edu.tw/wordpress/software_download/). Clicking these links, you will be directed to the online interface window. **Users do not need to learn/understand R to run iNEXT.4steps Online.** The interactive web application was built using the Shiny (a web application framework).

iNEXT.4steps contains four steps of a complete biological analysis for Taxonomic, Phylogenetic or Functional diversity based on Hill numbers:

### (1) Sample completeness profile

The first step provides the sample completeness profile which represents the proportion of observed species in the population (Chao et al., 2020). It can correspond to order  $q$ , which is a weight index. When order  $q$  tends to zero, then we will give more weight to rare species. If order  $q$  tends to unity, then we will equally treat each species, which is also called sample coverage at unity. In contrast, if order  $q$  tends to larger than unity, we will give more weights to abundant species.

(2) Size-based rarefaction and extrapolation analysis and the asymptotic diversity profile

The second step computes the estimated diversities for standardized samples with a common sample size to compare diversity estimates for equally-large (with a common sample size) samples; it is based on the seamless rarefaction and extrapolation (R/E) sampling curves of Hill numbers for  $q = 0, 1$  and  $2$  (Chao and Jost, 2012; Chao et al. 2014). This type of sampling curve plots the diversity estimates with respect to sample size. This step also provides the estimated asymptotic diversity profiles. It is based on statistical estimation of the true Hill number of any order  $q \geq 0$ ; see Chao and Jost (2015) for statistical estimation details.

(3) Non-asymptotic coverage-based rarefaction and extrapolation analysis

The third step computes the estimated diversities for standardized samples with a common sample completeness to compare diversity estimates for equally-complete (with a common sample coverage) samples; it is based on the seamless rarefaction and extrapolation (R/E) sampling curves of Hill numbers for  $q = 0, 1$  and  $2$  (Chao and Jost, 2012; Chao et al. 2014). This type of sampling curve plots the diversity estimates with respect to sample coverage.

(4) Evenness among species abundances

The four step calculates the estimated evenness to measure whether the assemblage is uniform or not (Chao and Ricotta, 2019). In this step, we provide sorted five main classes according to different transformation by species and diversity, and propose a “standardized coverage” (named  $C_{max}$ ) as a judged criterion.  $C_{max}$  means that we compute the diversity estimates for the minimum sample coverage among all samples extrapolated to double reference sizes. Under  $C_{max}$ , users can guarantee the accuracy of Evenness.

The step 2 and 3 can also be promoted to phylogenetic diversity and functional diversity under attribute diversity framework; see Chao et al. (2019, 2021) to get more details.

## **How to cite**

If you publish your work based on results from iNEXT.4steps Online, you should

make references to the following methodology paper:

- Chao, A., Y. Kubota, D. Zelený, C.-H. Chiu, C.-F. Li, B. Kusumoto, M. Yasuhara, S. Thorn, C.-L. Wei, M. J. Costello, and R. K. Colwell (2020). Quantifying sample completeness and comparing diversities among assemblages. *Ecological Research*, 35, 292-314.

## Data

iNEXT.4steps Online supports two types of data:

- Individual-based abundance data: Data for each assemblage/site include sample species abundances in an empirical sample of individuals (called a reference sample).
- Sampling-units-based incidence data: For each assemblage/site, data for a reference sample consist of species presence/absence (or detection/non-detection) in each of multiple sampling units.

See iNEXT.4steps Online User's Guide for data input formats.

## Running procedures for “iNEXT.4steps” analysis

Species identification names are irrelevant in taxonomic diversity assessment but necessary in phylogenetic and functional diversity assessment, so they must be included in your original data to conform to the iNEXT.4steps Online format. The following Steps 1 and 3 are required procedures (selections are placed in blocks); Step 2 is optional.

**Step 1.** “Data Setting” (on the left hand side of the window screen)

(1a) Choose diversity type:  Taxonomic,  Phylogenetic or  Functional diversity.

(1b) Require only for “Phylogenetic”: choose the phylogenetic type  PD for effective total branch length, or  meanPD for effective number of equally divergent lineages.

(1c) Require only for “Functional”: select  AUC for an overall functional diversity which integrates all threshold values between 0 to 1, or select  tau and input the level of threshold for functional diversity under specified threshold value (default level of threshold = dmean, the mean distance between any two individuals randomly selected from the pooled

assemblage).

(1d) Select  to see the illustrative example data or  to load your own data.

(1e) Select data type:  or ; see above for data input formats.

(1f) Select demo data sets.

(1g) Then the names/labels for the uploaded assemblages/sites will be automatically shown in the window below; you can select one set or multiple sets for comparison.

**Step 2.** “General Setting” for statistical procedures

(2a) Optional: specify the number of bootstrap replications to compute s.e. and confidence intervals for each estimator. (Default bootstrap replications = 0). You can type in “0” to skip all bootstrapping to save running time.

(2b) Optional: specify the level for the confidence interval (default level = 0.95).

**Step 3.** Press the  button to get the output.

Note 1: We use a bootstrap resampling method to compute the s.e. and confidence interval of any estimator involved in the analysis. The default number of bootstrap replications is 50 to save running time. You may specify a larger number (say, 100) to obtain more accurate results for publication purposes, but it will take a longer time to get the output.

Note 2: If you just want to take a glance at the pattern of rarefaction/extrapolation sampling curves and diversity profiles without requiring confidence intervals, then type in “0” for the number of bootstrap replications; in this case, the computation of s.e. and confidence intervals will be skipped so that the output can be promptly shown.

Note 3: 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.

## Output for “iNEXT.4steps”

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

- In the “**Data Summary**” tab panel, basic data information, such as sample size, observed species richness and estimated sample coverage refer to the chosen diversity type are shown for the reference sample.
- In the “**View Data**” tab panel, the content of selected data is listed.
- In the “**4.steps Summary**” tab panel, the summary outputs of four steps analysis are displayed. These results can be downloaded by clicking “Save as CSV file” at the bottom of the displayed table.
- In the “**Figure Plots**” tab panel, the combined figure of four steps analysis is shown. This figure can be downloaded by clicking “Save as PNG file” at the bottom of the displayed figures.
- In the “**Figure Plots (individual)**” tab panel, five individual figures of four steps analysis are shown: sample completeness profile, size-based rarefaction/extrapolation, asymptotic and empirical diversity profile, coverage-based rarefaction/extrapolation, and evenness profile. These figures can be downloaded by clicking “Save as PNG file” at the bottom of the displayed figures.
- In the “**Calculated Value**” tab panel, the detailed values for each plot are listed in the table. These tables can be downloaded by clicking “Save as CSV file” at the bottom of the displayed table.
- In the “**Introduction**” panel, users can view a brief introduction to iNEXT.4steps Online and a summary of the running procedures.
- In the “**User Guide**” panel, a link will direct users to this user guide.

## References

The following papers for pertinent background on rarefaction/extrapolation and related statistical analyses. These papers can be directly downloaded from Anne Chao’s website.

Chao, A., Chiu, C.-H., Villéger, S., Sun, I.-F., Thorn, S., Lin, Y.-C., Chiang, J. M. and Sherwin, W. B. (2019). An attribute-diversity approach to functional

diversity, functional beta diversity, and related (dis)similarity measures. *Ecological Monographs*, 89, e01343. 10.1002/ecm.1343.

Chao, A., Gotelli, N. G., 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 biodiversity studies. *Ecological Monographs* 84, 45-67.

Chao, A., Henderson, P. A., Chiu, C.-H., Moyes, F., Hu, K.-H., Dornelas, M and Magurran, A. E. (2021). Measuring temporal change in alpha diversity: a framework integrating taxonomic, phylogenetic and functional diversity and the iNEXT.3D standardization. *Methods in Ecology and Evolution*, 12, 1926-1940.

Chao, A. and Jost, L. (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology*, 93, 2533-2547.

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.

Chao, A. and Ricotta, C. (2019). Quantifying evenness and linking it to diversity, beta diversity, and similarity. *Ecology*, 100(12), e02852.

Chao, A., Y. Kubota, D. Zelený, C.-H. Chiu, C.-F. Li, B. Kusumoto, M. Yasuhara, S. Thorn, C.-L. Wei, M. J. Costello, and R. K. Colwell (2020). Quantifying sample completeness and comparing diversities among assemblages. *Ecological Research*, 35, 292-314.