EstimateS

Statistical Estimation
of Species Richness
and Shared Species
from Samples

Home | About EstimateS | EstimateS Download | EstimateS User's Guide | Support | EstimateS & Biota | Links

About EstimateS
Major Features
Platforms
Who Uses EstimateS?
Publications That Cite EstS

EstimateS Download
EstS 9.1 Windows
EstS 9.1 Mac OS
EstS 8.2 Windows
EstS 8.2 Mac OS

EstimateS User's Guide
Introduction
Preparing a Data Input File
Running EstimateS
Additional Notes
What EstimateS Computes
Things You Should Know
References
Appendices

Support
Support by Email
Bug Reporting
Known Bugs
Notes for Mac OS
Notes for Windows
Your Suggestions
Recent Version History

EstimateS & Biota

Links

9 May 2019

Current Version
EstimateS 9.1.0

IMPORTANT: Operating system limitations of EstimateS 9.1.0 (2019)

- ***MacOS.*** This version does not operate fully under MacOS 10.12 (Sierra), nor under subsequent versions (High Sierra, Mojave....).

- **Windows**. This version does not operate fully under Windows 10, nor probably under any subsequent Windows operating systems.

- **"What shall I do?"** Anne Chao's website offers excellent R-based tools to compute most of what EstimateS computes (and much more!). You can download the R code, but you need not know any R to use most of Anne's tools, as they are also implemented in easy-to-use, on-screen input/output tools. There is a comprehensive downloadable User's Guide.

  http://chao.stat.nthu.edu.tw/wordpress/software_download/

- **"Why can't you fix it?"** I first builit EstimateS more than twenty years ago, long before open source software (like R) was available. So it has always run under 4th Dimension, a proprietary development environment. To make it run under current Mac and Windows operating systems would require purchasing about USD$5000 worth of 4D development tools and licenses. I am now retired from teaching and living on a fixed income, so that is not practical (unless some generous benefactor would like to fund it). Anne Chao's approach is the way to go, for now and the future, and Anne has always been my key statistical advisor for EstimateS. ~Robert Colwell

Major new features of EstimateS 9 (2016)

- ***A comprehensively revised User's Guide*** covering all the new features in EstimtateS 9 and all the traditional ones of previous versions.

- An entirely new capability for handing ***individual-based rarefaction with true (unconditional) confidence intervals*** (and of course, sample-based rarefaction, the core of all previous versions of EstimateS).

- ***Rarefaction of richness estimators and diversity indices for individual-based data*** (as well as sample-based data, as in previous versions).

- Non-parametric ***extrapolation of rarefaction curves*** for both sample-based and individual-based data (Colwell et al. 2012).

- ***Batch input and export*** option for both sample-based and individual-based datasets.

- Options for computing, displaying, and exporting subsets of results for ***evenly-spaced intervals*** for rarefaction and extrapolation of samples or individuals (interval-sampling or *knots*).
- Automatic support for ***International and US number formats.***
- **EstimateS 9 for Windows** runs under Windows 8, Windows 7, Vista, and XP.
- **EstimateS 9 for Mac OS** runs under OS 10.5 (Leopard) through 10.8 (Mountain Lion).
- **EstimateS 9 is blazing fast,** compared to earlier versions.

With these classic features of previous versions

- Computes Colwell and Mao's smooth species accumulation curves (sample-based rarefaction curves) with true confidence intervals, based on analytical formulas (Colwell et al. 2004).
- Computes a wide range of species richness estimators for sample-based abundance and incidence (presence/absence) data (Chao, Jackknife, ICE, ACE and others).
- Computes asymmetric (log-transformed) confidence intervals for Chao1 andChao2.
- Computes diversity indices (Shannon, Simpson, Fisher's alpha, Hill numbers).
- All richness estimators and diversity indices are computed for every level of sample accumulation, averaged over resamplings.
- Choice of resampling with or without replacement.
- Computes Chao's shared species estimator for sample pairs.
- Computes Chao's Sørensen and Jaccard similarity estimators (Chao et al. 2005).
- Computes classic Jaccard, Sørensen, Bray-Curtis, and Morista-Horn similarity indices for sample pairs.

Click here to download EstimateS 9.0.0 now.

Homepage references (many more in the User's Guide)

Colwell, R. K., A. Chao, N. J. Gotelli, S.-Y. Lin, C. X. Mao, R. L. Chazdon, and J. T. Longino. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation, and comparison of assemblages. *Journal of Plant Ecology* **5**:3-21. *Read it online* or *download pdf.*

Colwell, R. K., C. X. Mao, & J. Chang. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* **85**, 2717-2727. *Download pdf.* Spanish Version: *Download pdf.*

Chao, A., R. L. Chazdon, R. K. Colwell, and T.-J. Shen. 2005. A new statistical approach for assessing compositional similarity based on incidence and abundance data. *Ecology Letters* **8**:148-159. *Download pdf.* Spanish Version: *Download pdf.*

EstimateS

Statistical Estimation of Species Richness and Shared Species from Samples

Home | About EstimateS | EstimateS Download | EstimateS User's Guide | Support | EstimateS & Biota | Links

About EstimateS
Major Features
Platforms
Who Uses EstimateS?
Publications That Cite EstS

EstimateS Download
EstS 9.1 Windows
EstS 9.1 Mac OS
EstS 8.2 Windows
EstS 8.2 Mac OS

EstimateS User's Guide
Introduction
Preparing a Data Input File
Running EstimateS
Additional Notes
What EstimateS Computes
Things You Should Know
References
Appendices

Support
Support by Email
Bug Reporting
Known Bugs
Notes for Mac OS
Notes for Windows
Your Suggestions

EstimateS & Biota

Links

About EstimateS

EstimateS is a free software application for Windows and Macintosh operating systems that computes a variety of biodiversity statistics, estimators, and indices based on biotic sampling data. Some features require species relative abundance data, others only species presence/absence data. A comprehensive *User's Guide* is available online at this site.

Major Features of EstimateS 9

Nonparametric rarefaction and extrapolation of species accumulation curves

- Computes expected species accumulation curves for ***rarefied*** reference samples (both sample-based ***rarefaction*** curves and individual-based ***rarefaction*** curves, in the terminology of Gotelli & Colwell 2001, 2011), with 95% unconditional confidence intervals, using the analytical formulas of Colwell et al. (2004) and Colwell et al. (2012).
- Computes expected species accumulation curves for ***augmented*** reference samples (both sample-based ***extrapolation*** curves and individual-based ***extrapolation*** curves), with 95% unconditional confidence intervals, using the analytical formulas of Colwell et al. (2012).

Species richness estimators and diversity indices

- Computes non-parametric, asymptotic species richness estimators for abundance-based data (Chao 1 and ACE) and for incidence-based data (Chao2, ICE, first and second order Jackknives, and Bootstrap).
- For Chao1 and Chao 2, computes log-linear confidence intervals.
- Computes Shannon (both *H* and *e^H*) and inverse Simpson (Hill numbers 1 and 2), and Fisher's alpha diversity indices.
- All estimators and indices are computed for rarefied samples or individuals to reduce the bias of comparisons.

Shared species estimation and biotic similarity indices, with estimators

- Computes Chao's estimator of total shared species (including observed and unseen shared species) for sample pairs.
- Computes classic Sørensen and Jaccard similarity indices for sample pairs based on incidence (presence-absence) or abundance data sets
- Computes Chao's abundance-based Sørensen and Jaccard similarity indices, with statisically rigorous estimators and their standard errors, for sample pairs based on abundance or on replicated incidence (presence-absence) data sets.
- Computes Morisita-Horn and Bray-Curtis (= "Sørensen quantitative") similarity indices for sample pairs based on abundance data sets.

Importing data and exporting results

- For input, accepts tab-delimited, plain text files (produced by programs such as Excel, using "Save As Tab-delimited Text"), in either of two matrix formats (species-by-sample or , sample-by-species) or either of two "triplet" formats ([species index, sample index, value]; or [sample index, species index, value]).
- For input, accepts pre-formatted sample-by-species matrices exported by *Biota: The Biodiversity Database Manager*.
- Exports all results as tab-delimited text files, easily opened in Excel, statistical, or graphing applications.
- As an option, records and exports results from *n* individual randomizations to a text file, allowing computation of precision, accuracy, and other analyses (Walther and Moore 2005), using Excel or other applications.

Platforms

Two versions of EstimateS are currently offered.

- EstimateS 9.00 runs under Windows 8 and Mac OS 10.8 and earlier operating systems and has many new features.
- EstimateS 8.20 runs under Windows 7 and MacOS 10.6 and earlier operating systems and maintains the traditional interface and tools that EstimateS has offered since 2005.

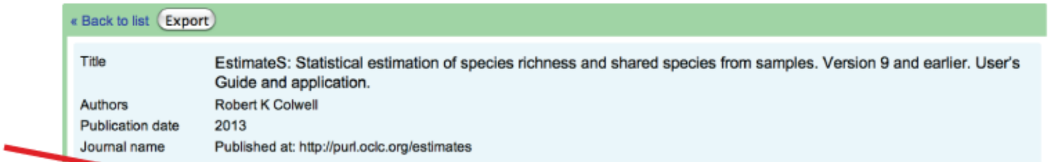
Who Uses EstimateS?

Between 1997 and 2013, EstimateS was downloaded by more than 60,000 users in more than 100 countries around the globe. They use EstimateS in conservation biology, biotic inventories, natural area assesssment and monitoring, biogeography, macroecology, microbial ecology, molecular biology, ethnobotany, archaeology and many other fields. You can get an idea of the diversity of uses by scanning the titles of the references that appear below.

Publications That Cite EstimateS

As of June, 2013, more than 4400 publications were known to have cited EstimateS. You can see a histogram of EstimateS citations, by year, on Google Scholar. On the Google Scholar page, click the "Cited by...." link (red oval, below) to see the complete, searchable list. Because of exsitence of this constantly updated list, we no longer list individual papers here.

However, if you have cited EstimateS in a published work that is *not* listed by Google Scholar in this list, and you would be willing to have it included, please contact Robert K. Colwell with details.



Title	EstimateS: Statistical estimation of species richness and shared species from samples. Version 9 and earlier. User's Guide and application.
Authors	Robert K Colwell
Publication date	2013
Journal name	Published at: http://purl.oclc.org/estimates

Total citations: Cited by 4406

Citations per year

Scholar articles

EstimateS.
RK Colwell - 1997
Cited by 4406 - Related articles

EstimateS: Biodiversity Estimation Software
RK Colwell - 2006

EstimateS

Statistical Estimation of Species Richness and Shared Species from Samples

Home | About EstimateS | EstimateS Download | EstimateS User's Guide | Support | EstimateS & Biota | Links

About EstimateS
Major Features
Platforms
Who Uses EstimateS?
Publications That Cite EstS

EstimateS Download
EstS 9.1 Windows
EstS 9.1 Mac OS
EstS 8.2 Windows
EstS 8.2 Mac OS

EstimateS User's Guide
Introduction
Preparing a Data Input File
Running EstimateS
Additional Notes
What EstimateS Compute
Things You Should Know
References
Appendices

Support
Support by Email
Bug Reporting
Known Bugs
Notes for Mac OS
Notes for Windows
Your Suggestions
Recent Version History

EstimateS & Biota

Links

Download EstimateS 9.1.0

EstimateS 9.1.0 for Windows: Downloading and Installation Instructions

For Windows 10, Windows 8, Windows 7, Vista, and XP

1. Hold down the right mouse button and click on the link below.

   Download EstimateS 9.1.0 for Windows (56 MB zip)

2. Choose "Save Target As" from the popup menu that appears.
3. Place and save the zipped file where you would like to run EstimateS (e.g. the Programs directory, but you can put it anywhere you like).
4. Double-click the zipped file to unpack EstimateS (it may happen automatically), which will be inside a folder called "EstimateS 9 Windows."
5. Delete the zip file if you wish.
6. Launch EstimateS from the *EstimateSWin910.exe* file in the EstimateS 9 Windows folder. You can make a shortcut and move it elsewhere it you want.
7. See the *EstimateS User's Guide* (at this website) for instructions on using EstimateS.

EstimateS 9.1.0 for Mac OS: Downloading and Installation Instructions

For OS 10.9 (Mavericks), 10.8 (Mountain Lion), OS 10.7 (Lion), OS 10.6 (Snow Leopard), and OS 10.5 (Leopard)

1. Hold down the Control key and click on the link below.

   Download EstimateS 9.1.0 for Mac OS (63 MB zipped)

2. Choose "Download Linked File As..." from the popup menu that appears.
3. Place and save the zipped file where you would like to run EstimateS (e.g. the Applications folder, but you can put it anywhere you like).
4. Double-click the zipped file to unpack EstimateS (it may happen automatically), which will be inside a folder called "EstimateS 9 Mac."
5. Delete the zip file if you wish.
6. Launch EstimateS by double-clicking its icon inside the Estimates 9 Mac folder, or drag the EstimateS icon to the Dock to create an alias in the Dock and launch it there.
7. See the *EstimateS User's Guide* (at this website) for instructions on using EstimateS.

Download EstimateS 8.2.0

EstimateS 8.2.0 for Windows: Downloading and Installation Instructions

For Vista, Windows XP, Windows 2000 and most installations of Windows 7

1. Hold down the right mouse button and click on the link below.

   Download EstimateS 8.2.0 for Windows (4.9 MB)

2. Choose "Save Target As" from the popup menu that appears.
3. Create and name a new, temporary folder (e.g. "EstimateS temp") on the Desktop or elsewhere. (You will delete it later.)
4. Save the downloaded file (SetupEstimateSWin800.exe) in the new folder.
5. Launch the installer (SetupEstimateSWin800.exe) by (left) double-clicking its icon. The installer places EstimateS in the Programs folder (by default), and creates a shortcut in the Programs menu (Start --> Programs).
6. Launch EstimateS from the shortcut in the Programs menu or from its icon in the Estimates folder inside the Programs folder.
7. You may delete the temporary folder (e.g. "EstimateS temp") that you created in Step 3.
8. See the *EstimateS User's Guide* (at this website) for instructions on using EstimateS.

EstimateS 8.2.0 for Mac OS: Downloading and Installation Instructions
For OS 10.2 to 10.6 only. Will not run under later systems

1. Hold down the Control (Ctl) key and click on the link below.

   Download EstimateS 8.2.0 for Mac OS (21.3 MB)

2. Choose "Download Linked File" from the popup menu that appears.
3. In your Downloads folder, launch the installer (InstallEstimates800Mac) by double-clicking its icon. The installer places EstimateS in the Applications folder (by default).
4. Launch EstimateS from the its icon in the Estimates folder inside the Applications folder, or drag its icon to the Dock to create an alias in the Dock and launch it there.
5. You may delete the installer (InstallEstimates800Mac).
6. See the *EstimateS User's Guide* (at this website) for instructions on using EstimateS.

EstimateS

Statistical Estimation
of Species Richness
and Shared Species
from Samples

Home | About EstimateS | EstimateS Download | EstimateS User's Guide | Support | EstimateS & Biota | Links

About EstimateS
Major Features
Platforms
Who Uses EstimateS?
Publications That Cite EstS

EstimateS Download
EstS 9.1 Windows
EstS 9.0 Mac OS
EstS 8.2 Windows
EstS 8.2 Mac OS

EstimateS User's Guide
Introduction
Preparing a Data Input File
Running EstimateS
Additional Notes
What EstimateS Computes
Things You Should Know
References
Appendices

Support
Support by Email
Bug Reporting
Known Bugs
Notes for Mac OS X
Notes for Windows
Your Suggestions

EstimateS & Biota

Links

EstimateS 9.1.0 *User's Guide*

Last Revised June 14, 2013

Website: http://purl.oclc.org/estimates or http://viceroy.eeb.uconn.edu/estimates

Table of Contents

Introduction
  Samples and Species, Abundance and Incidence
  Single and Multiple Datasets
  The Fundamental Design of EstimateS: Diversity
  The Fundamental Design of EstimateS: Shared Species and Similarity

Preparing a Data Input File for EstimateS
  EstimateS Filetypes: The Load Data Input Screen
    The Four Input Filetypes
      Filetype 1. *Sample-based incidence or abundance data: One set of replicated sampling units* (classic EstimateS input)
      Filetype 2. *Sample-based incidence or abundance data: Multiple sets of replicated sampling units* (batch input).
      Filetype 3. *Individual-based abundance data: One individual-based abundance sample*
      Filetype 4. *Individual-based abundance data: Multiple individual-based abundance samples* (batch input).

  Data Input Formats
    The Five Data Input Formats
      Format 1. *Species (rows) by Samples(columns)*
      Format 2. *Samples (rows) by Species (columns)*
      Format 3. (Sample-based filetypes only). *Species, Sample, Abundance triplets*
      Format 4. (Sample-based filetypes only). *Sample, Species , Abundance triplets*
      Format 5. (Sample-based filetypes only). *Biota format.*

Running EstimateS
  Loading the Data Input File

  Setting and Running the *Diversity* Options
    The Randomization and Rarefaction Tab
      Sample order randomization for estimators and indices
      Extrapolation of rarefaction curves (richness only)
      Estimation points (knots) for rarefaction and extrapolation
    The Estimators and Indices Tab
      Diversity indices (Fisher's alpha, Shannon, Simpson)
      Chao1 and Chao2 bias correction
      Coverage-based estimators (ACE, ICE, Shared Species)
      Randomization protocol for estimators and indices
    The Other Options Tab
      Individual run export
      Random number generator for randomizations
      Individual shuffing (Sample-based filetypes only)
      Settings usage (Saving settings)
    Launching the Computations
    Exporting the Results

- Setting and Running the *Shared Species* Options (Sample-based filetypes only)
  - Coverage-based Estimators (ICE, ACE, Shared Species)
  - Similarity Indices and Estimators
  - Settings usage (Saving settings)
  - Launching the Computations
  - Exporting the Results

Additional Notes
- Comparing Species Accumulation Curves: Rarefaction and Extrapolation of Reference Samples
  - Rarefaction and extrapolation
  - Confidence intervals for rarefaction and extrapolation
  - Statistical inference
  - Comparing sample-based abundance data
  - Rarefaction and extrapolation vs. asymptotic richness estimation

- Asymptotic Richness Estimators
  - Chao1 and Chao2 richness estimators
  - Coverage-based richness estimators ICE and ACE

- Estimating total Species Richness by Functional Extrapolation (Sample-based Filetypes only)

- Indices of Species Diversity and Hill Numbers

- Non-integer Sampling Data (Percent Cover, Basal Area, Biomass, etc) and EstimateS

What EstimateS 9 Computes
- *Table 1*: Diversity Statistics
- *Table 2*: Shared Species Statistics

Things You Should Know Before You Begin
- Caveat Receptor
- Citing EstimateS
- What You Must Agree To: Copyright and Fair Use
- Sharing EstimateS With Others

References Cited

Appendices
- *Appendix A:* Control Parameters for Automated Input
- *Appendix B:* Nonparametric Estimators of Species Richness
- *Appendix C:* Coverage-Based Estimators of Shared Species
- *Appendix D*: Chao's Abundance-based Jaccard and Sorensen Similarity Indices

---

Introduction

EstimateS 9 is a free software application for Windows and Macintosh operating systems, designed to help you ass
and compare the diversity and composition of species assemblages based on sampling data. EstimateS computes
variety of biodiversity statistics, including rarefaction and extrapolation, estimators of species richness, diversity indi
Hill numbers, and similarity measures. *For an overview of major features, **click here.***

Samples and Species, Abundance and Incidence

In this Guide, the term *sample* refers to any list of *species* or other taxa from a locality, site, quadrat, trap, time unit, clone library, or some other entity.

Some estimators and indices require *counts of individuals* (or gene copies) for each species in single sample, or in each of a set of samples. Such data are called *individual-based abundance data*. Other estimators and indices requ
only presence/absence (occurrence) data for each species in each a set of related (or replicate) samples. (These related samples are sometimes called *sampling units* in the literature, and in this Guide.) Such data are called *samp*
based incidence data. When a dataset consists of species abundance data for a set of related samples (*sample-bas*
abundance data), the dataset can be treated, sample-by-sample, as individual-based abundance data, or converted
sample-based incidence data.

When comparing the biotic (species or higher taxa) similarity of two or more localities (or habitats, treatments, seas
etc.), you can do so either using abundance data or by using summed incidence data (frequencies of occurence, pooled among samples) for each or two or more sample sets. More information ***here***.

Single and Multiple Datasets

EstimateS 9 allows you to analyze either a *single dataset* or a *multiple datasets*, one after another, in a single data input file (batch input). *Each* dataset may consist of either individual-based abundance data (a single sample of abundance data) or sample-based incidence or abundance data (several related samples of incidence or abundanc data). More information ***here***.

The Fundamental Design of EstimateS: Diversity

EstimateS helps you account for the inevitable confounding effects of sample size (or sampling effort) on biodiversit data by several different strategies. Consider a *reference sample*: either a single, individual-based abundance samp of *n* individuals, or a set of *t* related sampling units for which incidence data have been recorded.

Richness estimators. Based on a reference sample (as defined above), EstimateS computes several widely-used statistical estimators of asymptotic species richness, the true number of species in the assemblage sampled. These estimators aim to reduce the effect of undersampling, which inevitably biases the observed species count. More information ***here***.

Rarefaction. Rarefaction is a resampling framework that selects, at random, 1, 2, ..., *n* individuals or i= 1, 2, ..., *t* sampling units (generally without replacement) until all individuals or sampling units in the reference sample have be accumulated. For each level of rarefaction, EstimateS computes a large number of biodiversity statistics. For specie richness, exact analytical methods are used to compute the expected number of species (and its *unconditional* standard deviation) for each level of accumulation. For other diversity measures, EstimateS resamples individuals o sampling units stochastically, based on a random-number-driven algorithm. The resampling process is repeated ma times, and the means (and conditional standard deviations) among resamples for each level of accumulation are reported. The effects of differences in sample size on diversity statistics for two or more samples can usually be substantially reduced by comparing a the same level of species accumulation. More information ***here***.

Extrapolation. Rarefaction, in effect, represents an interpolation between the value of a diversity measure assesse for the reference sample and zero (for individual-based abundance data) or the diversity of a typical single sampling unit (for sample-based indidence data). For species richness (only) EstimateS 9 introduces extrapolation from a reference sample to the expected richness (and its *unconditional* standard deviation) for a user-specified, augmente number of individuals or sampling units. The methods that EstimateS uses for richness extrapolation rely on statistic sampling models, not on fitting mathematical functions. They require an estimator for asymptotic richness as a "targ for the extrapolation; EstimateS uses Chao1 for individual-based abundance data and Chao2 for sample-based incidence data. More information ***here***.

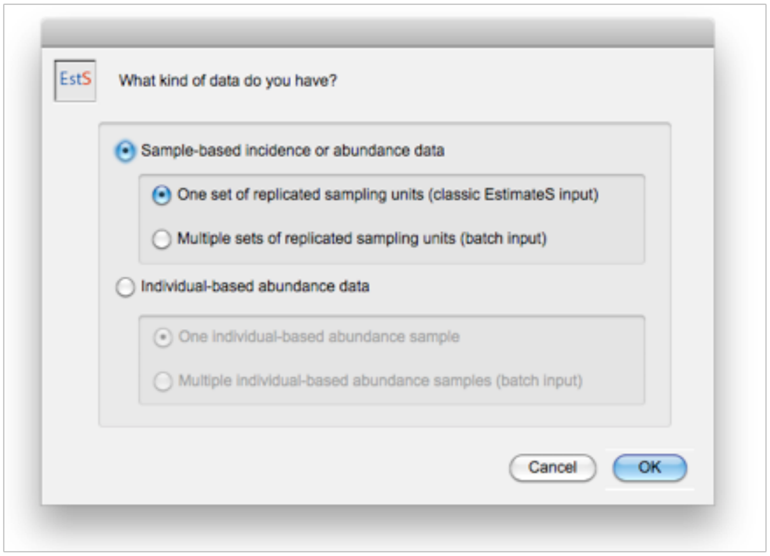
The Fundamental Design of EstimateS: Shared Species and Similarity

For sets of related sampling units, EstimateS computes several measures of compositional similarity, including traditional similarity indices as well as estimators of shared species and similarity indices that take shared, but unobserved species into account by statisical methods. These latter methods require species abundance data for a of related samples (sample-based abundance data) or for summed incidence data for two or more sets of sampling units. More information ***here***.

Preparing a Data Input File for EstimateS

EstimateS *Filetypes*: The Load Data Input File Screen

In *EstimateS* 9, the Load Data Input File command (from the File menu) presents a set of four *filetype* options. Spec *data input formats* for these filetype options are discussed in *a later section* of this Guide. This section describes the four filetypes and their uses.



What kind of data do you have?

● Sample-based incidence or abundance data
  ● One set of replicated sampling units (classic EstimateS input)
  ○ Multiple sets of replicated sampling units (batch input)
○ Individual-based abundance data
  ● One individual-based abundance sample
  ○ Multiple individual-based abundance samples (batch input)

Essential Notes

- All Input Files in EstimateS must be in *tab-delimited plain text* (sometimes called tab-separated-values, or TS Excel files cannot be read. Save them first as *tab-delimited text*.
- The Input File may have any name and may be located in any folder (directory).
- In the specifications below, required entries are indicated in ***italics.***
- In the specifications below, optional entries are indicated in **[square brackets].**
- For *each* filetype, two *example files* are installed with EstimateS, one with a symbolic definition of the filetype and the other with a numerical example.

The Four Input Filetypes

Filetype 1. Sample-based incidence or abundance data: One set of replicated sampling units (classic EstimateS input).

Example files: *Single-Sample-Based-Data-Filetype.xlsx* and *Single-Sample-Based-Data-Example.txt*

An additional sample input file for this filetype, named *Seedbank.txt*, is also installed with EstimateS. The Seedbank dataset (Butler & Chazdon 1998) is a classic benchmark dataset, used to compute the species richness estimators for that appear in Figures 1 and 2 and Table 1 of Colwell & Coddington (1994).

This default filetype supports *sample-based data* in a single incidence (or abundance) dataset (the classic *EstimateS* input format from Versions 8.2 and earlier). Just as in previous versions of *EstimateS,* if the data are abundance counts, they are converted to presence/absence (incidence) data for richness rarefaction and extrapolation, and to computeincidence-based richness estimators (e.g. Chao2, ICE) and similarity measures (e.g. classic Jaccard and Sørensen). Counts are treated as ordinary abundance data for abundance-based richness estimators (e.g. Chao1, ACE), diversity indices (e.g. Shannon, Simpson), and similarity measures (e.g. Chao-Jaccard, Chao-Sørensen, Morisita-Horn).

The two required header records (rows) for this filetype are:

Record #1 (Title Record): *Datafile Title* <tab> [*SampleSet*] <tab> [Format Code] <tab> [Skip rows] <tab> [Skip columns]

The first record (line) of the Input File must contain a title in the first field (column); any text will do. The second field of this Title Record should read (exactly) **SampleSet** (including the asterisks; no spaces). For compatibility with EstimateS 8.2 and earlier, a blank in the second column of the Title Record will be interpreted as reading **SampleSet**. Additional fields for the Format Code (*see later*), the number of header rows to skip, and the number of header columns to skip are all optional for this filetype, since they can be chosen onscreen when the file is being loaded.

Record #2 (Parameter Record): *Number of Species*<tab>*Number of Sampling Units in the Sample Set*

The second record (line) of the Input File must contain two obligatory control parameters: the number of species and the number of sampling units, separated by a <tab> character. Additional execution *control parameters* are optional, and can be more easily recorded by exporting a new copy of the input file after setting the parameters in EstimateS' Settings screens.

Record #3 etc.: The rest of the Input File contains the input data, which can appear in any one of *five* alternative formats.

2. Sample-based incidence or abundance data: Multiple sets of replicated sampling units (batch input).

Example files: *Multiple-Sample-Based-Data-Filetype.xlsx* and *Multiple -Sample-Based-Example.txt*

The second sample-based filetype supports batch input of multiple incidence or abundance datasets.

The first record in the Input File, the *Batch Record*, indicates that multiple datasets are expected, specifies the number of datasets it includes, and (optionally) names the batch:

Record #1 (Batch Record): _*MultipleSampleSets*_ <tab> *Number of Datasets* <tab> [BatchTitle]

After the Batch Record, the datasets simply follow one after the other in the Input File, with no empty records separating them.

Each dataset must be prepared exactly as specified for *single* sample-based dataset input (previous section), but the second field (*_SampleSet_*) and the third field (Format Code) in the Title Record are *required* for each dataset. The entries for [Skip rows] and [Skip columns] must be specified for each dataset if either is non-zero. The skip parameters will be interpreted as zeroes if omitted:

Title Record: *Datafile Title* <tab> _*SampleSet*_ <tab> *Format Code* <tab> [Skip rows] <tab> [Skip columns]

The Parameter Record is exactly as specified for single sample-based dataset input:

Record #2 (Parameter Record): *Number of Species* <tab> *Number of Sampling Units in the Sample Set*

In batch input, execution control parameters in subsequent columns of this record may be present, *but are ignored.* In batch mode, you chose the analysis options you want (once) from the graphical interface. The options you choose then apply to all datasets in the batch.

3. Individual-based abundance data: One individual-based abundance sample.

Example files: *Single-Individual-Based-Data-Filetype.xlsx* and *Single-Individual-Based-Example.txt*

This first filetype for individual-based abundance data supports input for a single list (vector) of species abundances.

Individual-based rarefaction of abundance data is new in EstimateS 9. (*Coleman rarefaction* of *sample*-based abundance data, long available in earlier versions of EstimatedS, is a close approximation, but applies to richness only and lacks a proper, *unconditional* variance estimator.)

In addition to rarefaction and extrapolation for species richness (with *unconditional confidence intervals*), this input option can be used compute abundance-based richness estimators (e.g. Chao1, ACE) and diversity indices (e.g. Shannon, Simpson) for rarefied subsets of individuals.

The two required header records (rows) for this filetype are:

Record #1 (Title Record): *Datafile Title* <tab> _*Individuals*_ <tab> [Format Code] <tab> [Skip rows] <tab> [Skip columns]

The first record (line) of the Input File must contain a title in the first field (column); any text will do. The second field of the Title Record *must* read (exactly) _*Individuals*_ (including the asterisks; no spaces). Additional fields for the Format Code (*see later*), the number of header rows to skip, and the number of header columns to skip are all optional for this filetype, since they can be chosen onscreen when the file is being loaded.

Record #2 (Parameter Record): *Number of Species* <tab> *Number of Samples (1)*

The second record (line) of the Input File must contain two obligatory control parameters: the number of species and the number of samples, separated by a <tab> character. *The number of samples for this Filetype is always 1.* There are no additional control parameters for this filetype.

Record #3 etc.: The rest of the Input File contains the input data, which can appear in either of *two alternative formats (Format 1 or Format 2).*

4. Individual-based abundance data: Multiple individual-based abundance samples (batch input).

Example files: *Multiple-Individual-Based-Data-Filetype.xlsx* and *Multiple-Individual-Based-Example.txt*

This second individual-based filetype supports batch input of multiple incidence or abundance datasets.

The first record in the Input File, the *Batch Record*, indicates that multiple datasets are expected, specifies the number of datasets it includes, and (optionally) names the batch:

Record #1 (Batch Record): **MultipleIndividuals** <tab> *Number of Datasets* <tab> [BatchTitle]

After the Batch Record, the datasets simply follow one after the other in the Input File, with no empty records separating them.

Each dataset must be prepared exactly as specified for *single* individual-based dataset input (previous section), but both the second field (**Individuals**) and the third field (*Format Code*) in the Title Record are *required* for each dataset. The entries for [Skip rows] and [Skip columns] must be specified for each dataset if either is non-zero. The skip parameters will be interpreted as zeroes if omitted:

Title Record: *Datafile Title* <tab> **Individuals** <tab> *Format Code* <tab> [Skip rows] <tab> [Skip columns]

The Parameter Record is exactly as specified for single individual-based dataset input:

Record #2 (Parameter Record): *Number of Species*<tab>*Number of Samples (1)*

The second record (line) of each Input File must contain two obligatory control parameters: the number of species and the number of samples, separated by a <tab> character. *The number of samples for this Filetype is always 1.*

There are no additional control parameters for this filetype. In batch mode, you chose the analysis options you want (once) from the graphical interface. The options you choose then apply to all datasets in the batch.

Data Input Formats

The Five Data Input Formats

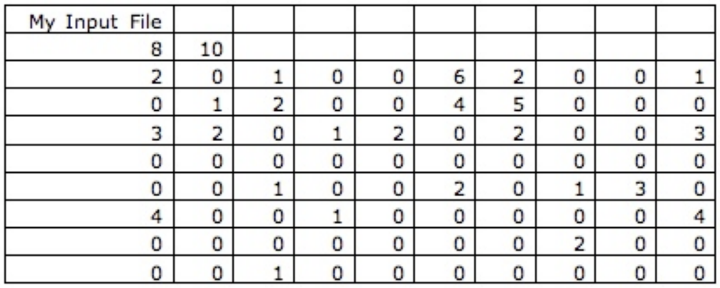
Once you have determined which of the four *input filetypes* you will be using, you need to decide which *data input format* you will use for the actual biodiversity data.

- Data input Formats 1 and 2 may be used with *any* input filetype.
- Data input Formats 3, 4, and 5 apply only to sample-based data (the first two *input filetypes*).

Format 1. Species (rows) by Samples(columns). For sample-based input filetypes, you will have one row for each species, one column for each sample. For individual-based input filetypes, you will have one row for each species i single column. The input file may contain any number of initial rows of column labels and/or initial columns of row labels, in which case you must tell EstimateS how many of each there are. (EstimateS simply skips over these specified label rows and columns.)

Note on Format 1: If your file includes one or more rows of column labels, they must *follow* the required Title and Parameter records and precede the data. If your file includes one or more columns of row labels, the required Title and Parameter records nonetheless begin in the *first* column.

Format 1 Example: Below is a simple example of an EstimateS sample-based Input File in Format 1, for a dataset called "My Input File" that includes data for 8 species (rows) in 10 samples (columns). The data are exactly the same as in the examples, below, for Formats 2 and 3. See the installed example files*Single-Sample-Based-Data-Filetype.xlsx*, *Single-Sample-Based-Example.txt,* and *Seedbank.txt*.



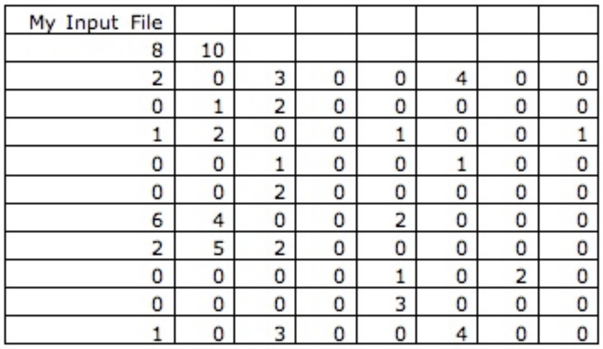
My Input File									
8	10								
2	0	1	0	0	6	2	0	0	1
0	1	2	0	0	4	5	0	0	0
3	2	0	1	2	0	2	0	0	3
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	2	0	1	3	0
4	0	0	1	0	0	0	0	0	4
0	0	0	0	0	0	0	2	0	0
0	0	1	0	0	0	0	0	0	0

Format 2. Samples (rows) by Species (columns). For sample-based input filetypes, you will have one row for each sample, one column for each species. For individual-based input filetypes, you will have one

column for each species in a single row. The input file may contain any number of initial rows of column labels and/or initial columns of row labels, in which case you must tell EstimateS how many of each there are. (EstimateS simply skips over these specified label rows and columns.)

Note on Format 2: If your file includes one or more rows of column labels, they must *follow* the required Title and Parameter records and precede the data. If your file includes one or more columns of row labels, the required Title and Parameter records nonetheless begin in the *first* column.

Format 2 Example: Below is a simple example of an EstimateS sample-based Input File in Format 2, for a dataset called "My Input File" that includes data for 8 species (columns) in 10 samples (rows). The data are exactly the same as in the example, above, for Format 1, and the example, below, for Format 3. See the installed example files *Single-Individual-Based-Data-Filetype.xlsx* and *Single-Individual-Based-Example.txt* for individual-based examples in Format 2.

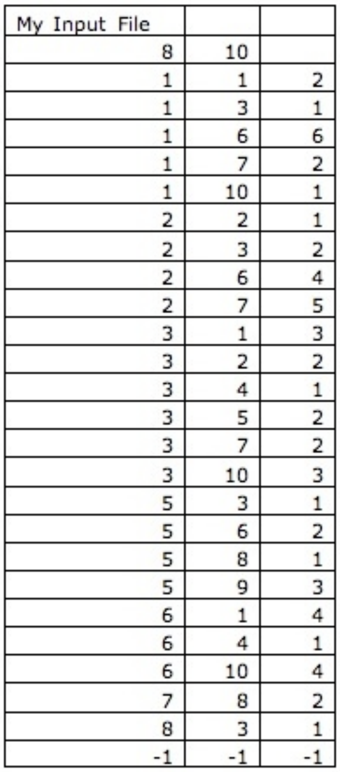


My Input File							
8	10						
2	0	3	0	0	4	0	0
0	1	2	0	0	0	0	0
1	2	0	0	1	0	0	1
0	0	1	0	0	1	0	0
0	0	2	0	0	0	0	0
6	4	0	0	2	0	0	0
2	5	2	0	0	0	0	0
0	0	0	0	1	0	2	0
0	0	0	0	3	0	0	0
1	0	3	0	0	4	0	0

Format 3 (Sample-based filetypes only). Species, Sample, Abundance triplets: the first column contains the speci number, the second the sample number, and the third the number of individuals (abundance) of that species in that sample. A final (extra) record with "-1" in each of these three columsn indicates end of input. This "triplet" format a common input format for statistical programs (e.g. SYSTAT.) You can list one row for every sample/species combination, or rows for only those combinations that have non-zero abundances. (The rest are automatically set to zero.) Using the triplet format and storing only non-zero abundance values requires far less file space than storing th full matrix. In fact, this may be the most practical way to store files larger than your spreadsheet will accept. As an option (see below), EstimateS can export a data matrix in this format, after reading it in using one of the other four formats listed here.

Note on Format 3: EstimateS expects no more than one record for each species **x** sample combination. If you have more than one, only the first is read. *A special record must terminate triplet files, with "-1" in each of these three colums to indicate end of input, as shown in the example below.*

Format 3 Example: Below is a simple example of an EstimateS Input File in Format 3, for a dataset called "My Input File" that includes data for 8 species (columns) in 10 samples (rows). The data are exactly the same as in the examples, above, for Formats 1 and 2.



My Input File		
8	10	
1	1	2
1	3	1
1	6	6
1	7	2
1	10	1
2	2	1
2	3	2
2	6	4
2	7	5
3	1	3
3	2	2
3	4	1
3	5	2
3	7	2
3	10	3
5	3	1
5	6	2
5	8	1
5	9	3
6	1	4
6	4	1
6	10	4
7	8	2
8	3	1
-1	-1	-1

Format 4 (Sample-based filetypes only). Sample, Species, Abundance triplets. The formatis just as for Format 3, the columns are ordered Sample, Species, Abundance.

> ***Note on Format 4:*** EstimateS expects no more than one record for each species x sample combination. If you have more than one, only the first is read. *A special record must terminate triplet files, with "-1" in each of these three colums to indicate end of input, as shown in the example, above.*

Format 5 (Sample-based filetypes only). Biota format. This format is output automatically by Biota, with appropria row and column labels. For other input files that include column or row labels, use Formats 1 or 2

Running EstimateS

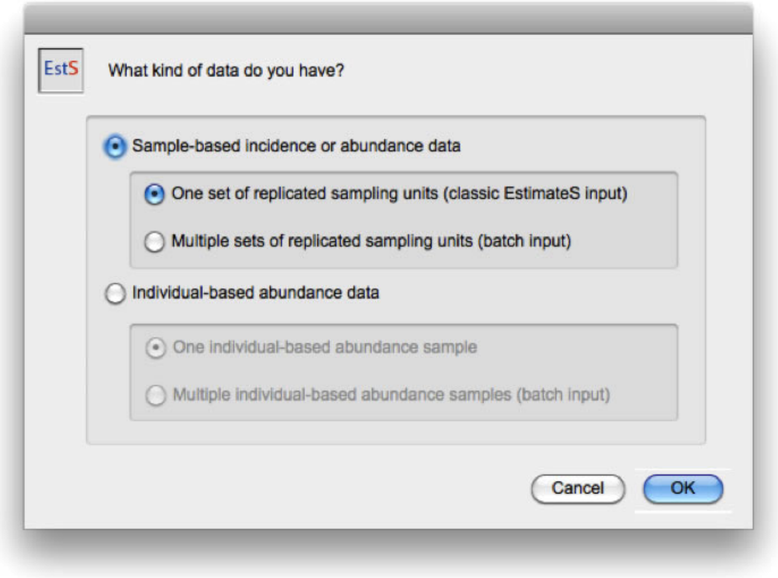
Loading the Data Input File

1. Launch EstimateS by double-clicking the EstimateS icon or application name (MacOS); or by launching Estimat from the Programs section of the Start menu or double-clicking the *EstimateS[version number].exe* file (Windows).

2. If a file navigation window appears asking you to select a "Data File," choose the file called ***Statistics.4DD***(Windows) or ***Statistics.data*** (Mac OS). This default file records the statistical **output** of Biota. *If you never see such a request, all the better! Just skip the rest of this step.* The Statistics file is of no practical use, but is required for EstimateS to function.

> ***Note 1:*** Do **not** try to load your **input** file at this point. If you cannot find the Statistic Data file, click the *New* button to create a new data output file. You can name it anything you wish, using the extension **.data** (Macintosh) or **.4DD** (Windows).
>
> ***Note 2:*** If you want to create a new output data file or find a different existing one, you can **force** the navigation window to appear as follows:
>
> > ***Windows:*** Select the EstimateS icon or application name, then choose open from the Windows File menu, while holding down the **Alt** key.
> >
> > ***Macintosh:*** Click and hold the **Option** key while launching EstimateS

2. **From the File menu in EstimateS, choose *Load Input File*.** The Filetype Selection dialog appears.



EstS What kind of data do you have?

● Sample-based incidence or abundance data
  - ● One set of replicated sampling units (classic EstimateS input)
  - ○ Multiple sets of replicated sampling units (batch input)

○ Individual-based abundance data
  - ● One individual-based abundance sample
  - ○ Multiple individual-based abundance samples (batch input)

First choose either the *Sample-based* or the *Incidence-based* option, then the appropriate *single dataset* or *multiple dataset* option. *Click here* to read about these four options. *Click here* learn how to prepare a Data Input File.

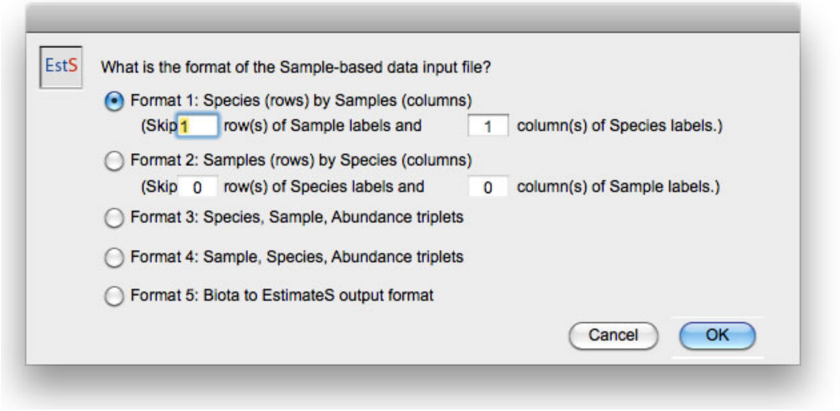
Note: If you are a first-time user of EstimateS, you might want use the default option ("One set of replicated sampling units") and choose the *Seedbank.txt* demonstation Data Input File that is installed with EstimateS, to explore the application.

3. **Click the OK button.** The Open File window appears.

4. **Find the Data Input File and open it.** What happens next depends upon which filetype you are loading.

a. For a single dataset, a confirmation screen appears, showing the parameter settings indicated by the Title Record and Parameter Record in the Data Input File (and default settings of several other parameters). Here is an example, for the default filetype option, "One set of replicated sampling units."

When you click the OK button, an input option diaolog appears, where you can indicate which Input Format you Data Input File uses, and tell EstimatesS how many (if any) rows of column headers and how many columns of row headers to skip. (The corresponding screen for the filetype "One individual-based abundance sample" is similar, but offers only two Format options.)



EstS What is the format of the Sample-based data input file?

● Format 1: Species (rows) by Samples (columns)
(Skip 1 row(s) of Sample labels and 1 column(s) of Species labels.)

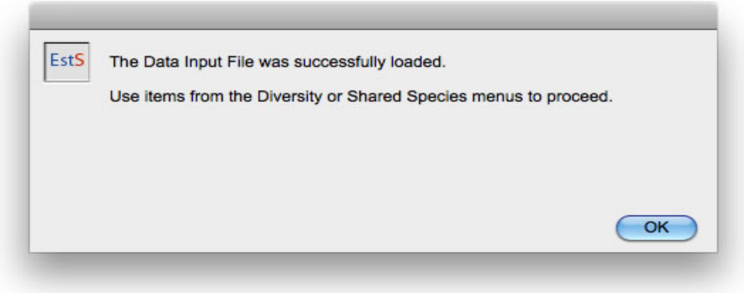
○ Format 2: Samples (rows) by Species (columns)
(Skip 0 row(s) of Species labels and 0 column(s) of Sample labels.)

○ Format 3: Species, Sample, Abundance triplets

○ Format 4: Sample, Species, Abundance triplets

○ Format 5: Biota to EstimateS output format

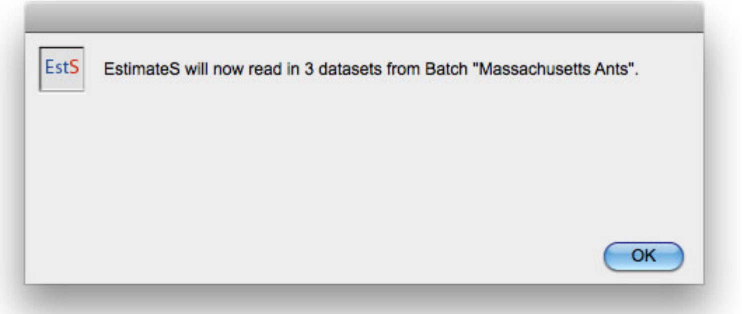
Once you are sure the settings are correct, click the *OK* button in the dialog. EstimateS completes the loading of the dataset and confirms that the file has been correctly loaded. (Input data errors will be flagged if they occur. Follow the onscreen instructions if this happens.)



The Data Input File was successfully loaded.

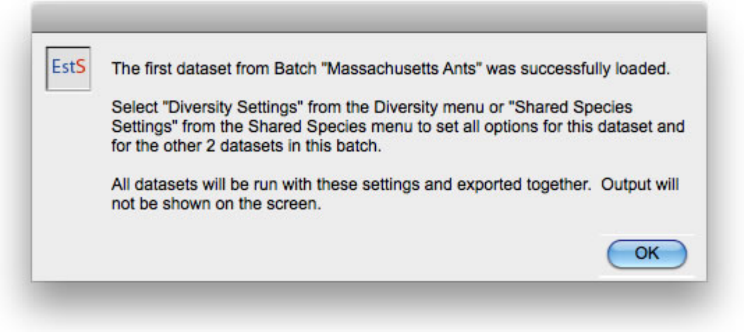
Use items from the Diversity or Shared Species menus to proceed.

b. For a batch (multiple) dataset, when you click OK in the Filetype Selection Dialog, a confirmation dialog appears with the Batch Name.



EstimateS will now read in 3 datasets from Batch "Massachusetts Ants".

When you click the OK button, a second confirmation diaolog appears, explaining that the datasets will be analyzed automatically and sequentially, after you set the analysis parameters for the first dataset. All datasets in the batch will be run with those parameter settings. Here is an example, for the sample-based batch option ("Multiple sets of replicated sampling units").



The first dataset from Batch "Massachusetts Ants" was successfully loaded.

Select "Diversity Settings" from the Diversity menu or "Shared Species Settings" from the Shared Species menu to set all options for this dataset and for the other 2 datasets in this batch.

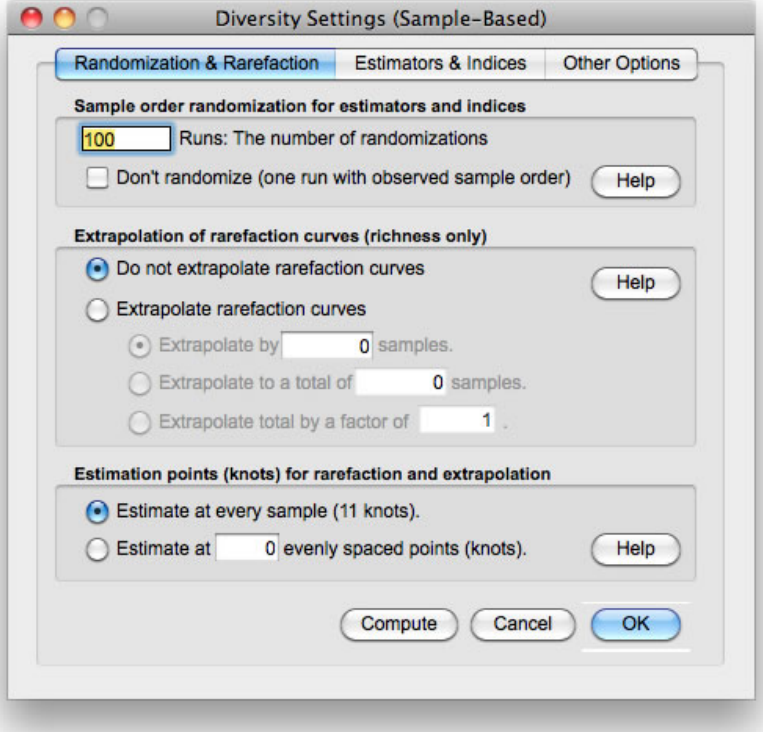
All datasets will be run with these settings and exported together. Output will not be shown on the screen.

Setting and Running the *Diversity* Options

Once the Data Input File has been loaded, you are ready to set or check the Diversity options (this section) and/or tl Shared Species options.

> ***Note:*** The Diversity Settings screensfor Sample-based and Individual-based filetypes are nearly identical. In this section the individual-based screen will be illustrated, with notes on differences in the abundance-based screen, where relevant.

1. From the Diversity menu, choose *Diversity Settings*. The Diversity Settings screen appears.



Sample order randomization for estimators and indices

100 Runs: The number of randomizations

☐ Don't randomize (one run with observed sample order)

Extrapolation of rarefaction curves (richness only)

● Do not extrapolate rarefaction curves

○ Extrapolate rarefaction curves

● Extrapolate by 0 samples.

○ Extrapolate to a total of 0 samples.

○ Extrapolate total by a factor of 1 .

Estimation points (knots) for rarefaction and extrapolation

● Estimate at every sample (11 knots).

○ Estimate at 0 evenly spaced points (knots).

The image above shows the default settings for the example Data Input File *Single-Sample-Based-Example.txt*, loaded immediately after launching EstimateS. Unless you indicate otherwise (in the *Other Options tab* of the Diversity Settings screen or in the Shared Species Settings screen), EstimateS will remember whatever settings you last used, and display those as the default, although they may be overridden by *Execution Control Parameters* in the Data Input File.Col

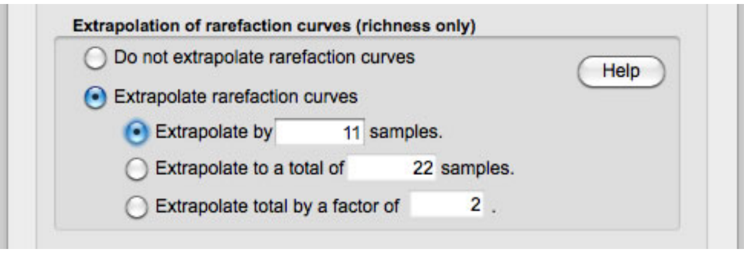
2. **Set the options on the *Randomization & Rarefaction* Tab** (illustrated above).

Sample order randomization for estimators and indices. *Runs* specifies the number of randomizations (resamples) to be carried out for rarefaction. If you want to evalulate asymptotic richness estimators or diversity indices at all levels of species accumulation (rarefaction) up to the size of the *reference sample*, you should choose a reasonable number of randomizations (100 is usually enough) to get smooth curves for the estimators and indices as a function of the number of samples (or individuals, for individual-based filetypes).

EstimateS computes rarefaction and extrapolation curves and their unconditional confidence intervals analytically, using the formulas of Colwell, Mao, & Chang (2004), Colwell et al. (2012), and Chao et al. (2013), for which no randomization is required or carried out. For sample-based rarefaction and extrapolation, EstimateS uses the Bernouilli prosucct model (Colwell et al. 2012). For individual-based rarefaction (beginning with EstimateS Version 9.1.0), computations follow the multinomial model for both rarefaction and the and extrapolation (Colwell et al. 2012).

Therefore, if all you want is a rarefaction curve, with or without extrapolation (no asymptotic richness estimators or diversity indices), check the *Don't randomize* checkbox .

Extrapolation of rarefaction curves (richness only). If you request *extrapolation from the reference sample*, by selection the "extrapolate rarefaction curves" option, EstimateS will estimate the expected number of species that would be found in an augmented sample using the nonparametric methods of (Colwell et al. (2012). Asymptotic richness estimators and diversity indices are *not* extrapolated.



Extrapolation of rarefaction curves (richness only)

○ Do not extrapolate rarefaction curves

● Extrapolate rarefaction curves

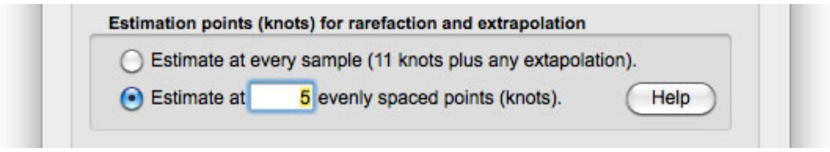
● Extrapolate by 11 samples.

○ Extrapolate to a total of 22 samples.

○ Extrapolate total by a factor of 2 .

You have three options (above) for specifying how far you wish to extrapolate the sample-based rarefaction curve beyond the size of the reference sample. You can: (1) augment the empirical sample set *by* a fixed number of samples, (2) augment the empirical sample set t*o a specified total* number of samples, or (3) augment the empirical sample set *by a specified factor* (e.g. 1.5x, 2x, 3x...). Extrapolation beyond doubling or tripling is not recommended, as the variance increases greatly.

Estimation points (knots) for rarefaction and extrapolation. EstimatesS gives you a choice between computing, displaying, and exporting rarefied (and extrapolated) richness, asymptotic richness estimators, and diversity indices for every sample increment (the classic EstimateS approach) or, instead, computing, displaying, and exporting these statsitics for a smaller number of sample increments, spaced at approximately even intervals along the rarefaction (and extrapolation) curve.



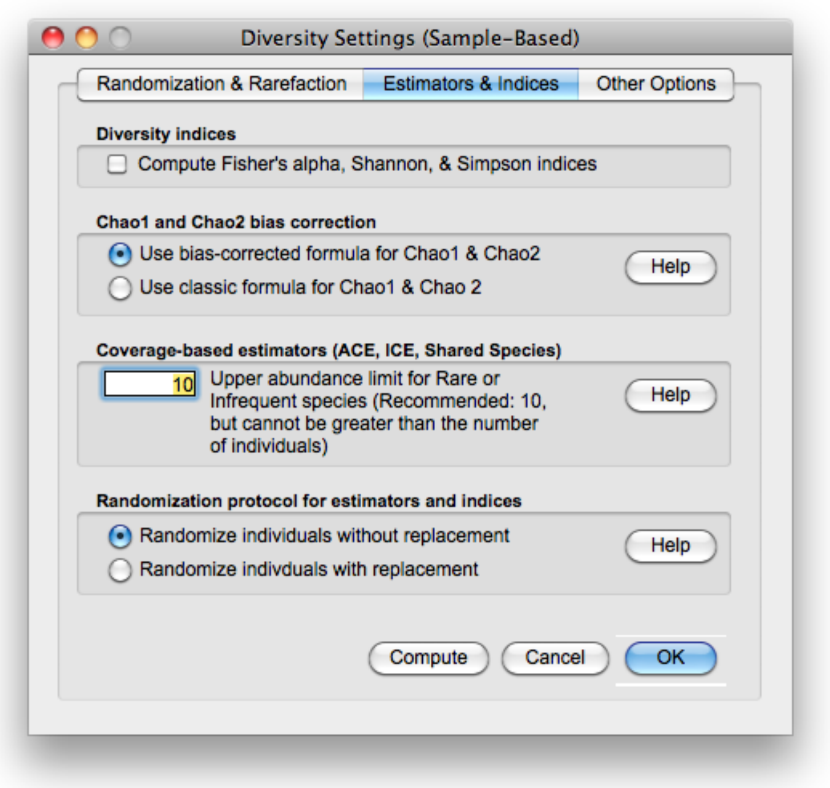
Estimation points (knots) for rarefaction and extrapolation

○ Estimate at every sample (11 knots plus any extapolation).

● Estimate at 5 evenly spaced points (knots).

The sampling points for the second approach are called "knots." EstimateS will always place a knot at the full reference (empirical) sample, even if the rarefaction curve is extrapolated. If you request extrapolation, a knot will be placed at the final sample of the extrapolated curve, as well. Because of these constraints and because knots must be integers, in many cases the spacing between knots will not be exactly even. If you don't like this, just choose the traditional option and compute, display, and export for every sample increment.

3. Set the options on the *Estimators and Indices* Tab.



Diversity Settings (Sample-Based)

Diversity indices

☐ Compute Fisher's alpha, Shannon, & Simpson indices

Chao1 and Chao2 bias correction

● Use bias-corrected formula for Chao1 & Chao2

○ Use classic formula for Chao1 & Chao 2

Coverage-based estimators (ACE, ICE, Shared Species)

10 Upper abundance limit for Rare or Infrequent species (Recommended: 10, but cannot be greater than the number of individuals)

Randomization protocol for estimators and indices

● Randomize individuals without replacement

○ Randomize indivduals with replacement

Diversity Indices (Fisher's alpha, Shannon, Simpson). By default, the C*ompute Fisher's alpha, Shannon, and Simpson indices* box is *unchecked*, so you must check it if you want these indices of diversity for rarefied subsamples of the *reference sample*. If you check this box, be sure to indicate multiple Runs (100 is suggested) on the Randomization and Rarefaction tab, so that the means among runs will produce a smooth rarefaction curve for the diversity indices.

EstimateS 9 computes *Shannon exponential, as well as the Shannon information statistic. Simpson diversity is computed in its inverse form.* Thus, EstimateS 9 computes the first three Hill numbers, for rarefied subsamples of the reference sample: *q* = 0 (richness), *q* = 1 (Shannon exponentia diversityl), and *q* = 2 (Simpson inverse diversity) (Jost 2006). Note that richness is computed analytically, whereas Shannon and Simpson diversities are computed by resampling.

Chao1 and Chao2 bias correction. By default, EstimateS uses the *bias-corrected form of the Chao1* and *Chao2* richness estimators in all cases (the recommended default). If you choose "Use classic formula for Chao1 and Chao2," instead, EstimateS uses the bias-corrected form only when either doubletons (Chao1) or duplicates (Chao2) are zero, and uses the approximate ("classic") formulas otherwise.

> ***Note:*** For some datasets (those with a coefficient of variation of the abundance or incidence distribution > 0.5), the Bias-corrected formula becomes inprecise. In these cases, EstimateS will post a message with Anne Chao's recommendation to chose the larger of Chao1 Classic and ACE, or Chao2 Classic and ICE.

Coverage-based estimators (ACE, ICE, Shared Species). The species richness estimators, ICE (Incidence Coverage-based Estimator) and ACE (Abundance Coverage-based Estimator) are modifications of the Chao & Lee (1992) estimators discussed by Colwell & Coddington (1994). Chazdon et al. (1998) introduced ICE and ACE to the ecological literature. See Appendix C of this User's Guide. The recommended (and default) upper limit for Rare or Infrequent species is 10 individuals or 10 samples, respectively.

For cases in which all Rare species are Singletons, ACE is undefined. Likewise, for cases in which all Infrequent species are Uniques, ICE is undefined. On the recommendation of Anne Chao, EstimateS uses the *bias-corrected form of the Chao1* and *Chao2*, respectively, for such cases.

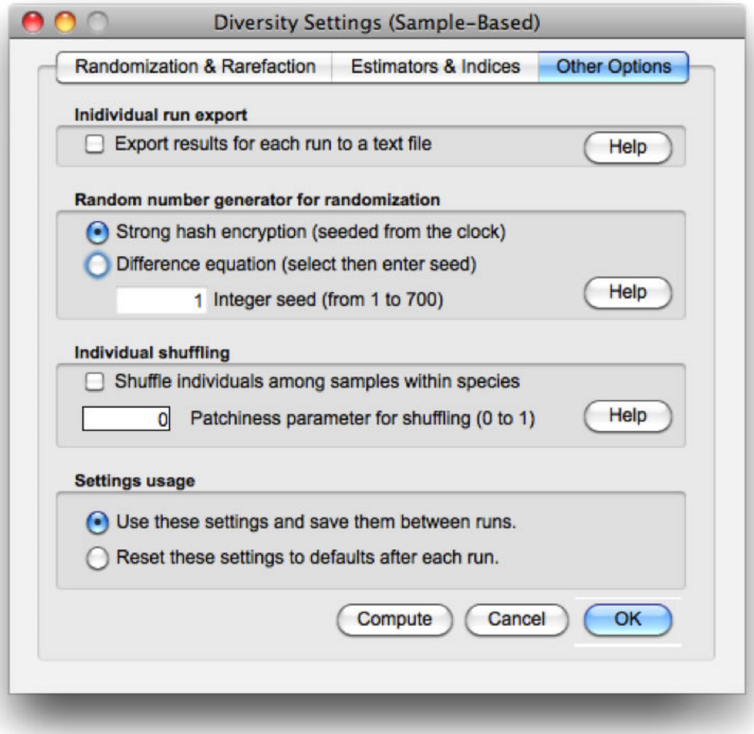
> ***Note:*** This setting also controls upper limit for Rare or Infrequent species for *Shared Species estimation.*

Randomization protocol for estimators and indices. If you specify *randomization of sample or individual order,* ***without replacement*** (the default, which is highly recommended), EstimateS selects a single sample (for sample-based filetypes) or a single individual (for individual-based filetypes) at random, computes the richness estimators (and diversity indices, if requested) based on that sample or individual, selects a second sample sample or individual, re-computes the estimators using the pooled data from both samples sample or individuals, selects a third, re-computes, and so on until all samples or individuals in the dataset are included. Samples or individuals are added to the analysis in random order, without replacement (each sample or individual is selected exactly once).

Each distinct randomization accumulates the samples or individuals in a different order, but all are included in each randomization. The final for species richness for the averaged, random-order species accumulation curve therefore matches, precisely, the total number of observed species. The drawback with this protocol is that the variance, among randomizations, of counts (individuals, singletons, etc.) and of estimators for which no analytical variance is provided, goes goes to zero at the right-hand end of the species accumulation curve. (Standard deviations based on variation among randomizations are identified as "runs" in EstimateS output. Standard deviations computed analytically include rarefied and extrapolated richness, for all filetypes, and standard deviations identified as "analytical" in EstimateS output.)

If you specify randomization of sample or individual order, ***with replacement**,* EstimateS follows the same procedure, but samples or individuals are added to the analysis in random order, with replacement (each sample or individual can appear in any pooled sample, some may appear in none). Each distinct randomization thus accumulates the samples or individuals in a different order, but in general, not all samples or individuals will be included, and some are likely to be chosen twice or more. Therefore, the final value of species richness for the averaged, random-order species accumulation curve generally is generally less the total number of observed species, since the missed samples or individuasl may represent species not found in the samples selected, for any given run. (In fact, the entire species accumulation/rarefaction curve generally lies below the corresponding curve produced by the without replacement option.) The advantage of randomizing samples with replacement is that the variance, among randomizations, of counts (individuals, singletons, etc.) and of estimators for which no analytical variance is provided, remains meaningful at the right- hand end of the species accumulation curve, and can thus be used to compare datasets.

4. Set the options on the *Other Options* tab.



Inidividual run export

☐ Export results for each run to a text file

Random number generator for randomization

● Strong hash encryption (seeded from the clock)

○ Difference equation (select then enter seed)

1 Integer seed (from 1 to 700)

Individual shuffling

☐ Shuffle individuals among samples within species

0 Patchiness parameter for shuffling (0 to 1)

Settings usage

● Use these settings and save them between runs.

○ Reset these settings to defaults after each run.

Individual run export. As an option, EstimateS records and exports results from *n* individual randomizations to a text file, allowing computation of precision, accuracy, and other analyses (Walther and Moore 2005), using Excel, R, or other applications. If you check the "Export results for each run to a text file" checkbox, when you click the *Compute* button (or choose *Compute Diversity* from the *Diversity* menu), EstimateS displays an expanatory message, and asks you to name and place the text file that will contain the exported results when the randomizations are complete. The data for each randomization appear in the same format as the summary Diversity results that EstimateS creates by default. (The summary results appear onscreen as usual, and may be exported as usual.) For large datasets, this option takes time, so be patient.

Random number generator for randomization. EstimateS offers two random number generator. The *Strong hash encryption* generator samples from a 160-bit strong hash (SHA) encryption function, seeded from the computer's clock. This procedure, developed by Jason Swain (personal communication), produces a non-repeating random number series that passes the most demanding tests.

The *Difference equation* alternative (Savitch (1992) is based on a seed number that you supply. Thus it permits EstimateS to generate precisely the same results on repeated sets of resampling runs with the same dataset. Unless you require precise repeatability, the strong hash encryption option is recommended.

If you would like to do a visual test of either random number generator, choose *Test Random Number Generator* from the *Special* menu.

Individual shuffling (*Sample-based filetypes only, with sample-based abundance data*). This tool allows you to explore the effects of spatial patchiness on species richness estimators, as discussed by Chazdon et al. (1998). If you check "Shuffle individuals among samples within species," EstimateS reassigns individuals at random to samples, within species, with a "tunable" degree of aggregation (patchiness).

> ***Note:*** Do not use this option without fully understanding it. It is a research and simulation tool, not an estimator.
>
> ***If the Patchiness parameter (A) is set to zero.*** Using the species abundance vector (marginal totals) for all samples pooled, each individual is re-assigned at random to a sample, within species. In other words, the distribution of individuals among species in the input matrix as a whole and the number of samples are maintained, but sample affiliations of individuals are randomized within species. Any patchiness of the original data is removed. (As expected, the mean of randomized sample accumulation curves is indistinguishable from the *Coleman curve*, which assumes spatial homogeneity, for this setting.)

If the Patchiness parameter (A) is set to a value greater than zero. In this case, the first individual of each species is assigned to a sample at random. The second (if there is one) is assigned to the same sample as the first with probability *A*, and to a randomly chosen sample with probability *(1-A)*. In other words, the larger you set *A*, the patchier the pseudo-distribution of individuals becomes. By "tuning" the patchiness of the distribution, you can investigate the effect on the performance of the richness estimators, using real relative abundance distributions. One could also enter made-up data sets that fit some particular relative abundance distribution(s).

Settings usage (saving settings). If you want to save your settings (the default) from one use of EstimateS to the next during a session, select "Use these settings and save them between runs." If you want to start with default settings the next time you open the Diversity or Shared Species settings screens, choose "Reset these settings to defaults after each run." Each time you launch EstimateS, all settings are returned to defaults.

5. Launch the Diversity computations.

To launch the Diversity computations directly, click the *Compute* button on the Diversity Settings screen, or click the *OK* button to save the settings, then choose *Compute Diversity Stats* from the Diversity menu. The results are displayed in the Diversity Statistics output screen.

6. Export the results of the Diversity computations.

To export the results of the Diversity computations to a tab-delimited text file, click the *Export* button at the bottom of the Diversity Statistics output screen or choose *Export Diversity Stats* from the Diversity menu. You can open the exported file in *Excel* or *R* or some other application to analyze and plot the data.

7. *(Optional)* Export the input data and all current parameter settings to a tab-delimited text file.

If you choose *Export Input File as Triplets* from the File menu. EstimateS creates a Format 3 input file, recording all parameter settings. You can reload this file at any time. The parameter settings are detailed in *Appendix A: Execution Control Parameters*.

Setting and Running the *Shared Species* Options (Sample-based filetypes only)

EstimateS computes a variety of statistics based on species shared between samples or between sets of replicated samples, including non-parametric estimators of the number of shared species (taking into account shared by unrecorded species), classic similarity indices, and non-parametric estimators of true similarity. All these meaasure require sample-based data. The Shared Species menu does not appear in the menu bar for individual-based data filetypes.

1. From the Shared Species menu, choose *Shared Species Settings*. The Shared Species Settings screen appears.

Shared Species Settings

Coverage-based estimators (ACE, ICE, Shared Species)

10 Upper abundance limit for rare or infrequent species (Recommended: 10)

Similarity indices and estimators

☑ Compute similarity indices: Jaccard (classic), Sorenson (classic), Chao-Jaccard Estimator, Chao-Sorensen Estimator, Morisita-Horn, and Bray-Curtis

☐ Input data are incidence frequencies, not abundances (Chao indices only)

☐ Compute bootstrap SEs for Chao indices only (takes time)

200 N for bootstraps (Chao indices only)

Settings usage

● Use these settings and save them between runs.

○ Reset these settings to defaults after each run.

2. **Set the options on the *Shared Species Settings screen*.** The image above shows the default settings.

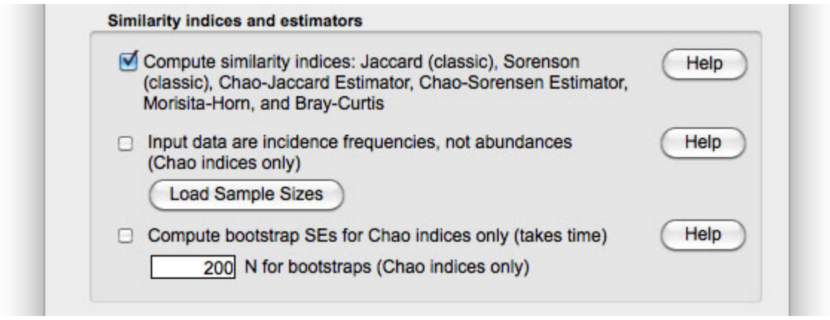
Coverage-based estimators (ACE, ICE, Shared Species). As discussed by Colwell & Coddington (1994), the problem of estimating the true number of species shared by two (or more) sites or biotas based on sample data presents a difficult but important challenge. The first statistical estimator of shared species was developed by Anne Chao and her colleagues (Chen et al. 1995 in Chinese; Chao et. al. 2000 in English), based on the same statistical strategy as ICE and ACE. Like ACE, the shared species estimator *V* requires abundance data. Just as ACE augments the observed number of species in a sample by a correction term dependent on the relative abundance of the rarest species (by default, those with fewer than 10 individuals) in the sample, *V* augments the observed number of shared species by a correction term based on the relative abundance of shared, rare species.

EstimateS computes Chao's shared species estimator for all pairs of samples in the input dataset (or datasets, for the multiple sample-based filetype). EstimateS also computes the ACE estimate of species richness for each sample. For cases in which all Rare species are Singletons, ACE is undefined.On the recommendation of Anne Chao, EstimateS uses the *bias-corrected form of the Chao1* and *Chao2* richness estimators, respectively, for such cases. A brief presentation of the mathematics behind the shared-species estimator appears in Appendix C of this Guide.

The recommended (and default) upper limit for Rare or Infrequent species is 10 individuals or 10 samples, respectively.

> ***Note:*** This setting also controls upper limit for Rare or Infrequent species for *ICE and ACE.*

Similarity indices and estimators. This panel has three checkboxes.



Similarity indices and estimators

☑ Compute similarity indices: Jaccard (classic), Sorenson (classic), Chao-Jaccard Estimator, Chao-Sorensen Estimator, Morisita-Horn, and Bray-Curtis

☐ Input data are incidence frequencies, not abundances (Chao indices only)

☐ Compute bootstrap SEs for Chao indices only (takes time)

200 N for bootstraps (Chao indices only)

Checkbox: Compute similarity indices. Checked by default, this box tells EstimateS to compute the similarity indices listed: Jaccard (classic), Sorenson (classic), Chao-Jaccard Estimator, Chao-Sorensen Estimator, Morisita-Horn, and Bray-Curtis.

EstimateS computes four classic indices of similarity, based on the raw data from the input file: the Classic Jaccard index, the Classic Sørensen incidence-based (qualitative, presence/absence) index, the Bray-Curtis index (= "Sørensen quantitative" index), and the Morisita-Horn index. Dozens of overlap indices exist in the literature; these were chosen based on the recommendations of Magurran (1998, 2004).

> ***Note:*** The Bray-Curtis (= "Sørensen quantitative") index and the Morisita-Horn index can be used with either integer or decimal (real number) input data. However, since EstimateS requires all data to be integer counts for estimator computation, all decimal data values are rounded to the nearest integer when imported into EstimateS. For this reason, values of the Sørensen Abundance-based index and the Morisita-Horn index computed by EstimateS will differ slightly from the corresponding indices computed for corresponding decimal data values, including Magurran's (1998) worked examples (Magurran 1988, pp. 165-166), which are based on decimal data.

Chao's Abundance-based Jaccard and Sørensen indices are based on the probability that two randomly chosen individuals, one from each of two samples (quadrats, sites, habitats, collections, etc.), both belong to species shared by both samples (but not necessarily to the same shared species). The estimators for these indices take into account the contribution to the true value of this probability made by species actually present at both sites, but not detected in one or both samples. This approach has been shown to reduce substantially the negative bias that undermines the usefulness of traditional similarity indices, especially with incomplete sampling of rich communities (Chao et al. 2005).

EstimateS computes the raw Chao Abundance-based Jaccard and Sørensen indices (not corrected for undersampling bias) as well as the estimators of their true values, so that you can assess the effect of the bias correction on the indices.

Checkbox: Input data are incidence frequencies. The default is to compute Chao-Jaccard & Chao-Sorensen Estimators using sample-based abundance data. Instead, it is possible to use replicated incidence data. In this case, the input data must be in terms of summed incidence frequencies, rather than abundances. Each column of the EstimateS Input File then represents the summed incidence frequencies from a different *Species X Samples* incidence matrix. All the original matrices must represent exactly the same global set of species, even if not all species are present in every matrix.

> ***Note:*** EstimateS does not compute the summed incidence frequencies. You must compute them in Excel, R, or another application from the original incicence data.

To compute replicated incidence indices, EstimateS needs to know the number of samples that you pooled to get the summed frequencies, for each incidence matrix. To input these sample sizes, use the "Load Sample Sizes" button in this panel. The required format is as follows:

> **Filetype:** One set of replicated sampling units (classic EstimateS input).
>
> LINE 1: Dataset title
> LINE 2: [Number of sample sizes, N]
> LINE 3: Sample size 1
> LINE 4: Sample size 2
> LINE N+2: Sample size N
>
> **Filetype:** Multiple sets of replicated sampling units (batch input for t datasets).
> LINE 1: Dataset title
>
> LINE 2: [Number of sample sizes, N1, for Dataset 1] <tab>[Number of sample sizes, N2, for Dataset 2] <tab>…<tab>[Number of sample sizes, Nt, for Dataset t]
>
> LINE 3: [Sample size 1, for Dataset 1] <tab>[Sample size 1, for Dataset 2] <tab>…<tab>[Sample size 1, for Dataset t]
>
> LINE 4: [Sample size 2, for Dataset 1] <tab>[Sample size 2, for Dataset 2] <tab>…<tab>[Sample size 2, for Dataset t]
>
> LINE Nmax+2: [Sample size Nmax, for Dataset 1] <tab>[Sample size Nmax, for Dataset 2] <tab>…<tab>[Sample size Nmax, for Dataset t]
>
> > ***Note:*** If not all datasets have the same number of sample sizes, you must fill in the empty cells of the input matrix with zeroes. An example input file is installed with Estimates, called *Multiple-Sample-Based-Example-Sample-Sizes.txt*, to be used with *Multiple-Sample-Based-Example.txt* as the Input Data File. (The sample sizes are hypothetical and do not reflect the original ant data.)

Checkbox: Compute bootstrap SEs for Chao indices only. If you check this box, EstimateS will estimate the standard errors for the ChaoJaccard and ChaoSørensen similarity estimators, allowing statistically rigorous comparison of two or more similarity index values. Standard errors for the Chao-Jaccard & Chao-Sorensen Estimators are computed by a bootstrap procedure, which requires resampling the observed data for pairs of samples and recomputing the estimators *N* times. You can specify *N* in the entry area labeled "*N* for bootstaps." See Chao et al. (2005) for details.

This procedure takes time. Anne Chao's suggested value for N for published results is 200 resamples, but you could use a smaller number for exploratory work.

To get the 95% Confidence Intervals, compute Chao-Jaccard-Est plus or minus 1.96*Chao-Jaccard-Est-SD, or Chao-Sorensen-Est plus or minus 1.96*Chao-Sorensen-Est-SD. (SE = SD because an infinite degrees of freedom is assumed.)

Settings usage. If you want to save your settings (the default) from one use of EstimateS to the next during a session, select "Use these settings and save them between runs." If you want to start with default settings the next time you open the Diversity or Shared Species settings screens, choose "Reset these settings to defaults after each run." Each time you launch EstimateS, all settings are returned to defaults.

3. Launch the Shared Species computations.

To launch the Shared Species computations directly, click the *Compute* button on the Shared Species Settings screen, or click the *OK* button to save the settings, then choose *Compute Shared Species Stats* from the Shared Species menu. The results are displayed in the Shared Species Statistics output screen.

4. Export the results of the Shared Species computations.

To export the results of Shared Species computations to a tab-delimited text file, click the *Export* button at the bottom of the Shared Species Statistics output screen or choose *Export Shared Species Stats* from the Shared Species menu. You can open the exported file in *Excel* or *R* or some other application to analyze and plot the data.

5. *(Optional)* Export the input data and all current parameter settings to a tab-delimited text file.

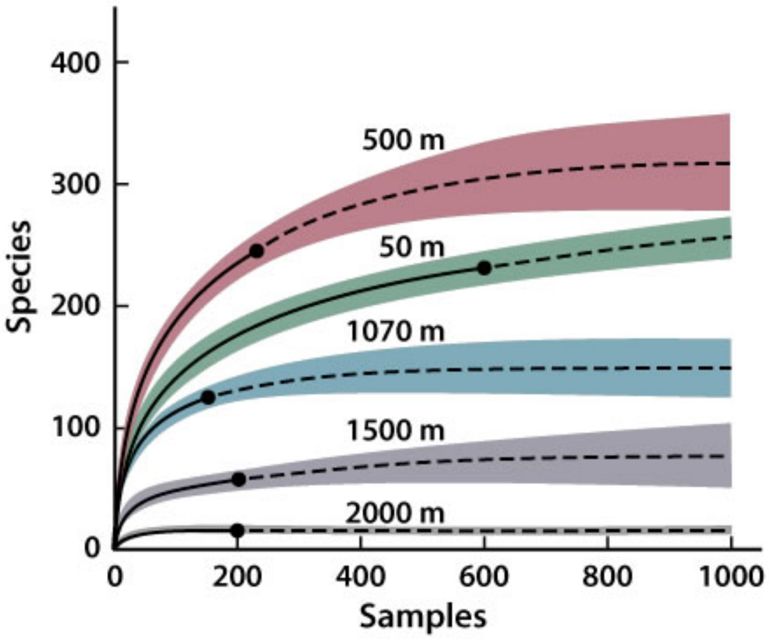
If you choose *Export Input File as Triplets* from the File menu. EstimateS creates a Format 3 input file, recording all parameter settings. You can reload this file at any time. The parameter settings are detailed in *Appendix A: Execution Control Parameters.*

Additional Notes

Comparing Species Accumulation Curves: Rarefaction and Extrapolation of Reference Samples

Rarefaction and extrapolation. EstimateS 9 introduces a new methodology for comparing the richness of *referenc* *samples* of biodiversity data. A reference sample is either a single, individual-based abundance sample of *n* individu or a set of *t* related sampling units for which incidence data have been recorded.

For four decades (Heck et al. 1983), biologists (and others) have used rarefaction to equalize the information conter individual-based abundance samples. Although sample-based rarefaction is at least as old (see Chiarucci et al. 200 it was not widely known or used until recently (Colwell et al. 2004, and in Estimates since 2004). Until the introductic of linked rarefaction and extrapolation curves (Colwell et al. 2012), based on a set of appropriate statistical samplinc models (rather than *functional curve-fitting*, like Michaelis-Menten or other functions), biologists were forced to comp richness of rarefied references samples at the sample size (in individuals or number of sampling units) of the *smalle* reference sample. The necessity of having to "throw away" data for the larger samples has long frustrated biologists but that frustration can now come an end, because there is a pot of gold at the end of the Rainbow (below).



With statistically sound extrapolation now possible (Colwell et al. 2012, nicknamed "The Rainbow" by its authors), thanks to the statistical genius of Anne Chao and her students, biologists and other users of rarefaction can now rigorously extrapolate the smaller samples, and compare them with the full reference sample for larger (and often th largest) samples in a dataset. A sample-based example for the species richness of ants at several elevations along transect in Costa Rica appears above (Longino and Colwell 2011). Reference samples are indicated by solid circles rarefaction by solid lines, extrapolation by dashed lines.

Confidence intervals for rarefaction and extrapolation. Of course, statistical comparison requires estimates not of richness itself, but of its variance, which we must know to estimate confidence limits. There are two ways to estim the variance of rarefied richness: *conditional* on the reference sample, or *unconditional*, treating the reference samp as a representative sample from a larger assemblage. Rarefaction curves with conditional confidence limits, which necessarily "converge to zero" variance at the reference sample, can answer only a very limited question: "Could *smaller* Reference Sample *A* have been drawn from the *larger* reference sample *B*?" With unconditional variance, reference samples can, in principle, be compared in the same way one would compare samples in a *t*-test or an ANOVA, asking whether or not two or more reference samples differ significantly at some specified P-value. Becaus richness is inherently sample-size dependent, however, any such comparison must be done at equivalent sample sizes, which is why we rarefy (and extrapolate).

An estimator of the unconditional variance for sample-based rarefaction was introduced by Colwell et al. (2004) and implemented in Estimates the same year. An estimator of the unconditional variance for individual-based rarefaction long missing from the biodiversity statistics tool-chest, was finally introduced by Colwell et al. 2012, and is implemer in EstimateS 9. For extrapolation, Shen et al. (2003) developed an unconditional variance estimator, also built into EstimateS 9, which Colwell et al. 2012 showed links smoothly with the unconditional variance estimators for rarefact despite being based on entirely different mathematics.

The computation of "open" unconditional confidence intervals for rarefaction and extrapolation assumes that some species in the assemblage sampled remain undetected, when all individuals or sampling units are pooled (the reference sample). An estimator of asymptotic richness is used to assess this assumption. For sample-based data, estimator is Chao2; for individual-based data, it is Chao1.

In the current release of EstimateS (Version 9.1.0), if Chao1 or Chao2 is equal to the observed number of species (S(obs)), the accumulation of species is assumed to have reached an asymptote, and the unconditional confidence interval closes to zero (around S(obs)). For the same reason, extrapolated richness is simply S(obs) for all sample sizes beyond the the reference sample. In terms of singletons and doubletons, for individual-based data (or uniques and duplicates, for sample-based data), the asymptote is reached when either there are no singletons (or uniques) in the pooled sample, or there is exactly one singleton (or one unique) and no doubletons (no duplicates). the formulas for Chao1 and Chao1 in Appendix B. In a future version, a new approach developed by Chao et al. (20 will be implemented, for these special cases, that estimates an "open" confidence interval even when neither singlet nor doubletons, for individual-based data (neither uniques nor duplicates, for sample-based data) are present in the reference sample.

Statistical inference. With regard to statistical inference, Colwell et al. 2012 write: "Even when based on unconditic variances, the use of confidence intervals to infer statistical significance (or lack of it) between samples is not straightforward. In general, lack of overlap between 95% confidence intervals (mean plus or minus 1.96 s.e.) does indeed guarantee significant difference in means at P< or = 0.05, but this condition is overly conservative: samples f normal distributions at the P = 0.05 threshold have substantially overlapping 95% confidence intervals. Payton et al. (2004) show that, for samples from two normal distributions with approximately equal variances, overlap or non-over of 84% confidence intervals (mean plus or minus 1.41 s.e.) provide a more appropriate rule of thumb for inferring a difference of mean at P = 0.05, and this approach has been suggested by two of us for comparing unconditional confidence intervals around rarefaction curves (Gotelli and Colwell 2011). Unfortunately, the statisticians among us (Anne Chao., C. X. Mao, and S.-Y. Li) doubt that this approach is likely to be accurate for the confidence intervals around rarefaction (or extrapolation) curves, so the matter of a simple method must be left for further study. Meanwh non-overlap of 95% confidence intervals constructed from our unconditional variance estimators can be used as a simple but conservative criterion of statistical difference."

Comparing sample-based abundance data. To compare sample-based abundance data, in terms of species richr instead of species density, Chazdon et al. (1998) and Gotelli & Colwell (2001) recommend rescaling the expected *sample-based* species accumulation curves (and their 95% confidence intervals) by *individuals*, instead of leaving th scaled by samples. To allow this rescaling to produce smooth curves, EstimateS computes the *expected* number of individuals for each sampling level, instead of taking the mean for number of individuals, among resampling runs. If there are *N* individuals, total, in *T* samples, total, the expected number of individuals in *t* samples is just $[t/ T)] * N$; th are the values tabled by EstimateS in the *Individuals* column of the output.

Coleman Rarefaction Curves

Like previous versions of EstimateS, Version 9 computes Coleman rarefaction curves (Coleman 1981, 1982) for *sample-based abundance data* (*Filetypes 1 and 2*). These curves estimate the number of species in 1, 2, ... T samp on the assumption that all individuals in all samples are randomly mixed (Chazdon et al. 1998). In other words, the Coleman curve in EstimateS for Filetypes 1 and 2 is a form of individual-based rarefaction, applied to sample-based data. In fact, for individual-based rarefaction (*Filetypes 3 and 4*), EstimateS 9 follows a Poisson model for rarefactio mathematically identical to Coleman's classic area-based sampling model (Colwell et al. 2012). If you summed abunances across samples in a sample-based abundance dataset (Filetype 1), and ran the totals as a single reference sample of Filetype 3 (indidividual-based rarefaction), the results would be indentical.

Rarefaction and Extrapolation vs. Asympotic Richness Estimation

Neither sample-based *rarefaction* curves nor individual-based *rarefaction* curves are estimators of the true species richness of the assemblage that a *reference sample* represents, in the same sense as the *asympotic richness estimators* that EstimateS computes. Whereas Chao1, Chao2, ACE, ICE or Jack1, for example, estimate *total speci richness*, including species not present in any sample, *rarefaction* curves estimate species richness for a *sub-sampl the pooled total* species richness, based on an empirical reference sample.

In contrast, the tools implemented in EstimateS 9 for *extrapolation* (Colwell et al. 2012) from a reference sample req a "target richness" that estimates the asymptotic number of species in the source assemblage, including species no documented by the reference sample. As explained in detail by Colwell et al. (2012), Chao1 was chosen to estimate the target richness for individual-based data and Chao 2 does so for sample-based data. For this reason, extrapolat may underestimate the expected richness of an augmented sample for hyperdiverse communities, for which Chao1 and Chao2 (and all other!) asymptotic richness estimators tend to increase with (reference) sample size.

Asymptotic Species Richness Estimators

The literature on species richness estimators continues to grow in several directions. Key reviews in the 1990s inclu Bunge & Fitzpatrick (1993) and Colwell & Coddington (1994). For a more recent review of the field, see Chao (2004 which, like most key papers cited in this User's Guide, can be downloaded as pdf file. Gotelli and Colwell (2011) also review the subject.

Chao1 and Chao2 Richness Estimators. In EstimateS, a comprehensive battery of both classic and bias-correcte forms of the richness estimators Chao1 and Chao 2 is computed along with log-linear 95% confidence intervals, as suggested by Chao (1987). These asymmetrical confidence intervals, which are based on the assumption that *log*(*S* - S_{obs}) is normally distributed, have the common-sense property that the lower confidence bound cannot be less tha the observed number of species, S_{obs}. See Appendix B for details. If you need a doubly-bounded richness estimator with a fixed upper bound, see Eren et al. (in press) (not implemented in EstimateS). Special forms of the Chao1 and Chao2 estimators (and their variances) are computed by EstimateS for cases involving sampling data with few singletons or doubletons (or uniques and duplicates). See Appendix A. Beginning with EstimateS Version 9.1.0, all versions of the Chao1 and Chao2 estimators include small-sample adjustment factors of the form (n-1)/n.

Anne Chao provides this advice on adequate sample size for Chao1 and Chao2: "The Chao1 and Chao2 estimators are universally valid *lower bounds* of species richness. They can be applied to any species abundance distribution and any sample size. In general, these two lower bounds are close to species[asymptotic richness if san size is sufficiently large, in which case the two estimators can be used as species richness estimators. A rough guideline for “sufficient” sample size: the estimated sample completeness should be at least 50%. For Chao 1, this means the proportion of singletons should be less than 50%, i.e., F1/n < 50%. For Chao 2, this means the proportion uniques should be less than 50%, i.e., Q1/M < 50%, where M is the total number of incidences.

Coverage-Based Richness Estimators ICE and ACE.The species richness estimators, ICE (Incidence-based Coverage Estimator) and ACE (Abundance-base Coverage Estimator) are modifications of the Chao & Lee (1992) estimators discussed by Colwell & Coddington (1994). Chazdon et al. (1998) introduced ICE and ACE to the ecolog literature. For that paper, they found it necessary and useful to change the notation for the variables involved in the other estimators, to allow a unified system of notation covering the new estimators. This new notation is referenced Table 1 and detailed in the Appendix C of this User's Guide, replacing the notation of Colwell & Coddington (1994). Chazdon et al. (1998), which can be downloaded as pdf file, for details and rationale.

Estimating Total Species Richness by Functional Extrapolation (Sample-based filetypes only)

> ***Note:*** With the development of extrapolation methods based on statistical sampling models, I would no longer recommend function-fitting extrapolation for most purposes. The data points used to fit them are non-independent and serially correlated, and do not permit the estimation of a rigorous confidence interval. This section of the User's Guide has been retained for those who may need it.

Many different curvilinear functions, asymptotic and non-asymptotic, might fit a species accumulation curve (Soberó Llorente 1993, Colwell & Coddington 1994, Colwell et al. 2004). As a richness estimation option, EstimateS compute (mostly as a legacy; see the Note, above) the asymptotic function most commonly used, the Michaelis-Menten funct (Colwell & Coddington 1994).

EstimateS computes two different Michaelis Menten (MM) richness estimators. In both, the data that EstimateS produces represent the estimated MM asymptote based on one, two, three...*T* samples (see Colwell & Coddington 1994, Fig. 1). The difference is that the first method (MMRuns) computes estimates for values for each pooling level each randomization run, then averages over randomization runs. If you have some samples that are much richer tha others, randomization runs that, by chance, add a rich sample early in the curve are likely to produce enormous estimates of richness, since the rich sample "shoots" the fitted MM curve suddenly skyward. Thus, MMRuns data ar often rather erratic for small numbers of samples, even when 100 runs are randomized.

The second method (MMMeans) computes the estimates for each sample pooling level just once, based on the analytical rarefaction curve for S(est). Since this curve is computed analytically, it is quite smooth, thus the MM Mea estimates are much less erratic than for the MMRuns method. This method is therefore generally recommended ove MMRuns.

> ***Note:*** Although *means* of S(est) among resampling runs are no longer used to compute MMMeans in Estimates 7 and later, the name *MMMeans* has been retained to make clear that it is the same as the estimator of that name in previous versions of EstimateS.

Indices of Species Diversity and Hill Numbers

In addition to rarefaction, extrapolation, and species richness estimators, all of which assess species richness as a measure of diversity, EstimateS computes the four most widely used indices of species diversity that combine information on richness and relative abundance in different ways (Magurran 2004; Jost 2006, 2007). They are *Fishe alpha* (the alpha parameter of a fitted logarithmic series distribution), *Shannon diversity* (using natural logarithms), *exponential Shannon diversity*, and *Simpson diversity* (the "inverse" form). The last two, like species richness itself, in units of equivalent, equally abundant species. For example, an exponential Shannon index or Simpson index of 4 based on a sample of 10 species of unequal abundance, means that the same value of the index would arise from a sample of 4 species of equal abundance. In terms of sensitivity to rare species, richness is the most sensitive, Simp diversity the least, and Shannon diversity intermediate. These three (when Shannon is its exponential form) represe particular points in a continuum of diversity indice, called Hill numbers, that share the same mathematical form (Jost 2006, 2007). Fisher's alpha is not part of this continuum.

EstimateS does not compute these indices unless you ask it to. Check the Diversity Indices checkbox on the Other Options tab of the Diversity Settings screen to enable this option.

As with species richness estimators, EstimateS computes these four indices for each level of sample pooling, from c sample up to the total number in your dataset, allowing you to see whether and when each index stabilizes with increasing numbers of samples. Because of the balance each strikes between richness and evenness, Fisher's alph and Simpson will almost inevitably stabilize faster (for smaller sample sizes) then Shannon, and Shanon will stabiliz faster than richness. This pattern does not mean that one is "better" than another; they measure different things (Jos 2006).

Samples or individuals are added to the pool at random. The Runs parameter (on the Randomizations tab of the Diversity Settings screen) specifies how many randomizations EstimateS carries out to compute the mean and bootstrap (conditional on the reference sample) standard deviation (for all but Fisher's alpha, for which an unconditic SD is computed) for the indices at each level of pooling. You can also specify whether you want the samples to be added to the pool *with or without replacement*.

Non-integer Sampling Data (Percent Cover, Basal Area, Biomass, etc) and EstimateS

To understand the issues with non-integer data, we need to distinguish between data that are intrinsically non-intege numbers (e.g. percent cover, basal area, biomass, etc.), integer *abundance data* (counts of discrete individuals), an *replicated incidence data* (presence/absence in replicated sampling units, such as quadrats, transects, traps, nets, plankton hauls, etc.). Like abundance data, replicated incidence data are integer "counts" (number of samples in wh a species occurs) and represent a powerful approach to estimating richness and a assessing biotic similarity. If there any way to convert your non-integer data to replicated incidence data, you can use nearly all of EstimateS's tools ar statistics.

EstimateS expects integer data (no decimal markers in the input data), because most of the biodiversity statistics it computes are based on sampling models for counts (either individuals or incidences), and make no sense for non-integer data. There are a few exceptions: Shannon and Simpson diversity indexes are based on proportions, so no integer data make sense for these indices. If that is all you want, you can multiply all your input data by some const to get "integer" data, and run EstimateS on these values. But be aware that only Simpson and Shannon diversities make any sense, and you must ignore everything else!

What EstimateS 9 Computes

Table 1, below, lists the variables and statistics that EstimateS 9 computes from the *Diversity* menu. Table 2 lists the variable and statistics computed from the *Shared Species* menu.

Table 1: Diversity Statistics. Accumulated species and individuals, richness estimators, species diversity indices a related variables computed by EstimateS 9. In the output screen (and exported text files), values for accumulated species, richness estimators, and diversity indices appear *for each level of accumulatio*n, from a single sampling uni a single individual up to the full *reference sample*. The statistics listed are reported as analytically computed expecte values, or as mean values averaved over the number of randomizations you specify, for statistics that have no analytical rarefaction known. Formulas for the estimators appear in Appendix B .

Filetype	**Variable**	**Estimator**	**Reference**
Sample-based	Samples (*t*)	Number of sampling units accumulated	*m* in Chazdon et al. (1998) *h* in Colwell et al. (2004) *t* in Colwell et al. (2012)
Sample-based	Individuals (computed)	*[t/T]*N*, where T is the number of sampling units in the reference sample and *N* is the total number of individuals in all *T* samples (makes sense for sample-based *abundance* date only)	Gotelli and Colwell (2001) Gotelli and Colwell (2011)
Sample-based	S(est) (analytical)	Expected number of species in *t* pooled samples, given the reference sample (analytical).	**Rarefaction:** *MaoTau* in earlier versions of EstimateS (< v. 9), Eq. 5 in Colwell et al. (2004), Eq. 17 in Colwell et al. (2012) **Extrapolation:** Eq. 18 in Colwell et al. (2012)
Sample-based	S(est) 95% CI Lower Bound	Lower bound of 95% Confidence Interval for S(est)	**Rarefaction:** Eq. 6 in Colwell et al. (2004) **Extrapolation:** Eq. 19 in Colwell et al. (2012)
Sample-based	S(est) 95% CI Upper Bound	Upper bound of 95% Confidence Interval for S(est)	**Rarefaction:** Eq. 6 in Colwell et al. (2004) **Extrapolation:** Eq. 19 in Colwell et al. (2012)
Sample-based	S(est) SD (analytical)	Standard deviation of S(est) (analytical) (SD = SE)	**Rarefaction:** Eq. 6 in Colwell et al. (2004) **Extrapolation:** Eq. 19 in Colwell et al. (2012)
Sample-	S Mean	Number of species in *t* pooled	*Sobs Mean* in earlier versions of

based	(runs)	samples, given the reference sample (mean among runs)	EstimateS (< v. 9)
Individual-based	Individuals (*m*)	Number of individuals	*m* in Colwell et al. (2012)
Individual-based	S(est) (analytical)	Expected number of species represented among *m* individuals, given the reference sample (analytical).	**Rarefaction:** Eq. 4 in Colwell et al. (2012) **Extrapolation:** Eq. 9 in Colwell et al. (2012), slightly modified (to match Eq. 18, on Anne Chaos' advice)
Individual-based	S(est) 95% CI Lower Bound	Lower bound of 95% Confidence Interval for S(est)	**Rarefaction:** Eq. 7 in Colwell et al. (2012) **Extrapolation:** Eq. 10 in Colwell et al. (2012)
Individual-based	S(est) 95% CI Upper Bound	Upper bound of 95% Confidence Interval for S(est)	**Rarefaction:** Eq. 7 in Colwell et al. (2012) **Extrapolation:** Eq. 10 in Colwell et al. (2012)
Individual-based	S(est) SD (analytical)	Standard deviation of S(est) (analytical) (SD = SE)	**Rarefaction:** Eq. 7 in Colwell et al. (2012) **Extrapolation:** Eq. 10 in Colwell et al. (2012)
Individual-based	S Mean (runs)	Number of species represented among *m* individuals, given the reference sample (mean among runs)	
All filetypes	Singletons Mean	Number of singletons (species with only one individual) in *t* pooled samples or among *m* individuals (mean among runs)	*a* in Colwell & Coddington (1994) *F1* in Chazdon et al. (1998) *f*1 in Colwell et al. (2012)
All filetypes	Singletons SD (runs)	Standard deviation of Singletons, among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.
All filetypes	Doubletons Mean	Number of doubletons (species with only two individuals) in *t* pooled samples or among *m* individuals (mean among runs)	*b* in Colwell & Coddington (1994) *F2* in Chazdon et al. (1998) *f*2 in Colwell et al. (2012)
All filetypes	Doubletons SD (runs)	Standard deviation of doubletons, among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.
Sample-based	Uniques Mean	Number of uniques (species that occur in a only one sample) in *t* pooled samples (mean among runs)	*L* in Colwell & Coddington (1994) *Q1* in Chazdon et al. (1998) *Q1* in Colwell et al. (2012)
Sample-based	Uniques SD (runs)	Standard deviation of Uniques, among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.

Sample-based	Duplicates Mean	Number of duplicates (species that occur in a only two samples) in *t* pooled samples (mean among runs)	*M* in Colwell & Coddington (1994) *Q2* in Chazdon et al. (1998) *Q2* in Colwell et al. (2012)
Sample-based	Duplicates SD (runs)	Standard deviation of duplicates, among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.
Sample-based & Individual-based	ACE Mean	Abundance Coverage-based Estimator of species richness (mean among runs)	Chao et al. (2000), Chazdon et al. (1998)
Sample-based & Individual-based	ACE SD (runs)	Standard deviation of ACE, among randomizations of sample order or individual order	This is a bootstrap SD, based on variation in sample order among randomizations.
Sample-based	ICE Mean	Incidence Coverage-based Estimator of species richness (mean among runs)	Chao et al. (2000), Chazdon et al. (1998)
Sample-based	ICE SD (runs)	Standard deviation of ICE, among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.
All filetypes	Chao1 Mean	Chao 1 richness estimator (mean among runs)	Chao (1984), with special cases as detailed in Appendix B.
All filetypes	Chao1 95% CI Lower Bound	Chao 1 log-linear confidence interval lower bound (mean among runs)	Chao (1987), see Appendix B.
All filetypes	Chao1 95% CI Upper Bound	Chao 1 log-linear confidence interval upper bound (mean among runs)	Chao (1987), see Appendix B.
All filetypes	Chao1 SD (analytical)	Chao 1 standard deviation (by Chao's formulas)	Chao (1987) (not Chao 1984). Note: The formula in Colwell & Coddington (1994) is incorrect. See Appendix B for the correct formula and for special cases.
Sample-based	Chao2 Mean	Chao 2 richness estimator (mean among runs)	Chao (1984, 1987), with special cases as detailed in Appendix B.
Sample-based	Chao2 95% CI Lower Bound	Chao 2 log-linear confidence interval lower bound (mean among runs)	Chao (1987), see Appendix B.
Sample-based	Chao2 95% CI Upper Bound	Chao 2 log-linear confidence interval upper bound (mean among runs)	Chao (1987), see Appendix B.
Sample-	Chao2 SD	Chao 2 standard deviation (by	Chao (1987) Note: The formula in Colwell

based	(analytical)	Chao's formula)	& Coddington is incorrect. See Appendix B for the correct formula and for special cases.
Sample-based	Jack1 Mean	First-order Jackknife richness estimator (mean among runs)	Burnham & Overton(1978, 1979), Smith & van Belle (1984), Heltshe & Forrester (1983)
Sample-based	Jack1 SD (runs)	First-order Jackknife standard deviation	This is a bootstrap SD, based on variation in sample order among randomizations.
Sample-based	Jack2 Mean	Second-order Jackknife richness estimator (mean among runs)	Burnham & Overton(1978, 1979), Smith & van Belle (1984), Palmer (1991)
Sample-based	Jack2 SD (runs)	Standard deviation of Jack2, among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.
Sample-based	Bootstrap Mean	Bootstrap richness estimator (mean among runs)	Smith & van Belle (1984)
Sample-based	Bootstrap SD (runs)	Standard deviation of Bootstrap, among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.
Sample-based	MMRuns Mean	Michaelis-Menten richness estimator: estimators averaged over randomizations (mean among runs)	Raaijmakers (1987)
Sample-based	MMMeans (1 run)	Michaelis-Menten richness estimator: estimators computed once for analytica rarefaction curve, computed by Eq. 5 in Colwell et al. (2004)	Raaijmakers (1987), Colwell et al. (2004)
Sample-based	Cole Rarefaction	Coleman rarefaction (number of species expected in *t* pooled samples, assuming individuals distributed at random among samples)	Coleman (1981), Coleman et al. (1982)
Sample-based	Cole SD	Coleman standard deviation (analytical)	Coleman (1981), Coleman et al. (1982)
All filetypes	Alpha Mean	Fisher's alpha diversity index	Magurran (2004), Hayek & Buzas (1996)
All filetypes	Alpha SD (analytical)	Fisher's alpha standard deviation	Magurran (1988), Hayek & Buzas (1996)
All filetypes	Shannon Mean	Shannon diversity index (mean among runs), natural logarithms	Magurran (2004, page 238)

All filetypes	Shannon SD (runs)	Standard deviation of Shannon index among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.
All filetypes	Shannon Exp Mean	Exponential Shannon diversity index (mean among runs)	Magurran (2004, page 149); Jost (2006)
All filetypes	Shannon Exp SD (runs)	Standard deviation of Exponential Shannon index among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.
All filetypes	Simpson (Inverse) Mean	Simpson (inverse) diversity index (mean among runs)	Magurran (1988, eq. 2.27), Magurran (2004, p. 115), Hayek & Buzas (1996); Jost (2006)
All filetypes	Simpson (Inverse) SD (runs)	Standard deviation of Simpson (inverse) index among randomizations of sample order	This is a bootstrap SD, based on variation in sample order among randomizations.

Table 2: Shared Species Statistics. Shared Species estimators, classic similarity indices, Chao's abundance-base Jaccard and Sorensen similarity indices and their estimators, and related variables computed by EstimateS 9. In the output screen (and exported text files), values for these statistics and variables appear for each possible pair of samples. The formula for the shared species estimator appears in Appendix C , and the formulas for Chao's abundance-based Jaccard and Sorensen similarity indices, and their estimators and variances appears in Appendix

Note: The statisics in Table 2 are computed only for *sample-based abundance data.*

Variable	**Estimator**	**Reference**
First Sample		*j* in Appendix C
Second Sample		*k* in Appendix C
Sobs First Sample	Observed number of species in the First Sample	
Sobs Second Sample	Observed number of species idiv id="masthead"div id="masthead"n the Second Sample	
Shared Spp Observed	Observed number of species shared by First and Second samples	
ACE First	Estimated number of species in the First Sample: ACE	Chao, Ma, and Yang (1993), Chazdon et al. (1998)
ACE Second	Estimated number of species in the Second Sample: ACE	Chao, Ma, and Yang (1993), Chazdon et al. (1998)

Chao Shared Estimated	Estimated number of species shared by First and Second samples: *V(est)*	Chen et al. 1995
Jaccard Classic	Classic Jaccard sample similarity index	Chao et al. (2005, eq. 1)
Sørensen Classic	Classic Sørensen incidence-based (qualitative) sample similarity index	Chao et al. (2005, eq. 2)
Chao-Jacc-Raw Abundance-based	Chao's Jaccard Raw (uncorrected for unseen species) Abundance-based similarity index	Chao et al. (2005, eq. 5)
Chao-Jacc-Est Abundance-based	Chao's estimator (corrected for unseen species) for Chao's Jaccard Abundance-based similarity index	Chao et al. (2005, eq. 9)
Chao-Jacc-EstSD Abundance-based	Standard Deviation of Chao's estimator (corrected for unseen species) for Chao's Jaccard Abundance-based similarity index	Chao et al. (In press)
Chao-Jacc-Est Incidence-based	Chao's estimator (corrected for unseen species) for Chao's Jaccard similarity index for replicated Incidence-based data	Chao et al. (2005, eq. 13)
Chao-Sor-EstSD Indidence-based	Standard Deviation of Chao's estimator (corrected for unseen species) for Chao's Jaccard similarity index for replicated Incidence-based data	Chao et al. (In press)
Chao-Sor-Raw Abundance-based	Chao's Sørensen Raw (uncorrected for unseen species) Abundance-based similarity index	Chao et al. (2005, eq. 6)
Chao-Sor-Est Abundance-based	Chao's estimator (corrected for unseen species) for Chao's Sørensen Abundance-based similarity index	Chao et al. (2005, eq. 10)
Chao-Sor-EstSD Abundance-based	Standard Deviation of Chao's estimator (corrected for unseen species) for Chao's Sørensen Abundance-based similarity index	Chao et al. (In press)

Chao-Sor-Est Incidence-based	Chao's estimator (corrected for unseen species) for Chao's Sørensen similarity index for replicated Incidence-based data	Chao et al. (2005, eq. 14)
Chao-Sor-EstSD Indidence-based	Standard Deviation of Chao's estimator (corrected for unseen species) for Chao's Sørensen similarity index for replicated Incidence-based data	Chao et al. (In press)
Morisita-Horn	Morisita-Horn sample similarity index	Magurran (1988, eq. 5.10), Magurran (2004, page)
Bray-Curtis	Bray-Curtis (=Sørensen quantitative) sample similarity index	Magurran (1988, eq. 5.9), Magurran (2004, page)

Things You Should Know Before You Begin

Caveat Receptor

I have done my best to check all features of EstimateS 9 for usability and all computations and algorithms for accura but the final responsibility for ensuring that your results are correct must rest with you.

In general, you should have little trouble understanding the output, by referring to Colwell et al. (2012), Gotelli & Col (2001), Gotelli & Colwell (2011), Chao et. al. (2005), Colwell & Coddington (1994), Chazdon et al. (1998), Colwell et (2004), , or if necessary the references in Tables 1 and 2.

Citing EstimateS

If you appreciate the effort that has gone into EstimateS, please credit the application and its author in any publishe work that makes use of results from EstimateS, citing EstimateS as an electronic publication and giving the Estimate persistent URL (PURL) website address (http://purl.oclc.org/estimates) if the journal permits it. (This "permanent" address automatically transfers the visitor to http://viceroy.eeb.uconn.edu/EstimateS or any subsequent host. Here i one possible form for a References Cited entry:

> Colwell, R. K. 2013. EstimateS: Statistical estimation of species richness and shared species from samples. Version 9. User's Guide and application published at: http://purl.oclc.org/estimates.

If the journal or book editor will not permit an entry in the References Cited section, you might try this text citation: "...computed using EstimateS (Version 9, R. K. Colwell, http://purl.oclc.org/estimates)...."

Failing that, you may be reduced to: "...computed using EstimateS (Version 9, R. K. Colwell, unpublished)...," perha slipping in the EstimateS website address (http://purl.oclc.org/estimates) in the Acknowledgment section.

I would be most grateful if you would kindly send a reprint of any paper based on your use of the program. Send a p to colwell@uconn.edu

What You Must Agree To: Copyright and Fair Use

EstimateS is a freeware application. By downloading and using EstimateS, you must agree not to distribute Estimate in any commercial form.

You are most welcome to use EstimateS in any way you like for your own research, as long as such use is acknowledged as outlined above.

Sharing EstimateS With Others

To keep track of EstimateS users and to make sure that the latest version is in use, it is preferable that each new us downloads and registers his or her own copy of EstimateS from http://viceroy.eeb.uconn.edu/estimates or http://purl.oclc.org/estimates, rather than sharing someone else's (e.g. your) copy.

If you do share the program with a colleague, please be sure to make clear that the User's Guide is available online http://viceroy.eeb.uconn.edu/estimates or http://purl.oclc.org/estimates, to save needless email support questions.

References Cited

References marked "Download pdf" are available here for downloading.

Brewer, A., & M. Williamson. 1994. A new relationship for rarefaction. *Biodiversity and Conservation* **3**:373-379.

Bunge, J., & M. Fitzpatrick. 1993. Estimating the number of species: A review. *Journal of the American Statistical Association* **88**, 364-373.

Burnham, K.P. & W.S. Overton. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 623-633.

Burnham, K.P. & W.S. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60**, 927-936.

Butler, B. J., & R. L. Chazdon. 1998. Species richness, spatial variation, and abundance of the soil seed bank of a secondary tropical rain forest. *Biotropica* **30**:214-222. *Download pdf.*

Chao, A. 1984. Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statist* **11**, 265-270. *Download pdf.*

Chao, A. 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 7 791. *Download pdf.*

Chao, A. 2005. Species richness estimation, Pages 7909-7916 in N. Balakrishnan, C. B. Read, and B. Vidakovic, e *Encyclopedia of Statistical Sciences*. New York, Wiley. *Download pdf.*

Chao, A., R. L. Chazdon, R. K. Colwell, and T.-J. Shen. 2005. A new statistical approach for assessing compositiona similarity based on incidence and abundance data. *Ecology Letters* **8**:148-159. *Download pdf.* Spanish Version: Download pdf.

Chao, A., R. L. Chazdon, R. K. Colwell, and T.-J. Shen. 2006. Abundance-based similarity indices and their estimati when there are unseen species in samples. *Biometrics* **62,** 361-371. *Download pdf.*

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2013, *online early.* Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studie *Ecological Monographs.*

Chao, A., W.-H. Hwang, Y.-C. Chen, and C.-Y. Kuo. 2000. Estimating the number of shared species in two communities. *Statistica Sinica* **10**:227-246. *Download pdf.*

Chao, A. & S.-M Lee. 1992 Estimating the number of classes via sample coverage. *Journal of the American Statistic Association* **87**, 210-217. *Download pdf.*

Chao, A., M.-C. Ma, & M. C. K. Yang. 1993. Stopping rules and estimation for recapture debugging with unequal fail rates. *Biometrika* **80**, 193-201. *Download pdf.*

Chazdon, R. L., R. K. Colwell, J. S. Denslow, & M. R. Guariguata. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. Pp. 285-309 in F. Dallmeier and J. A. Comiskey, eds. *Forest biodiversity research, monitoring and modeling: Conceptual background and Old W case studies.* Parthenon Publishing, Paris. *Download pdf.*

Chen, Y.-C., W.-H. Hwang, A. Chao, & C.-Y. Kuo. 1995. Estimating the number of common species. Analysis of the number of common bird species in Ke-Yar Stream and Chung-Kang Stream. (In Chinese with English abstract.) *Jou of the Chinese Statistical Association* **33,** 373-393.

Chiarucci, A., G. Bacaro, D. Rocchini, and L. Fattorini. 2008. Discovering and rediscovering the sample-based rarefaction formula in the ecological literature. *Community Ecology* **9**:121-123.

Coleman, B.D. 1981. On random placement and species-area relations. *Mathematical Biosciences* **54**, 191-215.

Coleman, B.D., Mares, M.A., Willig, M.R. & Hsieh, Y.-H. 1982. Randomness, area, and species richness. *Ecology* **6** 1121-1133.

Colwell, R. K. 2006. *Biota: The biodiversity database manager, Version 3.*

Colwell, R. K., A. Chao, N. J. Gotelli, S.-Y. Lin, C. X. Mao, R. L. Chazdon, and J. T. Longino. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation, and comparison of assemblages. *Journal of Plant Ecology* **5**:3-21. Download pdf. Read online.

Colwell, R. K., & J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society* (Series B) **345**, 101-118. *Download low resolution pdf.* or *download high resolutio pdf.*

Colwell, R. K., C. X. Mao, & J. Chang. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* **85**, 2717-2727. *Download pdf.* Spanish Version: *Download pdf.*

Gotelli, N., & R. K. Colwell. 2001. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparis of species richness. *Ecology Letters* **4** , 379-391. *Download pdf.*

Gotelli, N. J. and R. K. Colwell. 2011. Estimating species richness. Pages 39-54 in A. E. Magurran and B. J. McGill, editors. *Frontiers in measuring biodiversity.* Oxford University Press, New York

Hayek, L. C., & M. A. Buzas. 1996. *Surveying natural populations.* Columbia University Press, NY.

Heck, K.L., Jr., van Belle, G. & Simberloff, D. 1975. Explicit calculation of the rarefaction diversity measurement and determination of sufficient sample size. *Ecology* **56**, 1459-1461.

Heltshe, J. & Forrester, N.E. 1983 . Estimating species richness using the jackknife procedure. *Biometrics* **39**, 1-11.

Jost, L. 2006. Entropy and diversity. *Oikos* 113:363.

Jost, L. 2007. Partitioning diversity into independent alpha and beta components. *Ecology* 88:2427-2439.

Lee, S.-M., and A. Chao. 1994. Estimating population size via sample coverage for closed capture-recapture model *Biometrics* **50**, 88-97. *Download pdf.*

Longino, J. T. and R. K. Colwell. 2011. Density compensation, species composition, and richness of ants on a neotropical elevational gradient. *Ecosphere* 2(3):art29, doi:10.1890/ES10-00200.1. *Download pdf. Online here.*

Mao, C. X., R. K. Colwell, and J. Chang. 2005. Estimating species accumulation curves using mixtures. Biometrics 61:433–441. *Download pdf.*

Magurran, A. E. 1988. *Ecological diversity and its measurement.* Princeton University Press, Princeton, N. J.

Magurran, A. E. 2004. *Measuring biological diversity*. Blackwell.

Palmer, M.W. 1991. Estimating species richness: The second-order jackknife reconsidered. *Ecology* **72**, 1512-1513.

Payton, M. E., M. H. Greenstone, and N. Schenker. 2004. Overlapping confidence intervals or standard error interva What do they mean in terms of statistical significance? 6 pp. *Journal of Insect Science*, 3:34. Available online: insectscience.org/3.34.

Raaijmakers, J. G. W. 1987. Statistical analysis of the Michaelis-Menten equation. *Biometrics* **43**, 793-803.

Savitch, Walter J. 1992. *Turbo Pascal : an introduction to the art and science of programming.* 3rd ed. Benjamin/Cummings, Redwood City, Calif.

Shen, T.-J., A. Chao, and C.-F. Lin. 2003. Predicting the number of new species in further taxonomic sampling. *Ecol* 84:798-804.

Smith, E.P. & van Belle, G. 1984. Nonparametric estimation of species richness. *Biometrics* **40**, 119-129.

Soberón, J., & J. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. *Conservation Biology* **7** , 480-488.

Ugland, K. I., J. S. Gray, & K. E. Ellingsen. 2003. The species-accumulation curve and estimation of species richnes *Journal of Animal Ecology* **72** , 888-897.

Walther, B. A., and J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* **28**, 815-8

Appendices

Appendix A: Contol Parameters for Automated Input

This Appendix applies only to *sample-based incidence or abundance filetypes* (the *classic EstimateS input filetype* a its *batch version*). *Most users can simply skip this section* and use the graphical query screens, during input, or the graphical Settings screens to set options, once your data have been input to EstimateS. All options described in this Appendix may be set, instead, from the onscreen graphical user interface. These Execution Control Parameters are intended primarily for repeated or automated data entry and execution.

For information on **Record 1** (and the **Batch Record**) for sample-based filetypes, *click here.*

Record 2: Parameter Record (all this on one line, each element separated by a <tab> character from the nex

Required: Number of species

Required: Number of samples (sampling units)

> ***Note:*** The remaining, optional parameters are intended to be used for repeated analyses. It is much easier to set these options from graphical query and settings screens during input, or in the graphical Settings screens, once your data have been input to EstimateS.

Optional: [AbMax]: This parameter is ignored in EstimateS 7+; it is retained only for backwards compatibility.

Optional: [Runs]: Number of randomizations to perform.

Optional: [Memory]: If this parameter is blank or zero, the SHA random number generator is used (seeded from the clock). An integer value > 0 in this field is interpreted as the "seed" for the difference equation random number generator. It must an integer, any value between 1 and 700.

Optional: [RareInfreqCut]: The number of abundance classes (singletons, doubletons, tripletons, etc.) or the number of incidence classes (uniques, duplicates, triplicates, etc.) to be included in the calculation of the CV estimates used in ICE, ACE, and shared species estimator V. Anne Chao (pers. comm.) recommends using 10 for this parameter. If this parameter is blank or zero, EstimateS set it to 10.

Optional: [DivIndexFlag]: If this flag is 1, EstimateS computes Fisher's alpha and the Shannon and Simpson indices. If this flag is blank or zero, these indices are not computed.

Optional: [RandFlag]: If this flag is set to 1, EstimateS does *not* randomize sample order and the Runs parameter is set automatically to 1. If this flag is blank or zero, Runs randomizations are carried out.

Optional: [Shuffle]: If this flag is set to 1, EstimateS randomizes the placement of individuals among samples, within species (Chazdon et al. 1998), using the Patchiness parameter to set aggregation. If this flag is blank or zero, no shuffling is done.

Optional: [Patchiness]: This variable must be between 0 and 1, inclusive. See details on the Patchiness parameter earlier in this Guide. The recommended default is zero.

Optional: [SimIndexFlag]: If this flag is set to 1, EstimateS computes the Jaccard, Sørensen, and Morisita-Horn indices. If this flag is blank or zero, the indices are not computed.

Optional: [FormatKey]: This variable specifies the input file format, and must be an integer between 0 and 5. EstimateS always allows you to specify the file format during data input, so you need not include this parameter. (It is set automatically to 3 in Format 3 files exported from EstimateS, and is set to 5 when reading Biota to EstimateS input files.)

Optional: [ChaoClassic]: If this flag is blank or zero, EstimateS uses the bias-corrected form of the Chao1 and Chao2 richness estimators in all cases (the recommended default). If this flag is set to 1, EstimateS uses the the bias-corrected form only when doubletons (Chao1) or duplicates (Chao2) are zero, and uses the approximate ("classic") formulas otherwise.

Optional: [Replace]: If this flag is blank or zero, EstimateS randomizes sample order **without** replacement. If this flag is 1, samples are selected for accumulation **with** replacement.

Optional: [SkipRows]: If this parameter is blank or zero, EstimateS assumes the input file contains no label rows. If set to N, EstimateS will skip N rows after reading the Title Record and the Parameter Record, then begin reading the incidence or abundance rows. (SkipRows may also be indicated in the Title Record, in EstimateS 9 and later)

Optional: [SkipColumns]: If this parameter is blank or zero, EstimateS assumes the input file contains no label columns. If set to N, EstimateS will skip the first N columns when reading each incidence or abundance row. (SkipColumns may also be indicated in the Title Record, in EstimateS 9 and later)

Optional: [ExportRuns]: If this parameter is blank or zero, EstimateS does not export the Diversity results for individual randomizations (runs). If set to 1, Diversity results for each randomization are exported. See Option to Export Results from Individual Randomizations.

Appendix B: Nonparametric Estimators of Species Richness

Please note that nonparametic estimators of species richness are *minimum estimators*: their computed values shoul be viewed as lower bounds of total species numbers, given the information in a sample or sample set.

Definition of variables

S_{est}	Estimated species richness, where *est* is replaced in the formula by the name of the estimator
S_{obs}	Total number of species observed in all samples pooled
S_{rare}	Number of rare species (each with 10 or fewer individuals) when all samples are pooled
S_{abund}	Number of abundant species (each with more than 10 individuals) when all samples are pooled
S_{infr}	Number of infrequent species (each found in 10 or fewer samples)
S_{freq}	Number of frequent species (each found in more than 10 samples)
m	Total number of samples
m_{infr}	Number of samples that have at least one infrequent species
F_i	Number of species that have exactly *i* individuals when all samples are pooled (F_1 is the frequency of singletons, *F₂* the frequency of doubletons)
Q_j	Number of species that occur in exactly *j* samples (Q_1 is the frequency of uniques, Q_2 the frequency of duplicates)
p_k	Proportion of samples that contain species *k*
N_{rare}	Total number of individuals in rare species
N_{infr}	Total number of incidences (occurrences) of infrequent species
C_{ace}	Sample abundance coverage estimator
C_{ice}	Sample incidence coverage estimator
γ^2_{ace}	Estimated coefficient of variation of the F_i for rare species
	Estimated coefficient of variation of the Q_i for infrequent species

γ^2_{ice}	

The estimators

Chao 1 and Chao2: Different equations are used to compute the Chao1 and Chao2 richness estimators, their estimated variance, and the corresponding log-linear 95% confidence intervals, depending on (1) the number of singletons and doubletons (in abundance-based data) or uniques and duplicates (for incidence-based data), and (2) settings you select "Chao 1 and Chao 2 bias correction" panel in the Estimators tab of the Diversity Settings screen (Diversity menu). The table below specifies the equations used in each case. The equations referred to appear belo the table. This section was developed in personal communication with Anne Chao, Institute of Statistics, National Ts Hua University, Taiwan, to whom I am most grateful. See the section on Chao1 and Chao2 in the main text of this User's Guide for information on sufficient sample size.

Estimator	Singletons (F_1) or Uniques (Q_1)	Doubletons (F_2) or Duplicates (Q_2)	Setting	Estimate	Variance	95% CI
Chao1	$F_1 > 0$	$F_2 > 0$	Classic	Eq. 1	Eq. 5	Eq. 13
			Bias-corrected	Eq. 2	Eq. 6	Eq. 13
	$F_1 > 1$	$F_2 = 0$	Either	Eq. 2	Eq. 7	Eq. 13
	$F_1 = 1$	$F_2 = 0$	Either	*S(obs)*	Eq. 8	Eq. 14
	$F_1 = 0$	$F_2 > 0$				
	$F_1 = 0$	$F_2 = 0$				
Chao2	$Q_1 > 0$	$Q_2 > 0$	Classic	Eq. 3	Eq. 9	Eq. 13
			Bias-corrected	Eq. 4	Eq. 10	Eq. 13
	$Q_1 > 0$	$Q_2 = 0$	Either	Eq. 4	Eq. 11	Eq. 13
	$Q_1 = 1$	$Q_2 = 0$	Either	*S(obs)*	Eq. 12	Eq. 14
	$Q_1 = 0$	$Q_2 > 0$				
	$Q_1 = 0$	$Q_2 = 0$				

Equations referenced in the table above:

$$(1)\ \hat{S}_{Chao1} = S_{obs} + \left(\frac{n-1}{n}\right)\frac{F_1^2}{2F_2}$$

$$(2)\ \hat{S}_{Chao1} = S_{obs} + \left(\frac{n-1}{n}\right)\frac{F_1(F_1-1)}{2(F_2+1)}$$

$$(3)\ \hat{S}_{Chao2} = S_{obs} + \left(\frac{m-1}{m}\right)\frac{Q_1^2}{2Q_2}$$

$$(4)\ \hat{S}_{Chao2} = S_{obs} + \left(\frac{m-1}{m}\right)\left(\frac{Q_1(Q_1-1)}{2(Q_2+1)}\right)$$

$$(5)\ \hat{\mathrm{var}}\left(\hat{S}_{Chao1}\right) = F_2\left[\frac{1}{2}\left(\frac{n-1}{n}\right)\left(\frac{F_1}{F_2}\right)^2 + \left(\frac{n-1}{n}\right)^2\left(\frac{F_1}{F_2}\right)^3 + \frac{1}{4}\left(\frac{n-1}{n}\right)^2\left(\frac{F_1}{F_2}\right)^4\right]$$

$$(6)\ \hat{\mathrm{var}}\left(\hat{S}_{Chao1}\right) = \left(\frac{n-1}{n}\right)\frac{F_1(F_1-1)}{2(F_2+1)} + \left(\frac{n-1}{n}\right)^2\frac{F_1(2F_1-1)^2}{4(F_2+1)^2} + \left(\frac{n-1}{n}\right)^2\frac{F_1^2F_2(F_1-1)^2}{4(F_2+1)^4}$$

$$(7)\ \hat{\mathrm{var}}\left(\hat{S}_{Chao1}\right) = \left(\frac{n-1}{n}\right)\frac{F_1(F_1-1)}{2} + \left(\frac{n-1}{n}\right)^2\frac{F_1(2F_1-1)^2}{4} - \left(\frac{n-1}{n}\right)^2\frac{F_1^4}{4\hat{S}_{Chao1}}$$

$$(8)\ \hat{\mathrm{var}}(S_{obs}) \approx \sum_{i\geq 1} F_i\left(e^{-i} - e^{-2i}\right) - \frac{1}{n}\left(\sum_{i\geq 1} i e^{-i} F_i\right)^2$$

$$(9)\ \hat{\mathrm{var}}\left(\hat{S}_{Chao2}\right) = Q_2\left[\frac{1}{2}\left(\frac{m-1}{m}\right)\left(\frac{Q_1}{Q_2}\right)^2 + \left(\frac{m-1}{m}\right)^2\left(\frac{Q_1}{Q_2}\right)^3 + \frac{1}{4}\left(\frac{m-1}{m}\right)^2\left(\frac{Q_1}{Q_2}\right)^4\right]$$

$$(10)\quad \hat{\text{var}}\left(\hat{S}_{Chao2}\right) = \left(\frac{m-1}{m}\right)\frac{Q_1(Q_1-1)}{2(Q_2+1)} + \left(\frac{m-1}{m}\right)^2 \frac{Q_1(2Q_1-1)^2}{4(Q_2+1)^2} + \left(\frac{m-1}{m}\right)^2 \frac{Q_1^2 Q_2 (Q_1-1)^2}{4(Q_2+1)^4}$$

$$(11)\quad \hat{\text{var}}\left(\hat{S}_{Chao2}\right) = \left(\frac{m-1}{m}\right)\frac{Q_1(Q_1-1)}{2} + \left(\frac{m-1}{m}\right)^2 \frac{Q_1(2Q_1-1)^2}{4} - \left(\frac{m-1}{m}\right)^2 \frac{Q_1^4}{4\hat{S}_{Chao2}}$$

$$(12)\quad \hat{\text{var}}(S_{obs}) \approx \sum_{i\geq 1} Q_i\left(e^{-i} - e^{-2i}\right) - \frac{1}{m}\left(\sum_{i\geq 1} ie^{-i}Q_i\right)^2$$

$$(13)\quad \text{Lower 95\% Bound} = S_{obs} + \frac{T}{K},\ \text{Upper 95\% Bound} = S_{obs} + TK,$$

$$\text{where } T = Chao - S_{obs}\text{ , and } K = \exp\left\{1.96\left[\log\left(1 + \frac{\hat{\text{var}}\left(\hat{S}_{Chao}\right)}{T^2}\right)\right]^{1/2}\right\}$$

$$(Chao \text{ may be either } Chao1 \text{ or } Chao2)$$

$$(14)\quad \text{Lower 95\% Bound} = \max\left(S_{obs}, \frac{S_{obs}}{(1-P)} - 1.96\frac{(\hat{\text{var}}\, S_{obs})^{1/2}}{(1-P)}\right),$$

$$\text{Upper 95\% Bound} = \frac{S_{obs}}{(1-P)} + 1.96\frac{(\hat{\text{var}}\, S_{obs})^{1/2}}{(1-P)},$$

$$\text{where}\quad P = \frac{\sum_{i\geq 1} F_i e^{-i}}{S_{obs}},\ \hat{\text{var}}(S_{obs})\text{ is given in Eq. (8) for Chao1;}$$

$$\text{and}\quad P = \frac{\sum_{i\geq 1} Q_i e^{-i}}{S_{obs}},\ \hat{\text{var}}(S_{obs})\text{ is given in Eq. (12) for Chao2.}$$

Jackknife 1: First-order jackknife estimator of species richness (incidence-based) (Burnham and Overton 1978,197 Heltshe and Forrester 1983)

$$\mathbf{S}_{\text{jack1}} = \mathbf{S}_{\text{obs}} + \mathbf{Q}_1\left(\frac{\mathbf{m}-\mathbf{1}}{\mathbf{m}}\right)$$

Jackknife 2: Second-order jackknife estimator of species richness (incidence-based) (Smith and van Belle 1984)

$$S_{jack2} = S_{obs} + \left[\frac{Q_1(2m-3)}{m} - \frac{Q_2(m-2)^2}{m(m-1)}\right]$$

Bootstrap: Bootstrap estimator of species richness (incidence-based) (Smith and van Belle 1984)

$$S_{boot} = S_{obs} + \sum_{k=1}^{S_{obs}} \left(1 - p_k\right)^m .$$

ACE: Abundance Coverage-based Estimator of species richness (Chao and Lee 1992, Chao, Ma, and Yang 1993)

First note that

$$S_{obs} = S_{rare} + S_{abund} .$$

The sample coverage estimate based on abundance data is

$$C_{ace} = 1 - \frac{F_1}{N_{rare}} ,$$

where

$$N_{rare} = \sum_{i=1}^{10} iF_i .$$

Thus, this sample coverage estimate is the proportion of all individuals in rare species that are not singletons. Then the ACE estimator of species richness is

$$S_{ace} = S_{abund} + \frac{S_{rare}}{C_{ace}} + \frac{F_1}{C_{ace}} \gamma_{ace}^2$$

where the estimate the coefficient of variation of the F_i's, is

$$\gamma_{ace}^2 = \max\left[\frac{S_{rare}}{C_{ace}} \frac{\sum_{i=1}^{10} i(i-1)F_i}{\left(N_{rare}\right)\left(N_{rare} - 1\right)} - 1, 0 \right] .$$

Note: The formula for ACE is undefined when all Rare species are Singletons ($F_1 = N_{rare}$, yielding $C = 0$). In this case, EstimateS computes the bias-corrected form of Chao1 instead (on Anne Chao's advice).

ICE: Incidence Coverage-based Estimator of species richness (Lee and Chao 1994)

First note that

$$S_{obs} = S_{infr} + S_{freq} .$$

The sample coverage estimate based on incidence data is

$$C_{ice} = 1 - \frac{Q_1}{N_{infr}} ,$$

where

$$N_{infr} = \sum_{j=1}^{10} jQ_j .$$

Thus, the sample coverage estimate is the proportion of all individuals in infrequent species that are not uniques. Then the ICE estimator of species richness is

$$S_{ice} = S_{freq} + \frac{S_{infr}}{C_{ice}} + \frac{Q_1}{C_{ice}} \gamma^2_{ice}$$

.

where the estimate the coefficient of variation estimates the coefficient of variation of the Q_j's, is

$$\gamma^2_{ice} = \max\left[\frac{S_{infr}}{C_{ice}} \frac{m_{infr}}{(m_{infr-1})} \frac{\sum_{j=1}^{10} j(j-1)Q_j}{(N_{infr})^2} - 1, 0\right]$$

.

Note: The formula for ICE is undefined when all Infrequent species are Uniques ($Q_1 = N_{infr}$, yielding $C = 0$). In this case, EstimateS computes the bias-corrected form of Chao2 instead (on Anne Chao's advice)

Appendix C: Coverage-based Estimator of Shared Species

This appendix and its implementation in EstimateS is based on Chao et al. (2000) and on personal communication
Anne Chao, Institute of Statistics, National Tsing Hua University, Taiwan.

Definition of variables

V_{est}	Estimated number of species shared by samples *j* and *k*
V_{obs}	Observed number of species shared by samples *j* and *k*
$V_{jk(abund)}$	Observed number of shared, abundant species (>10 individuals in sample j, in sample *k,* or in both)
$V_{jk(rare)}$	Observed number of shared, rare species (< or = 10 individuals in sample *j* AND < or = 10 individuals in sample *k*)
$X_1 \cdots X_i \cdots X_{V_{jk(rare)}}$	Number of individuals of rare, shared species *i* in sample *j*
$Y_1 \cdots Y_i \cdots Y_{V_{jk(rare)}}$	Number of individuals of rare, shared species *i* in sample *k*
$f_{1\cdot}$	Total number of singletons ($X_i = 1$) among rare, shared species in sample *j*
$f_{\cdot 1}$	Total number of singletons ($Y_i = 1$) among rare, shared species in sample *k*
f_{1+}	Number of rare, shared species that are singletons in sample j but have $Y_i > 1$ in sample *k*
f_{+1}	Number of rare, shared species that are singletons in sample k but have $X_i > 1$ in sample *j*

f_{11}	Number of rare, shared species that are singletons in both samples *j* and *k*
N_{1+}	Number of individuals in sample k for rare, shared species that are singletons in sample *j*
N_{+1}	Number of individuals in sample *j* for rare, shared species that are singletons in sample *k*
C_{jk}	Sample coverage for rare, shared species

The estimator

Sample coverage for rare, shared species is estimated by

$$C_{jk} = 1 - \frac{(N_{1+} + N_{+1} - f_{11})}{\sum_{i=1}^{V_{jk(rare)}} X_i Y_i},$$

where the summation is taken over all rare, shared species. An estimate of the true number of rare, shared species samples j and k, uncorrected for variation (among species) and covariation (among species between samples) in abundance is

$$\hat{V}_{jk(rare)} = \frac{V_{jk(rare)}}{C_{jk}}.$$

With variation and covariation in abundance taken into account, estimated true number of shared species for sampl and k (the result that EstimateS produces) is then

$$V_{jk(est)} = V_{jk(abund)} + \hat{V}_{jk(rare)} + \frac{1}{C_{jk}}\left(f_{1+}\gamma_j + f_{+1}\gamma_k + f_{11}\gamma_{jk}\right),$$

where the gamma terms are estimates of the coefficients of variation and covariation in abundance among rare, sha species. The gamma terms are computed as

$$\gamma_j = \frac{\hat{V}_{jk(rare)} s_{21}}{s_{10} s_{11}} - 1$$

$$\gamma_k = \frac{\hat{V}_{jk(rare)} s_{12}}{s_{01} s_{11}} - 1$$

$$\gamma_{jk} = \frac{\hat{V}_{jk(rare)} s_{22}}{s_{10} s_{01} s_{11}} - \frac{\hat{V}_{jk(rare)} s_{11}}{s_{10} s_{01}} - \gamma_j - \gamma_k$$

,

where, taking all summations over

$$i = 1 \ldots V_{jk(rare)}$$

we have

$$s_{10} = \sum X_i$$

$$s_{01} = \sum Y_i$$

$$s_{11} = \sum X_i Y_i$$

$$s_{12} = \sum X_i Y_i (Y_i - 1)$$

$$s_{21} = \sum X_i (X_i - 1) Y_i$$

$$s_{22} = \sum X_i (X_i - 1) Y_i (Y_i - 1)$$

.

Note: Sample size terms in the numerator and denominator of the gammas, of the form *n/(n-1)*, appear in Chao et a (2000). Since these ratios are effectively unity, they have been omitted above and for computational purposes in EstimateS.

Appendix D: Chao's Abundance-based Jaccard and Sorensen Similarity Indices and Their Estimators

This appendix and its implementation in EstimateS is based on Chao et al. (2005) and on personal communication w Anne Chao, Institute of Statistics, National Tsing Hua University, Taiwan.

Appendix D is a pdf document. Click here to display Appendix D in Acrobat or Acrobat Reader. For full details, download Chao et al. (2005).

EstimateS

Statistical Estimation
of Species Richness
and Shared Species
from Samples

Home | About EstimateS | EstimateS Download | EstimateS User's Guide | Support | EstimateS & Biota | Links

About EstimateS
Major Features
Platforms
Who Uses EstimateS?
Publications That Cite EstS

EstimateS Download
EstS 9.1 Windows
EstS 9.1 Mac OS
EstS 8.2 Windows
EstS 8.2 Mac OS

EstimateS User's Guide
Introduction
Preparing a Data Input File
Running EstimateS
Additional Notes
What EstimateS Computes
Things You Should Know
References
Appendices

Support
Support by Email
Bug Reporting
Known Bugs
Notes for Mac OS
Notes for Windows
Your Suggestions
Recent Version History

EstimateS & Biota

Links

EstimateS Support

The EstimateS User's Guide

EstimateS support starts, and usually ends, with the *EstimateS User's Guide*, which hasdeliberately been written in much more detail than most manuals, to provide you with full access to the information you need.

"How To" Questions: The User's Guide is the first place to go to find out if EstimateS can do something you want done. Use the comprehensive *Table of Contents* or your browser's search tool to see if you can find the answer you need.

Support by Email

E-mailed questions that can be answered by consulting the EstimateS User's Guide will receive the lowest priority for email support, or may not be answered at all. Please understand that EstimateS has thousands of users, and exactly one busy professor to keep up with support email.

Technical problems with EstimateS or problems with the EstimateS Website .You may request email help with problems you can't solve with the help of the *EstimateS User's Guide*. Please contact EstimateS's author, Robert Colwell at colwell@uconn.edu.

When preparing an email with a request for for support, please be sure to include the following:

1. **The word "EstimateS"** in the message Subject.
2. **Version of EstimateS.** From the File menu in EstimateS, select "About EstimateS." Record the Version and Date of the copy of EstimateS that you are using.
3. Your computer's **Operating System.**
4. **A detailed description** of what you are trying to do, or the problem you have doing it.

Please do not telephone the Author with support questions.

Bug Reporting

If EstimateS does something you did not expect, first try consulting the *EstimateS User's Guide* to see if the behavior is normal. If something happens that is clearly not intended, you may have found a bug. If a 4th Dimension error window appears, you have definitely found a bug.

If you believe you have discovered a bug**,** please email EstimateS's author, Robert Colwell at colwell@uconn.edu.

When preparing an email with a bug report, please be sure to include the following:

1. **The word "EstimateS"** in the message Subject.
2. **Version of EstimateS.** From the File menu in EstimateS, select "About EstimateS."

Record the Version and Date of the copy of EstimateS that you are using.

3. Your computer's **Operating System.**
4. **A description of the problem.** To be most useful the description of the problem should include a full account of the steps that led up to it. If at all possible, try making it happen again until you are sure what steps reliably produce the problem. If the problem cannot be reproduced, it cannot be fixed.
5. **Any data that were involved.** If the problem seems to concern a particular data set, please send the input file as an attachment.

Known Bugs and Problems

- ***ACE*** is not correctly computed for individual-based data. Chao1 is reported instead. This bug will be fixed in the next version.

Notes for Mac OS X

- ***None at this time***.

Notes for Windows

- **If you get a 4D *"Create a New Segment" Message or "Cannot save changes to file" message.*** If you get a message saying you cannot save changes to file, or an offer to "Create a new segment," your system is probably set up to write-protect files. You need to have *Write* privileges to use EstimateS. See your IT administrator.

Your Suggestions for EstimateS

EstimateS's current tools and features evolved over a period of nearly 20 years, partly under the guidance of EstimateS users. Suggestions for improvement are welcomed. Please email EstimateS's author, Robert Colwell at colwell@uconn.edu. Please include the word "EstimateS" in the Subject line of your message.

Recent Version History

New in EstimateS 9.1.0:

- Chao1 and Chao 2 computations have now been comprehensively revised to take account of more special cases and to reduce small-sample bias. See Appendix B in the User's Guide.
- Individual-based rarefaction is now carried out using the classic multinomial model, instead of the Poisson model, as in Version 9.0.0. Individual-based extrapolation continues to use the multinomial model.
- For Individual-based rarefaction, a bootstrap unconditional variance estimator for special cases has been implemented (all cases in which Sest = Sobs for the reference sample), based on Appendix G of Chao et al. 2013. Similar estimators for theses cases will be implemented for Individual-based extrapolation and for sample-based rarefaction and extrapolation in a future version.
- **Bug fixed:** The on-screen settings for Skip Columns and Skip Rows for Individual-based Data File input did not work. Fixed in 9.1.0.

New in EstimateS 9.0.0:

- ***A comprehensively revised User's Guide*** covering all the new features in EstimtateS 9 and all the traditional ones of previous versions.

- An entirely new capability for handing ***individual-based rarefaction with true (unconditional***) ***confidence intervals*** (and of course, sample-based rarefaction, the core of all previous versions of EstimateS).

- ***Rarefaction of richness estimators and diversity indices for individual-based data*** (as well as sample-based data, as in previous versions).

- Non-parametric ***extrapolation of rarefaction curves*** for both sample-based and individual-based data (Colwell et al. 2012).

- ***Batch input and export*** option for both sample-based and individual-based datasets.

- Options for computing, displaying, and exporting subsets of results for ***evenly-spaced intervals*** for rarefaction and extrapolation of samples or individuals (interval-sampling or *knots*).

- Automatic support for ***International and US number formats.***

- **EstimateS 9 for Windows** runs under Windows 8, Windows 7, Vista, and XP.

- **EstimateS 9 for Mac OS** runs under OS 10.5 (Leopard) through 10.8 (Mountain Lion).

- **EstimateS 9 is blazing fast,** compared to earlier versions.

New in EstimateS 8.2.0:

- In the Diversity Indexes output, the exponential form of Shannon diversity (e^H) and its conditional standard deviation are reported (as well as H itself, using natural logarithms, as in previous versions).

- Bugs fixed:
  - In Version 8.00 (but not in earlier versions), the application froze if the Diversity Indexes option was checked and one or more samples in the input dataset were completely empty (no individuals of any species).   This bug has been fixed in Version 8.20.
  - In Version 8.00 and Version 7.52, the Morisita-Horn similarity index computations experienced an untrapped overflow when input numbers were extremely large, yielding nonsensical results (including M-H values less than zero and greater than unity). This bug has been fixed in Version 8.20.
  - Under some conditions, Version 8 did not read *Biota* output files correctly. This bug has been fixed in Version 8.20.

New in EstimateS 8.0.0:

- Built around a new version of 4D Engine that ensures compatibility with the latest operating systems.
- Option to export results from *n* individual randomizations to a text file, allowing computation of precision, accuracy, and other analyses.
- Improved detection and reporting on input file problems.
- Corrected all known bugs in previous versions, including an error in confidence interval computations for Chao1 and Chao2 (with generally minor effects).

EstimateS

Statistical Estimation
of Species Richness
and Shared Species
from Samples

Home | About EstimateS | EstimateS Download | EstimateS User's Guide | Support | EstimateS & Biota | Links

About EstimateS
Major Features
Platforms
Who Uses EstimateS?
Publications That Cite EstS

EstimateS Download
EstS 9.1 Windows
EstS 9.1 Mac OS
EstS 8.2 Windows
EstS 8.2 Mac OS

EstimateS User's Guide
Introduction
Preparing a Data Input File
Running EstimateS
Additional Notes
What EstimateS Computes
Things You Should Know
References
Appendices

Support
Support by Email
Bug Reporting
Known Bugs
Notes for Mac OS
Notes for Windows
Your Suggestions
Recent Version History

EstimateS & Biota

Links

EstimateS and *Biota: The Biodiversity Database Manager*

- *Biota: Biodiversity Database Manager* is a cross-platform (Windows or Mac OS) application for specimen-based, biological data and collections management (http://viceroy.eeb.uconn.edu/biota).
- *Biota* is an excellent tool for managing the kinds of data that *EstimateS* analyzes.
- Among many other text file export options, *Biota* exports samples-by-taxon incidence or abundance matrixes to delimited text files, based on any selection of records at any taxonomic rank (Species through Kingdom) and any set of Collection (sample) or Locality records.
- You can tell *Biota* to export these matrices in a special format designed for *direct* input to EstimateS. Likewise, *EstimateS* has a special input option to read these files.
- *Biota 3* has been released under a GNU–GPU (General Purpose License). Fully enabled copies of the standalone desktop versions of Biota 3 (BiotaAppWin and BiotaAppMac) as well as the comprehensive User's Guide may now be freely downloaded, used, and distributed at no cost.
  Biota 3 is an updated version for current operating systems for both Windows and Macintosh. For more details, and to download, please visit the Biota website.

EstimateS

Statistical Estimation of Species Richness and Shared Species from Samples

Home | About EstimateS | EstimateS Download | EstimateS User's Guide | Support | EstimateS & Biota | Links

About EstimateS
Major Features
Platforms
Who Uses EstimateS?
Publications That Cite EstS

EstimateS Download
EstS 9.1 Windows
EstS 9.1 Mac OS
EstS 8.2 Windows
EstS 8.2 Mac OS

EstimateS User's Guide
Introduction
Preparing a Data Input File
Running EstimateS
Additional Notes
What EstimateS Computes
Things You Should Know
References
Appendices

Support
Support by Email
Bug Reporting
Known Bugs
Notes for Mac OS
Notes for Windows
Your Suggestions
Recent Version History

EstimateS & Biota

Links

Related Links

- ***Anne Chao's SPADE (Species Prediction And Diversity Estimation) and iNEXT (iNTerpolation and EXTrapolation).***

  SPADE and the SPADE User's Guide can be downloaded at the SPADE website.

  iNEXT can be *used online here*. iNEXT computes some of the same statistics as EstimateS, plus others (especially coverage-based rarefaction and extrapolation) that EstimateS does not compute.

  For iNext, the user provides a vector of abundances of individual species (abundance data) or incidences of individual species (incidence data). iNEXT computes the following species richness estimates and associated 95% confidence intervals:

  - Sample-size-based rarefaction and extrapolation: species richness estimates for rarefied and extrapolated samples up to a size specified by the user (i.e., an endpoint, see below). Refer to Colwell et al. (2012) for details.
  - Coverage-based rarefaction and extrapolation: species richness estimates for rarefied and extrapolated samples for sample coverage up to a coverage specified by the user (i.e., an endpoint, see below). Refer to Chao and Jost (2012) for details.

  iNEXT also plots the following three integrated sampling curves suggested in Chao et al. (2013) for unified sampling and estimation in species diversity analysis.

  - Sample-size-based rarefaction and extrapolation sampling curve: this curve plots the species richness estimates for a rarefied and extrapolated sample with respect to sample size up to a specified endpoint.
  - Sample completeness curve: this curve plots the sample completeness (as measured by sample coverage) with respect to sample size. The curve provides a bridge between sample-size- and coverage-based rarefaction and extrapolation.
  - Coverage-based rarefaction and extrapolation sampling curve: this curve plots the species richness estimates for rarefied sample and extrapolated sample with respect to sample coverage up to a specified endpoint.

- ***SAM: Spatial Analysis in Macroecology.*** SAM, downloadable freeware developed by Thiago Rangel and colleagues, is a compact but robust computer program designed as a package of statistical tools for spatial analysis, mainly for applications in Macroecology and Biogeography. SAM runs under Microsoft Windows as as a user-friendly, menu-driven, graphical interface computational program. SAM offers a wide spectrum of statistical methods currently used in Surface Pattern Spatial Analysis. Sam fills most current scientific and analytical needs of Macroecologists, Biogeographers and Geographical Ecologists who study broad scale biological patterns and processes.

- ***BioDiversity Pro.*** *BioDiversity Pro* was designed and developed by Neil McAleece. This package was devised jointly by P.J.D. Lambshead and G.L.J. Paterson of the The Natural History Museum in London and J.D. Gage of the Scottish Association for Marine Science, Oban, Scotland.
- ***Jari Oksanen's VEGAN Package for R.*** The *VEGAN* package is intended to help vegetation ecologists and other community ecologists to use *R*. It contains all major ordination methods, ecologically meaningful dissimilarity indices, tools to analysis of diversity, species richness and abundance models, plus numerous support functions
- If you would like to see a link to your software here, please write.