



Image - Scott Henderson

© Pisces Conservation 2007



SDR-IV Help

Measuring and understanding biodiversity

by R M H Seaby and P A Henderson

SDR is designed to help you measure and understand the variation in community composition that often exists between different locations.

Using either presence/absence data or quantitative data, it offers a wide range of both tried and tested and exciting new techniques to handle and analyse the complex data that ecological sampling often generates.

Using the latest data handling techniques, SDR can give you an insight into your data in seconds, and with vivid graphics, enable you to print and publish your results simply, quickly and easily.

SDR-IV Help

© **Pisces Conservation 2007**

All rights reserved. No parts of this work may be reproduced in any form or by any means - graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems - without the written permission of the publisher.

Products that are referred to in this document may be either trademarks and/or registered trademarks of the respective owners. The publisher and the author make no claim to these trademarks.

While every precaution has been taken in the preparation of this document, the publisher and the author assume no responsibility for errors or omissions, or for damages resulting from the use of information contained in this document or from the use of programs and source code that may accompany it. In no event shall the publisher and the author be liable for any loss of profit or any other commercial damage caused or alleged to have been caused directly or indirectly by this document.

Printed: November 2007 in Lymington, Hampshire

Publisher

Pisces Conservation Ltd.

Table of Contents

Foreword	I
Part I Introduction	2
Part II System requirements	5
Part III Installation	7
Part IV How to reference this program	9
Part V Tutorials and demonstration data sets	11
1 Essential steps for the analysis of diversity	11
2 Diversity Ordering tutorial	12
3 Species accumulation tutorial	15
4 Rank abundance plots	17
5 Comparing the index of two samples	20
Part VI Data exploration	23
1 Comparing samples	23
2 Summary Statistics	23
Part VII Using Species Diversity & Richness	27
1 Entering data	28
2 Maximum size of the data set	29
3 File open button	29
4 Creating a new data set within SDR	29
5 Creating a data file using a spread sheet	30
6 Add species	32
7 Add sample	32
8 Open an existing data file	32
9 Selecting the Samples to Analyse	34
10 Non-Integer Data	35
11 Editing data	35
12 Printing and exporting your results	36
13 Saving a data set	38
14 Closing the program	38
15 Maximum size of a variable	38
16 Bottom bar	38
Part VIII Alpha diversity	40
1 Shannon-Wiener Index	41

2 Simpson's Index	42
3 Species Number	43
4 Margalef D	43
5 Berger-Parker Dominance	44
6 McIntosh D	45
7 Brillouin D	45
8 Fisher's alpha	46
9 Q statistic	46
10 Menhinick Index	47
11 Strong's Index	47
12 Bootstrap estimation of confidence intervals	48
13 Jackknifing diversity	49
How jackknifing works	50
14 Testing for significant differences	50

Part IX Evenness 54

1 Smith and Wilson's view	55
2 Pielou J (All samples)	55
3 Pielou J (Each Sample)	55
4 McIntosh E	55
5 Brillouin E	56
6 Heip	56
7 Simpson's E	56
8 NHC	57
9 Camargo	57
10 Smith & Wilson B	57
11 Smith & Wilson 1-D	58
12 Smith & Wilson -lnD	58
13 Shannon maximum	58
14 Shannon minimum	58
15 Gini	59

Part X SHE Analysis 62

1 SHE analysis	62
----------------------	----

Part XI Diversity ordering 65

1 Renyi family	66
2 Right-tailed Sum Diversity Ordering	66

Part XII Species richness estimators 69

1 Heterogeneity test	70
2 Chao Quantitative	71
3 Chao Presence/absence	72
4 Chao & Lee 1	72

5	Chao & Lee 2	73
6	1st Order Jackknife	74
7	2nd Order jackknife	75
8	Bootstrap	75
9	Michaelis-Menten	76
10	Pooled rarefaction	77
11	Single sample rarefaction	78
12	Across sample rarefaction	80
13	Sample interpolation	81
14	Henderson	82
15	Number of Random Selections	84
16	Using the Michaelis-Menten method as a stopping rule	85
17	Using parametric models based on relative abundance	85

Part XIII Abundance 87

1	Species Number	87
2	Rank Abundance	87
3	Species Accumulation Curves	89
4	K dominance	89

Part XIV Fitting abundance distributions 92

1	Geometric	92
2	Log series	93
3	Truncated log-normal	94
4	Broken stick	95

Part XV Data Simulation 98

1	Data Simulation	98
---	-----------------------	----

Part XVI Beta diversity 101

1	Beta diversity indices	101
---	------------------------------	-----

Part XVII Special indices 104

1	BMWP score	104
	Habitat specific BMWP	106
	Revised BMWP	107
	Original BMWP	109
2	T-Group	110
3	ASPT index	110
4	Irish water quality rating	110
5	LIFE scores	114

Part XVIII References 116

1	References	116
---	------------------	-----

Part XIX Contact Pisces 119

1 Contact Pisces	119
Index	120

Licence Agreement

PISCES LICENSE AGREEMENT

This is a legal agreement between you the end user and PISCES Conservation Ltd. Lymington (PISCES). BY OPENING THIS PACKAGE YOU ARE AGREEING TO BE BOUND BY THE TERMS OF THIS AGREEMENT. IF YOU DO NOT AGREE TO THE TERMS OF THIS AGREEMENT PROMPTLY RETURN THE UNOPENED PACKAGE AND ALL ACCOMPANYING ITEMS (including written material) TO THE PLACE YOU OBTAINED THEM FOR A FULL REFUND.

1. GRANT OF LICENSE - This PISCES License Agreement ('License') permits you to use one copy of the PISCES software product acquired with this License (SOFTWARE) on any single computer, provided the SOFTWARE is in use on only one computer at any time. If you have multiple Licenses for the SOFTWARE then at any time, you may have as many copies of the SOFTWARE in use as you have Licenses. The SOFTWARE is 'in use' on a computer when it is loaded into the temporary memory (i.e. RAM) or installed into the permanent memory (e.g. hard disk, CD ROM, or other storage device) of that computer, except that a copy installed on a network server for the sole purpose of distribution to other computers is not 'in use'. If the anticipated number of users of the SOFTWARE will exceed the number of applicable Licenses then you must have a reasonable mechanism or process in place to assure that the number of persons using the SOFTWARE concurrently does not exceed the number of Licenses. If the SOFTWARE is permanently installed on the hard disk or other storage device of a computer (other than network server) and one person uses that computer more than 80% of the time it is in use then that person may also use the SOFTWARE on a portable or home computer.

2. COPYRIGHT - The SOFTWARE is owned by PISCES or its suppliers and is protected by all applicable national laws. Therefore, you must treat the SOFTWARE like any other copyrighted material (e.g. a book) except that if the software is not copy protected you may either (a) make one copy of the of the SOFTWARE solely for backup or archival purposes, or (b) transfer the SOFTWARE to a single hard disk provided you keep the original solely for backup or archival purpose. You may not copy the Product manual(s) or written materials accompanying the SOFTWARE.

3. OTHER RESTRICTIONS - You may not rent or lease the SOFTWARE, but you may transfer your rights under this PISCES License Agreement on a permanent basis provided you transfer all copies of the SOFTWARE and all written materials, and the recipient agrees to the terms of this Agreement. You may not reverse engineer, decompile or disassemble the SOFTWARE. Any transfer must include the most recent update and all prior versions.

LIMITED WARRANTY - PISCES warrants that (a) the SOFTWARE will perform substantially in accordance with the accompanying Product Manual(s) for a period of 90 days from the date of receipt; and (b) any PISCES supplied hardware accompanying the SOFTWARE will be free from defects in materials and workmanship under nominal use and service for a period of one year from the date of receipt. Any implied warranties on the SOFTWARE and hardware are limited to 90 days and one (1) year respectively or the shortest period permitted by applicable law, whichever is greater.

CUSTOMER REMEDIES - PISCES'S entire liability and your exclusive remedy shall be, at PISCES option, either (a) return of the price paid or (b) repair or replacement of the SOFTWARE or hardware that does not meet PISCES'S Limited Warranty, and which is returned to PISCES with a copy of your receipt. This Limited Warranty is void if failure of the SOFTWARE or hardware has resulted from accident, abuse or misapplication. Any replacement SOFTWARE will be warranted for the remainder of the original warranty period or 30 days, whichever is longer.

NO OTHER WARRANTIES - TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, PISCES DISCLAIMS ALL OTHER WARRANTIES. EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WITH RESPECT TO THE SOFTWARE, THE ACCOMPANYING PRODUCT MANUAL (S) AND WRITTEN MATERIALS, AND ANY ACCOMPANYING HARDWARE. THE LIMITED WARRANTY CONTAINED HEREIN GIVES YOU SPECIFIC LEGAL RIGHTS.

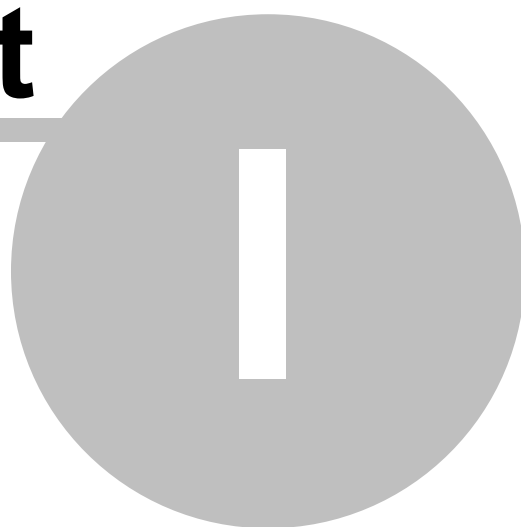
NO LIABILITY FOR CONSEQUENTIAL DAMAGES - TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW PISCES AND ITS SUPPLIERS SHALL NOT BE LIABLE FOR ANY OTHER DAMAGES WHATSOEVER (INCLUDING, WITHOUT LIMITATION, DAMAGES FOR LOSS OF BUSINESS PROFITS, BUSINESS INTERRUPTION, LOSS OF BUSINESS INFORMATION, OR OTHER PECUNIARY LOSS) ARISING OUT OF THE USE OF OR INABILITY TO USE THIS PISCES PRODUCT, EVEN IF PISCES HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. IN ANY CASE, PISCES'S ENTIRE LIABILITY UNDER ANY PROVISION OF THIS AGREEMENT SHALL BE LIMITED TO THE AMOUNT ACTUALLY PAID BY YOU FOR THE SOFTWARE.

This Agreement is governed by the laws of England.

Should you have any questions concerning this Agreement, or if you desire to contact PISCES for any reason, please use the address information enclosed in this product to contact PISCES or write:

PISCES Conservation Ltd
IRC House, The Square
Pennington, Lymington
Hampshire, England
SO41 8GN, UK
Tel 01590 674000

Part



1 Introduction

Species Diversity and Richness 4 is designed for professional ecologists and their students. The methods on offer range from the familiar, such as the calculation of various diversity indices and the fitting of common distributions, to more recently-developed techniques such as total species complement estimators. Taken together they provide a powerful suite of methods to explore, compare and analyse community structure.

In developing version 4 of this program we have consulted recent literature and have implemented the methods featured in Magurran (2004), *Measuring Biological Diversity*, and Southwood and Henderson (2002), *Ecological Methods* - 3rd edition.

Both these books are recommended for further information on the methods offered by *Species Diversity and Richness*.

The Species Diversity and Richness Version 4, first released in January 2006, includes the following additions and enhancements:

- An improved user interface and graphics
- No maximum size for the data set - all arrays are dynamically sized
- Increased range of evenness measures, (Smith & Wilson - B , $1/D$, $-\ln(d)$, and $1-D$, Camargo, NHC, Simpson E , Heip, etc).
- Addition of Menhinick, Gini and Strong's alpha diversity indices
- Addition of Harrison's beta diversity measures
- Jackknife estimation of standard error implemented
- Improved rarefaction methods and sample interpolation
- Improved help system with [tutorials](#)
- Improved BMWP scores for water quality assessment
- New species richness estimator for open systems studied long-term - Henderson
- A sample comparison feature quickly finds the species in common
- Maximum size of any observation increased to 2147483647.

Species Diversity and Richness version 3 offered considerable improvements over versions 1 & 2, with a redesigned menu system and improved graphics, data input and printing. Probably the greatest changes were the much-extended range of methods and statistical tests on offer. Additions included:

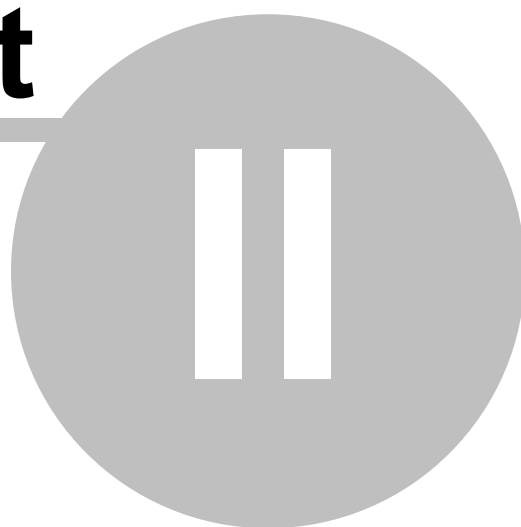
- bootstrap methods for estimating 95% confidence intervals for all diversity indices,
- a randomisation test for the significance of the difference in indices between samples,
- additional diversity measures including Fisher's a , the Q statistic and Brillouin indices,
- the right-tailed sum method for diversity ordering,
- the fitting and plotting of Geometric, log series, truncated log normal and broken stick abundance distributions to observations,
- the generation of simulated data conforming to a range of distributions
- beta diversity measures.

Version 3, completed in November 2002, offered further enhancements and additions, as itemised below:

- An improved user interface and graphics,
- Increased range evenness (equitability) measures NHC,
- 2nd order jackknife species richness measure,
- K-Dominance plots,

- SHE analysis for beta diversity,
- LIFE scores as used in water quality monitoring.

Part



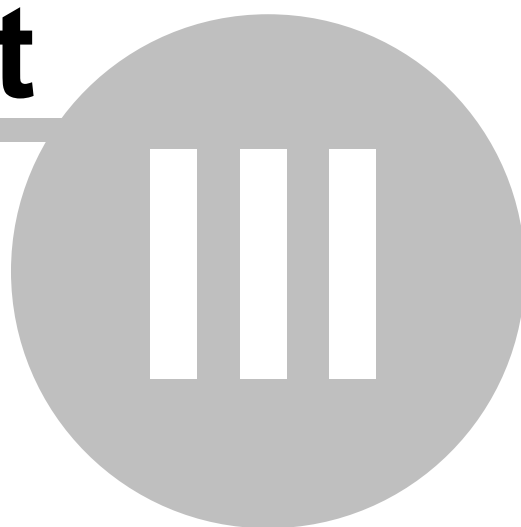
2 System requirements

- A Windows PC with CD-ROM drive.
- 11 MB of free hard disk space
- Windows XP / Vista operating system.

SDR 4 is specifically produced for Windows XP and Vista. We are not aware of any reasons why it should not work on older versions of Windows, such as ME, NT, 2000; however, we are not able to guarantee its performance under these operating systems.

We do not produce any software designed to run under the Mac OS, however, our programs should run satisfactorily using Windows emulation software. Similarly, though we do not produce software to run under Linux, our programs should run under WINE. We recommend downloading one of our demo versions to test this.

Part



3 Installation

- Place the SDR CD in the drive.
- The installation process should begin automatically; follow the on-screen instructions.
- If the CD does not auto-run, browse in Windows Explorer or My Computer, and double-click the file named **Setup.exe** in the root directory.

When installation is complete there will be a Species Diversity and Richness entry on your Start menu and a folder (directory) called *C:\Program Files\Diversity 4* which holds the Diversity program and associated files. An uninstall program will also be created, accessible from the Start menu should you wish to remove the program. If you wish to make a shortcut to the program on your desktop, go to the Species Diversity and Richness entry on the Start menu, right-click on it, and choose Send To: Desktop (create shortcut). Demonstration data files will be installed in the folder **My Documents/SDR Data**.

Part



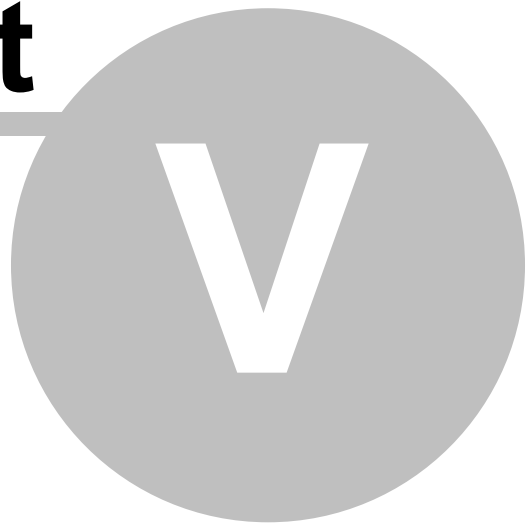
IV

4 How to reference this program

This program should be referenced as:

Seaby R. M. & Henderson, P. A. (2006) *Species Diversity and Richness Version 4*.
Pisces Conservation Ltd., Lymington, England.

Part



5 Tutorials and demonstration data sets

Upon installation of Species Diversity and Richness a number of demonstration data sets will also be placed on your hard disc. These data files are saved by default in the folder **My Documents/SDR Data**. These data sets are used in the help tutorial system.

Click on the links below to see the different tutorials

[Rank abundance plots](#) - this demo uses the file **Hinkley annual fish.csv**

[Species accumulation curves and smoothing](#) - this demo uses the file **hinkley_fish.csv**

[Checking for consistency in diversity indices](#) - this demo uses the file **Hinkley annual fish.csv**

[Comparing the index of two samples](#) - this demo uses the file **tropical butterfly demo.csv**

[A guide to a basic steps for the analysis of diversity](#)

5.1 Essential steps for the analysis of diversity

In the light of present knowledge the following may be taken as a guide to the analysis of diversity data for a single locality.

(1) Plot graph(s) of [log abundance on rank](#).

It has always been important to examine the form of your data and is easily done on a computer. Do the data form a straight line? Which are the species that depart most from it? Is there anything unusual in their biology (e.g. they could be vagrants)? The consideration of these graphs should indicate whether, exceptionally (Taylor et al., 1976), other models or indices should be used in addition to those outlined below. These graphs may be an excellent means of presenting the data for publication.

(2) Plot the [species accumulation curve\(s\)](#) and calculate total species richness, SMAX. The species accumulation curve gives insight into the sufficiency of the sampling effort. As there is no best method to calculate SMAX it is recommended that most of the methods given under [Species Richness Estimators](#) are explored.

(3) Determine alpha of the [log series](#).

(4) Calculate the [Berger- Parker dominance index](#)

(5) Calculate other parameters or statistics. This will be done where the graphs (1 above) reveal a special need.

(6) If sites are to be compared, undertake diversity ordering to ensure that they can be compared using a diversity index.

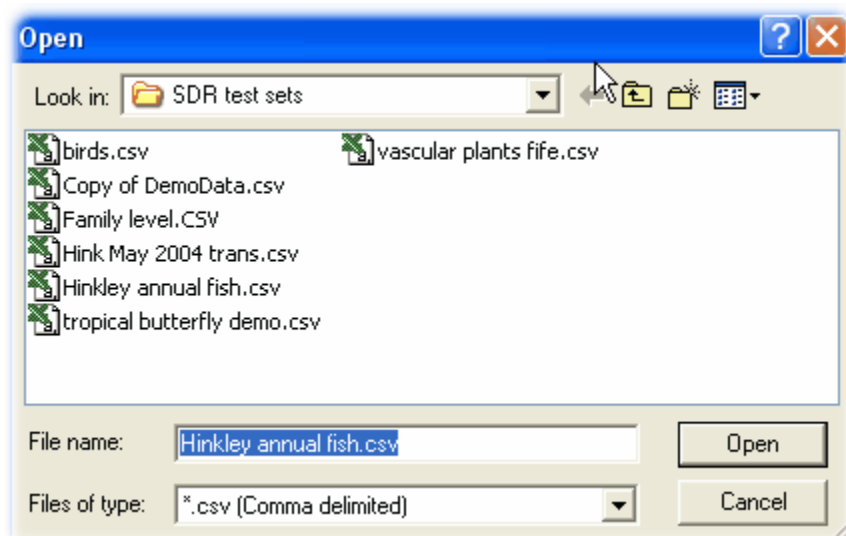
(7) Jackknife or bootstrap estimates as appropriate and run randomisation tests to compare indices.

5.2 Diversity Ordering tutorial

It is possible for a data set to give inconsistent results for alpha diversity indices, in that sample 1 will have a higher diversity than sample 2 using one index, but sample 2 will be higher than sample 1 with another index. This is because of the different weighting indices give to species number and relative species abundance. We show you here how to identify such an inconsistency.

First open the data set called **Hinkley annual fish.csv**

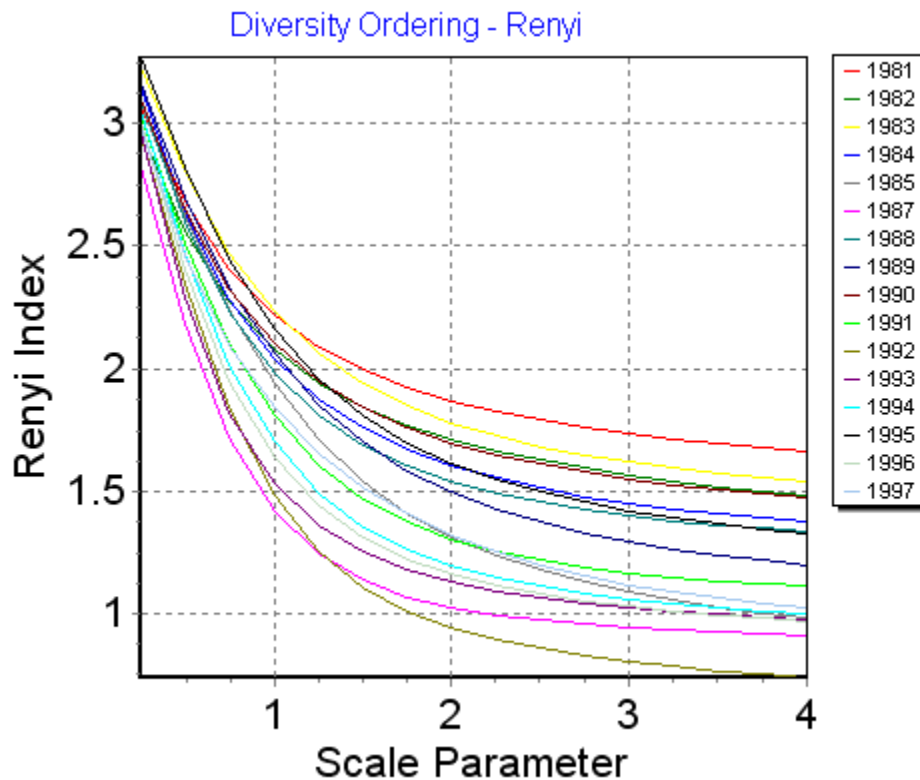
This data set is for the fish community of the Bristol Channel, England. It is derived from monthly sampling over a 25 year period and is one of the largest data sets of its type available.



Now from the Diversity Ordering tab choose Renyi.

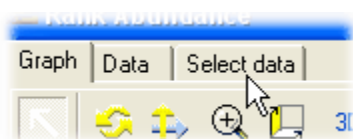


A graph showing the diversity ordering for each sample will be shown as follows:

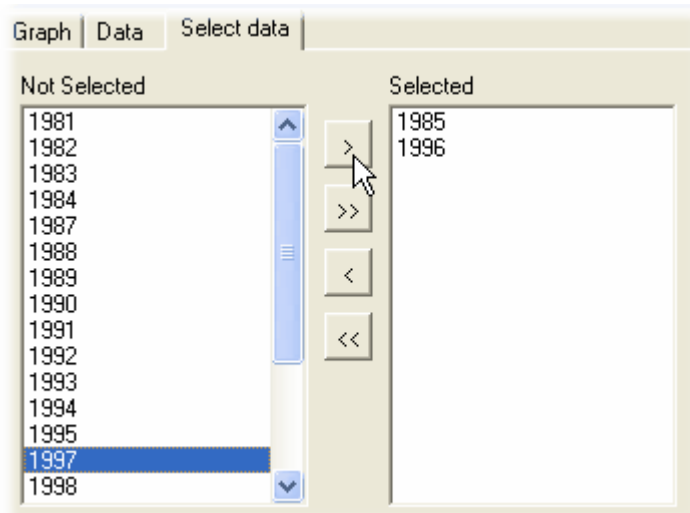


To identify inconsistencies in alpha diversity indices we must look for lines that cross. While this plot is confusing because of the number of lines that have been plotted (we will reduce them shortly) it can be seen that the bottom two lines (1987 and 1992) cross between a scale parameter of 1 and 2. This indicates that samples for these two years will be inconsistent for their Shannon, H, and Simpson's D indices.

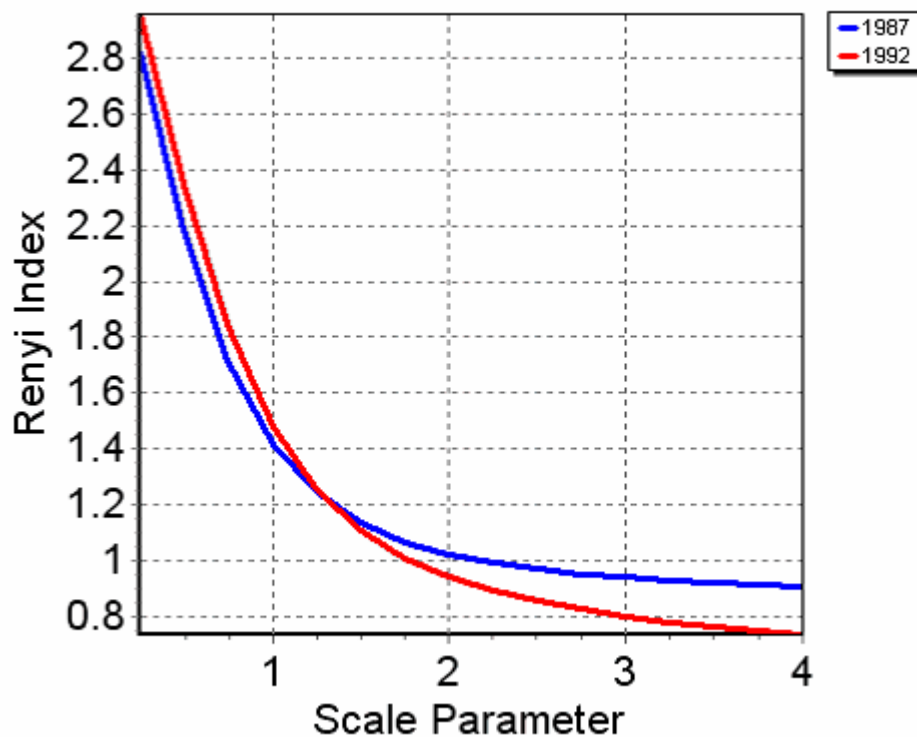
This plot is too cluttered, so you cannot see the difference between years. To select the samples (years) to plot, click on the Select data tab at the top of the graph.



Now select 1987 and 1992 only using the selection buttons between the two boxes which hold the not selected and selected samples.



Now click on the Graph tab to display the graph for 1987 and 1992. The resulting graph clearly shows the inconsistency in alpha diversity between the two years.



You can check that this inconsistency is true by looking up the alpha diversity Shannon and Simpson values for 1987 and 1992 using the Alpha diversity tab. The results are as follows:

1987: $H = 1.42$; $D = 2.778$

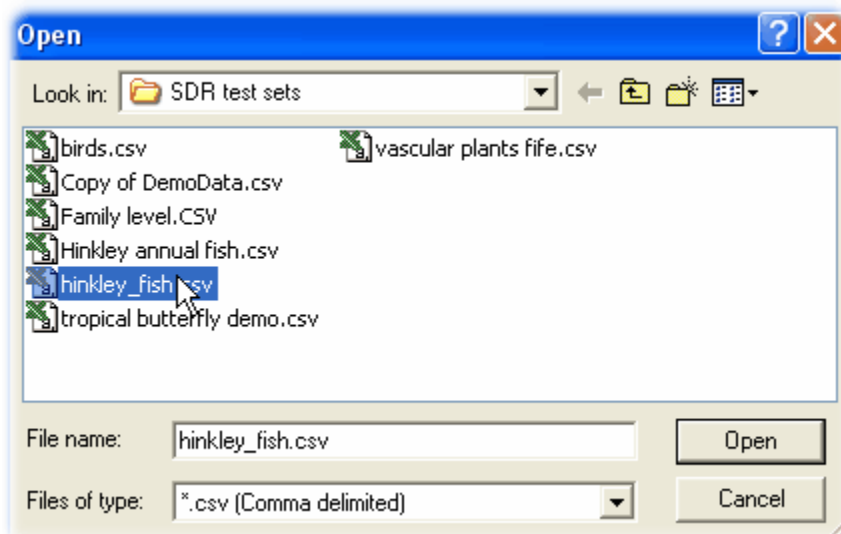
1992: $H = 1.49$; $D = 2.564$

5.3 Species accumulation tutorial

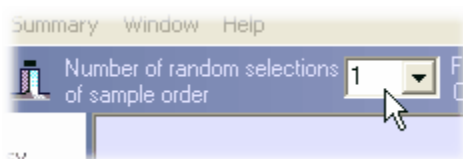
The species accumulation curve, which shows the increase in the taxa observed with sampling effort, is an excellent way to summarise the completeness of the sampling effort. In this tutorial we take you through creating observed and smoothed species accumulation curves.

First open the data set called **hinkley_fish.csv**

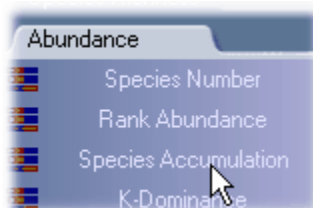
This data set is for the fish community of the Bristol Channel, England. It is a monthly record of fish samples over 184 months. It is one of the largest data sets of its type available.



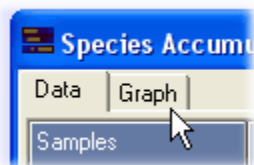
Now as we initially want to show the species accumulation curve as collected, go to the drop down box at top of the window labelled "Number of random selections of sample order" and change this number to 1. (The default is typically 10, which would automatically undertake 10 random permutations of sample order)



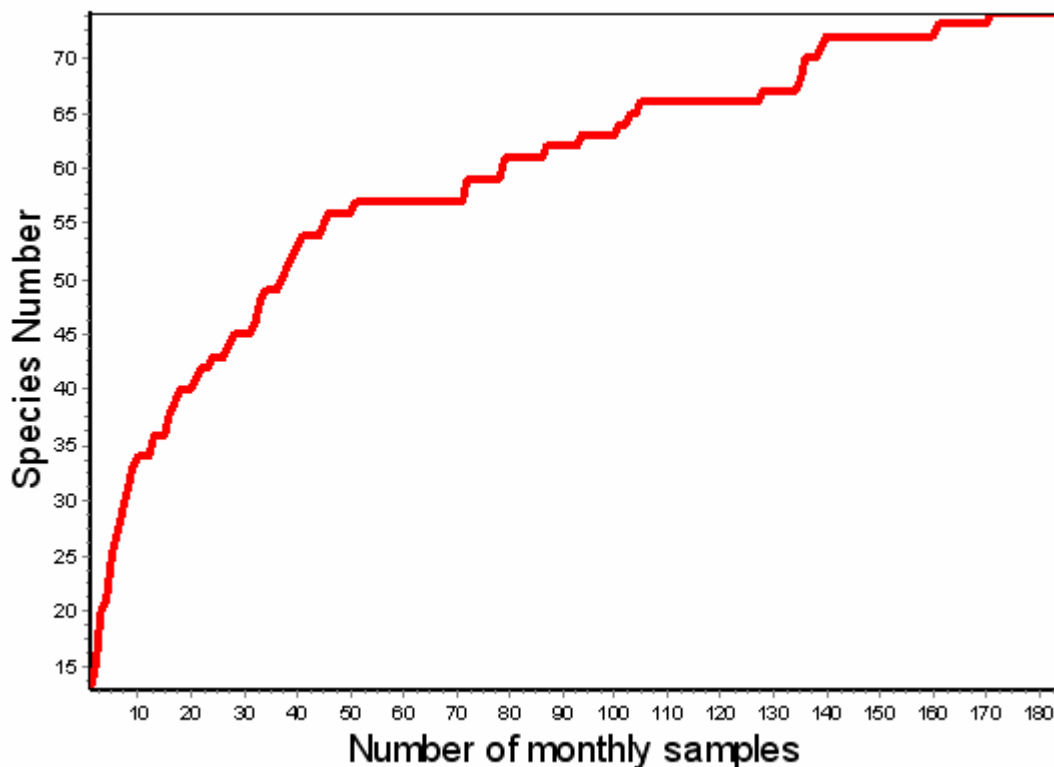
Now from the Abundance tab choose Species Accumulation.



A results window will open showing the results. Select the graph tab

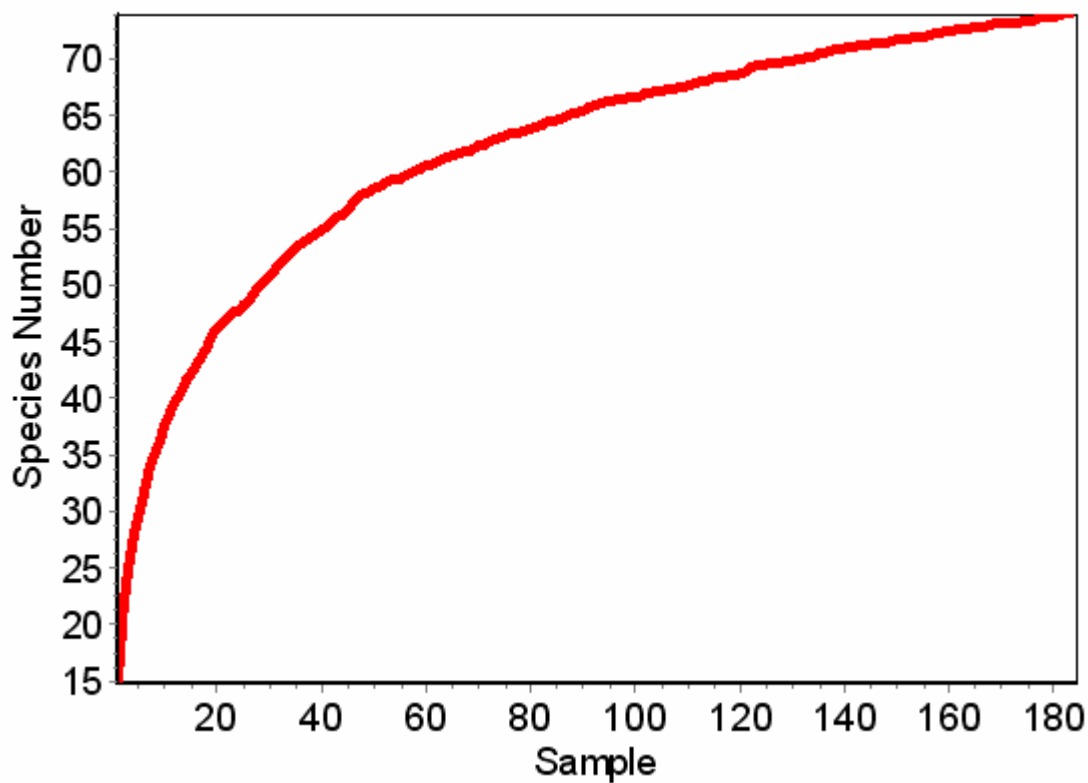


A graph showing the species accumulation will be shown as follows (for this graph we have changed the graph default fonts and line thickness using the [graph tools](#) above the plot):



Now this graph clearly shows the gradual deceleration in the acquisition of new species, however, the rate changes through time because of random sampling noise, climatic events etc. We can get a better idea of the general trend by randomly changing the sample order a number of times, and averaging the species accumulation curves produced.

To produce this smoothed curve go to the drop-down box at top of the window labelled "Number of random selections of sample order" and change this number to 30. The window showing the species accumulation curve will close down. Click again on Species accumulation under the Abundance tab, select the graph tab, and the following smoothed graph will appear.



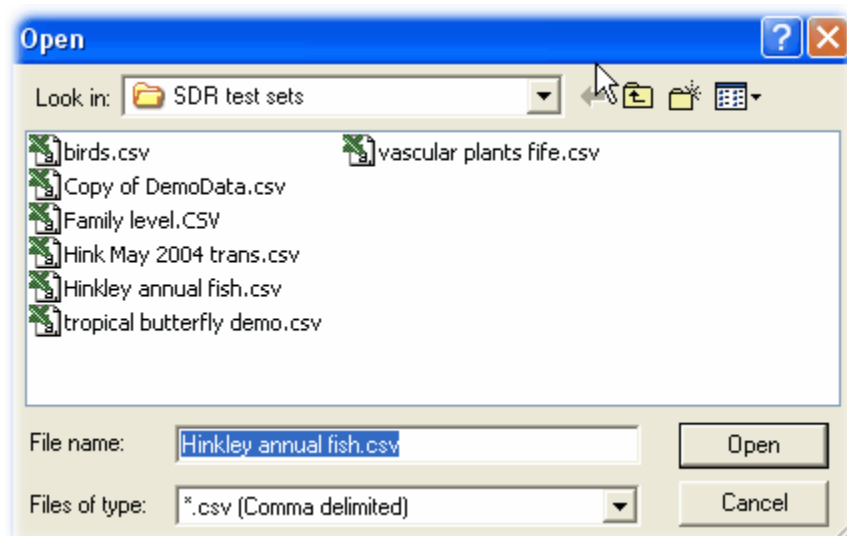
This graph shows the general trend of species acquisition with sampling effort, adjusted for random sample order.

5.4 Rank abundance plots

It is very useful to examine the pattern of relative abundance between species in your samples. The rank abundance plot is often a very useful way to summarise both equitability in abundance between species and species richness. Below we take you through all the stages of the analysis.

First open the data set called **Hinkley annual fish.csv**

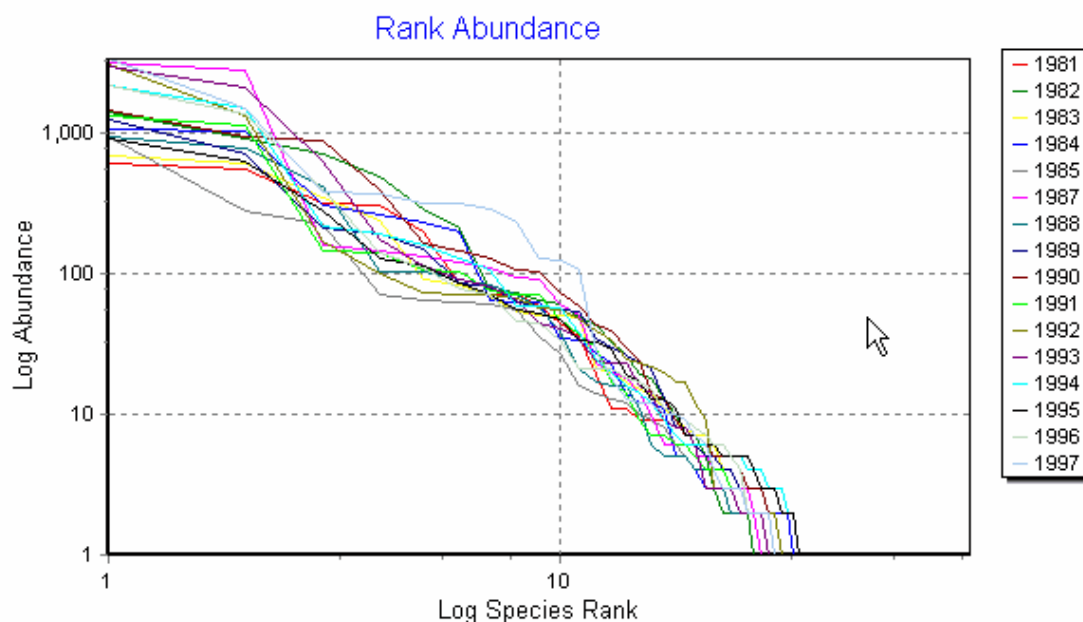
This data set is for the fish community of the Bristol Channel, England. It is derived from monthly sampling over a 25 year period and is one of the largest data sets of its type available.



Now from the Abundance tab choose Rank Abundance.

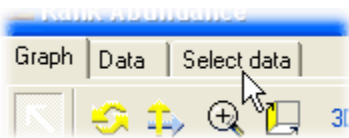


A graph showing the rank abundance plots for every year will be shown as follows:

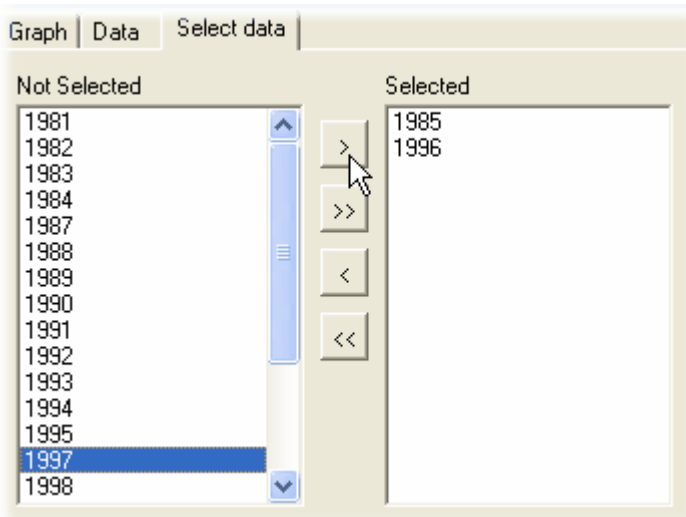


This shows a log-log plot, which is usually better than using normal axes because species typically vary in abundance over a number of orders of magnitude.

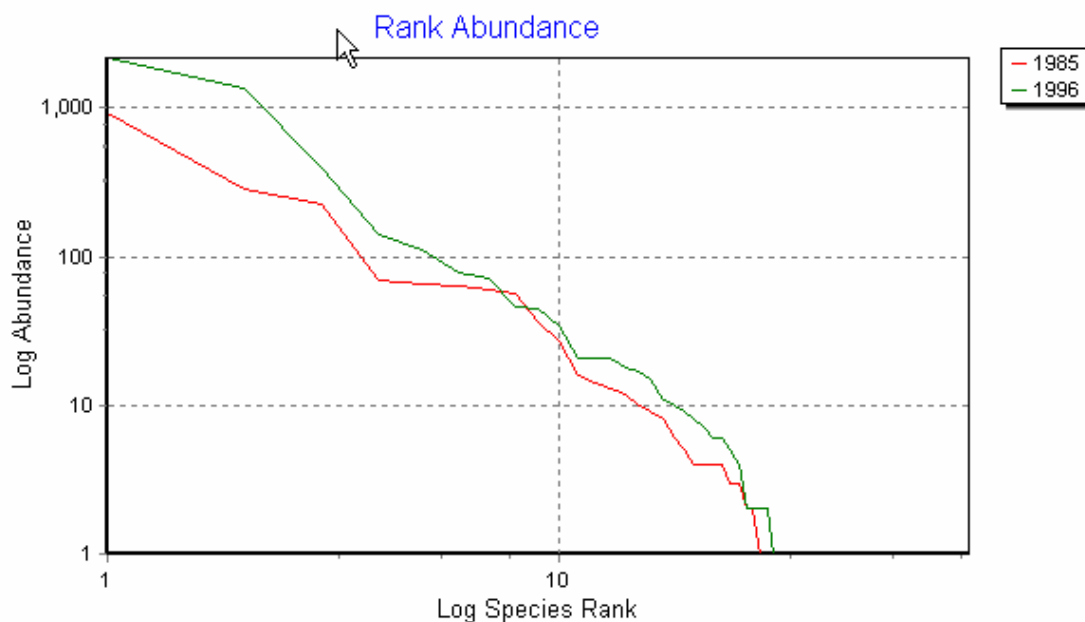
This plot is too cluttered, so you cannot see the difference between years. To select the samples (years) to plot click on the Select Data tab at the top of the graph.



Now select 1985 and 1996 only, using the selection buttons between the two boxes holding the not selected and selected samples.



Now click on the Graph tab to display the graph for 1985 and 1996. The resulting graph clearly shows the difference in fish abundance and richness between the two years.



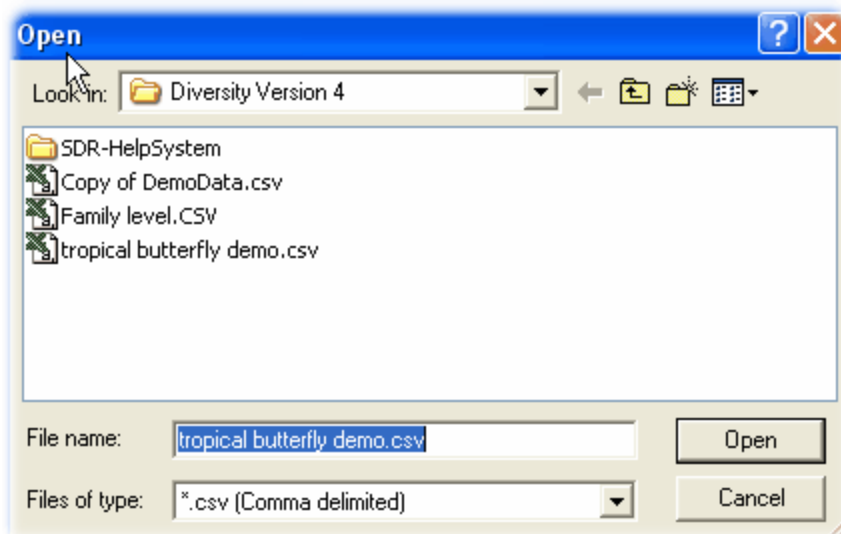
This change in fish community has been related to increases in average temperature and changes in the North Atlantic Oscillation.

5.5 Comparing the index of two samples

A common requirement is to answer the question - Is the observed difference in the diversity or equitability of two samples significant? We can test for significant differences using the [randomisation test of Solow \(1994\)](#). Below we take you through all the stages of the analysis.

First open the data set called **tropical butterfly demo.csv**

This data set is for the canopy and understory fruit-feeding butterfly communities of a South American rain forest (DeVries and Walla 2001).



Now from the Alpha Diversity tab choose the [Shannon Wiener Test](#) (We could have chosen any other index).



Now left button click on the names of the two samples canopy and understory. A "1" and "2" will appear next to the sample name.

The "1" identifies the first sample in the test results. You can change which is sample "1", as it is simply the first of the two samples you click on. The number assigned to a sample is only important if you are interested in a one-tailed test. This asks the question - Is sample 1 more diverse than sample 2?

Bootstrap		Bootstrap All	
Sample	H	Variance H	
2 Canopy	2.641	0.0003865	
1 Understory	2.983	0.0002548	

Having chosen the sample, now click on the tool bar icon labelled Comparison.



The computation will take some time as the program has to do many resamplings of the data. Finally, the results will be shown.

Delta, the difference between the two indices $2.983 - 2.641 = 0.34198$, is used by the test.

Generally you will be interested in a two-tailed test which asks the question - Is one sample index significantly different from the second sample index?

For our data, none of 10000 simulated pairs of data sets had a delta value with a magnitude greater than that observed, therefore there is no chance that the two samples have the same Shannon-Wiener diversity.

In answer to the question "is sample 1 significantly more diverse than sample 2?", the results of the one-tailed test are also highly significant. Sample 1 is shown statistically to be the more diverse.

In conclusion, using the Solow (1994) randomisation test we have demonstrated that the butterfly diversity of the tropical forest understory is significantly greater than that of the canopy.

Part

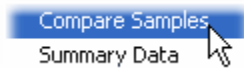


VI

6 Data exploration

6.1 Comparing samples

The taxa in two samples can be compared using the Summary menu: Compare Samples drop-down.



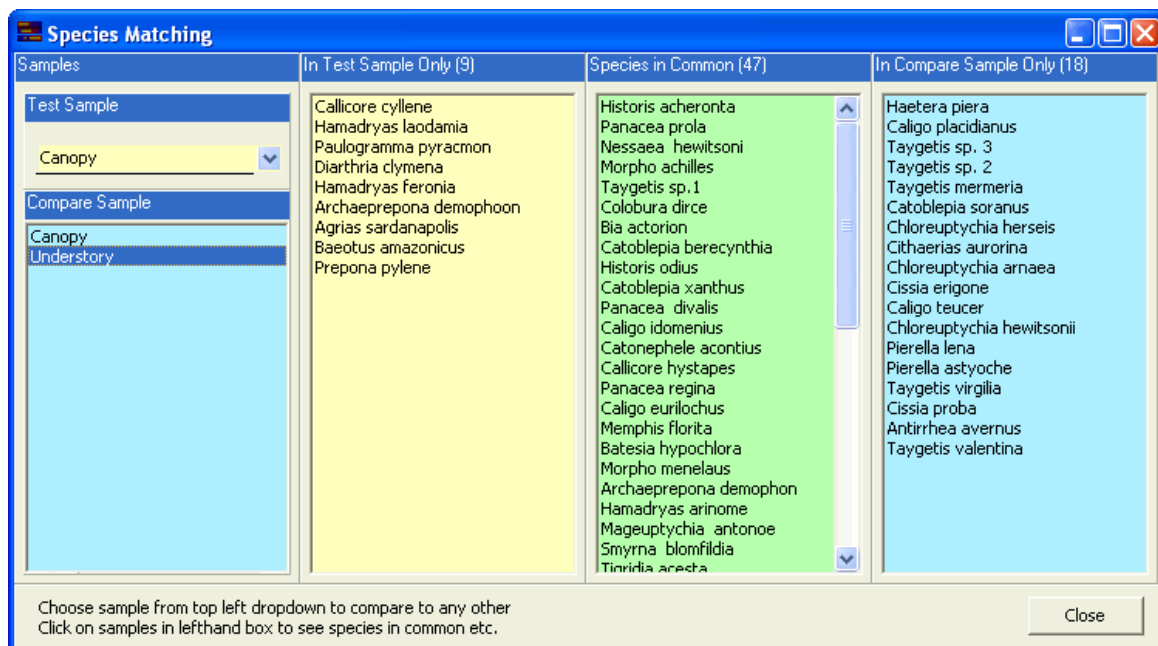
A window will open in which the two samples to compare can be chosen. You choose the two samples to compare from the drop-down box and the list at the left of the window.

Sample 1 - called the test sample is chosen from the drop-down menu under Test sample (in the example below this is labelled Canopy). Sample 2 is chosen from the list below Compare Sample by clicking on the name (in this example we have chosen Understory).

The species held in common are listed in the green central panel.

The species only in the test sample (sample 1) are in the left hand pale yellow panel. The species only in the compare sample (sample 2) are listed in the right-hand blue panel.

The numbers in brackets gives the number of species in each list.



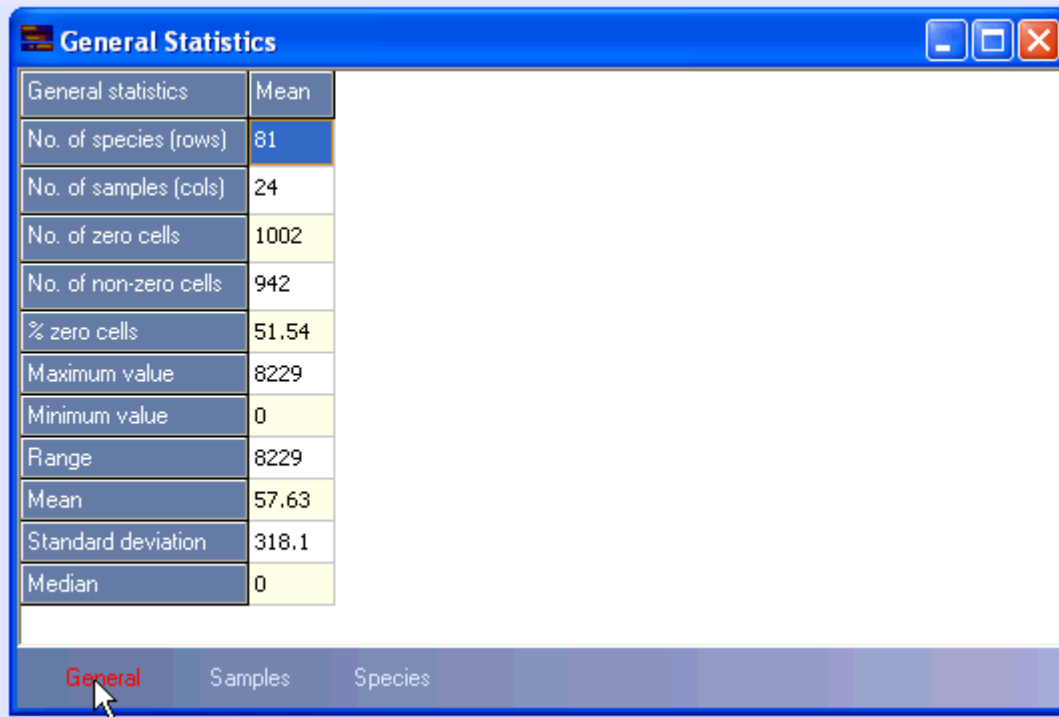
6.2 Summary Statistics

The statistics for your data set can be inspected using the Summary: Summary Data drop down.

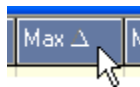
Compare Samples

Summary Data

This will open the General Statistics window, which gives statistics for the data set as a whole, for the Species, and for the Samples; you can select which set of statistics to view from the bottom bar of the General Statistics window.



These grids are sortable. To sort any column click on the topmost cell to activate the sort function. Clicking again will sort in the other direction, the third click will turn off sorting.



No. of species (rows) - This is the number of rows of data in the data set.

No. of samples (cols) - This is the number of columns in the data set.

No. of zero cells - This is the number of zero entries in the data matrix.

Non-Zeros - The number of non-zero entries

% Zeros - $(\text{Number of zeros} / \text{Total number of cells}) * 100$

Maximum value - This is the maximum value in the data matrix.

Minimum value - This is the minimum value in the data matrix.

Range - This is the difference between the maximum and minimum values.

Mean - This is the mean of all the values in the data matrix.

Standard deviation - This is the standard deviation of all the values in the data matrix.

Median - This is the median of all the values in the data matrix.

To examine the statistics for each sample click on Samples at the bottom of the window.



This will display statistics for each sample.

Sample Statistics													
Row	Mean	Median	Max	Min	Zeros	Non-zeros	% zeros	Sum	Sum Sqr	Total Variance	Sample Variance	Skewness	Kurtosis
Canopy	76.03	8.50	1882.00	0.00	18	56	24.32	5774.00	5029106.00	4578577.95	62720.25	5.91	40.28
Understory	80.03	14.00	984.00	0.00	9	65	12.16	5922.00	2809140.00	2335219.95	31989.31	3.26	13.88

The statistics for the samples can be displayed in similar fashion by clicking on Species.

Mean - This is the mean of all the values in each row or column of the data matrix.

Median - This is the median of all the values in each row or column of the data matrix.

Max - This is the maximum value in the data matrix.

Min - This is the minimum value in the data matrix.

Zeros - The number of zero entries in the data matrix

Non-Zeros - The number of non-zero entries

% Zeros - (Number of zeros/Total number of cells)*100

Sum - This is the sum of all the values in each row or column of the data matrix.

SumSqr - This is the sums of squares of all the values in each row or column of the data matrix.

Skewness - This is the skewness of all the values in each row or column of the data matrix.

Kurtosis - This is the kurtosis of all the values in each row or column of the data matrix.

For Samples only

Total Variance - This is the variance of all the values in each column of the data matrix.

Sample Variance - This is the estimate of the variance of all the values in each column of the data matrix.

For Species Only

Variance - This is the sample variance of all the values in each row of the data matrix.

Dispersion (I) - The dispersion statistic ie the variance over the mean.

Chi of I - The chi value of the dispersion statistic.

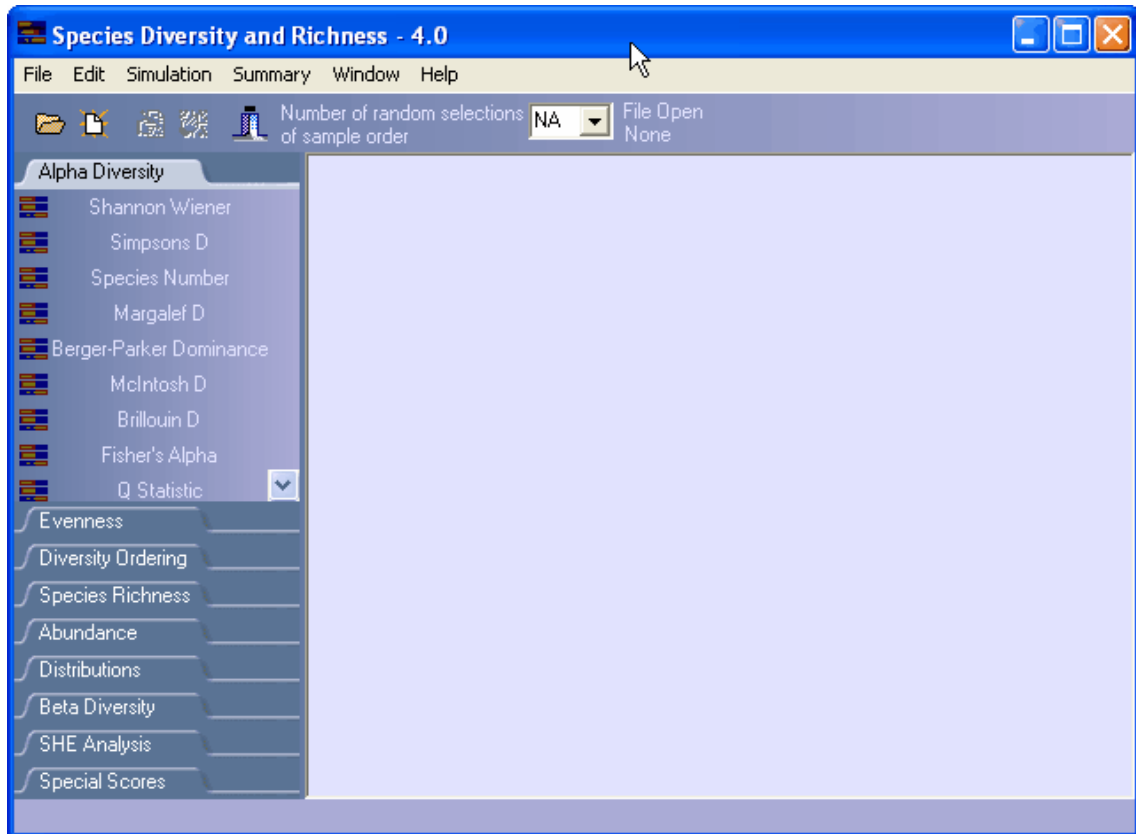
Dispersion Model - Whether the dispersion is random, aggregated or uniform.

Part

VII

7 Using Species Diversity & Richness

Start Diversity in the normal Windows fashion either by clicking on the desktop icon or from the Start menu. The program appears with no open data set, as shown below:



Along the top bar are a number of drop-down menus. These work in the same way as most standard Windows programs.

File: To open and create data sets, save files and print results.

Edit: To copy the active window.

Simulation: To create simulated data sets (see [Data Simulation](#)).

Summary: To view summary data and statistics relating to the data set currently in use, and compare species lists between samples or site.

Window: To arrange program windows.

Help: to enter the Help system. Context specific help can also be obtained for each method by opening its window and then clicking on the right hand mouse button, or by pressing the F1 key on your keyboard.

Below these pull-down menus is a series of buttons, each displaying a standard symbol:



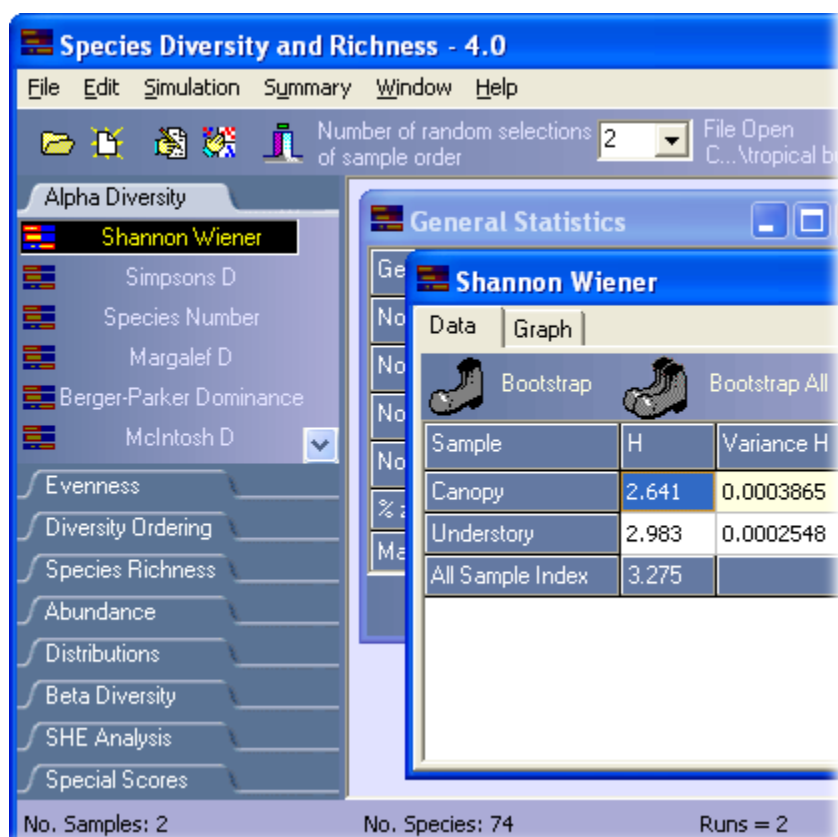
From left to right these are:

- Open an existing data file,
- Create a new data set,
- Edit the open data set (greyed out when no data set is open)

- Select which samples to analyse from the data set (greyed out when no data set is open),
- Close the program.

To the right of these buttons is a pull-down menu that allows the selection of the number of randomisations of the data that will be run during species richness estimation. See [Number of Random Selections](#) for more details.

Down the left hand side of the window are a series of tabbed panels relating to the different types of analysis available. When any of them is selected a list of possible methods will appear. If you hover the cursor over a method, it will be highlighted; to apply the method to the current data set, simply click on it. A new window will appear with the results, and the method's logo will change (image below):



SDR is designed to enable you to see the results of a number of calculations simultaneously. Each of the methods available in the program will be described in detail throughout this help system.

7.1 Entering data

Demonstration data sets

A number of [example data sets](#) are supplied with the program. These allow the user to test the program, and can be opened in a spreadsheet or word-processor program to examine the way the data are organised. The ***hinkley_fish.csv*** file holds the actual monthly captures of fish collected at Hinkley Point in the Severn Estuary over a 24-year period.

Creating data sets

Data sets can be created within *Species Diversity and Richness*, or by using a spreadsheet such as Microsoft Excel, or many other spreadsheets, word processors, database programs etc. We recommend that you organise large data sets using a spreadsheet like Excel as this will give access to a wide range of sorting and editing procedures to ease your task.

[Creating a data file using a spreadsheet](#)

[Creating a new data set within SDR](#)

7.2 Maximum size of the data set

SDR IV has been written to handle any size of data set. However, there will still be a maximum upper limit to the size of data set it can handle that will be defined by the processing power and available memory of your computer.

The program has been tested with up to 3500 species (rows) by 2500 samples (columns). It is not practicable to test rigorously larger data sets.

On a reasonably fast computer, most analyses will be performed in a few seconds; if not instantaneously. However, some analyses of data sets approaching the maximum permitted size can take a long time to run. The more time-consuming calculations include some of the species richness estimations, randomisation tests for a significant difference in indices between samples, and the creation of rank abundance graphs.

7.3 File open button

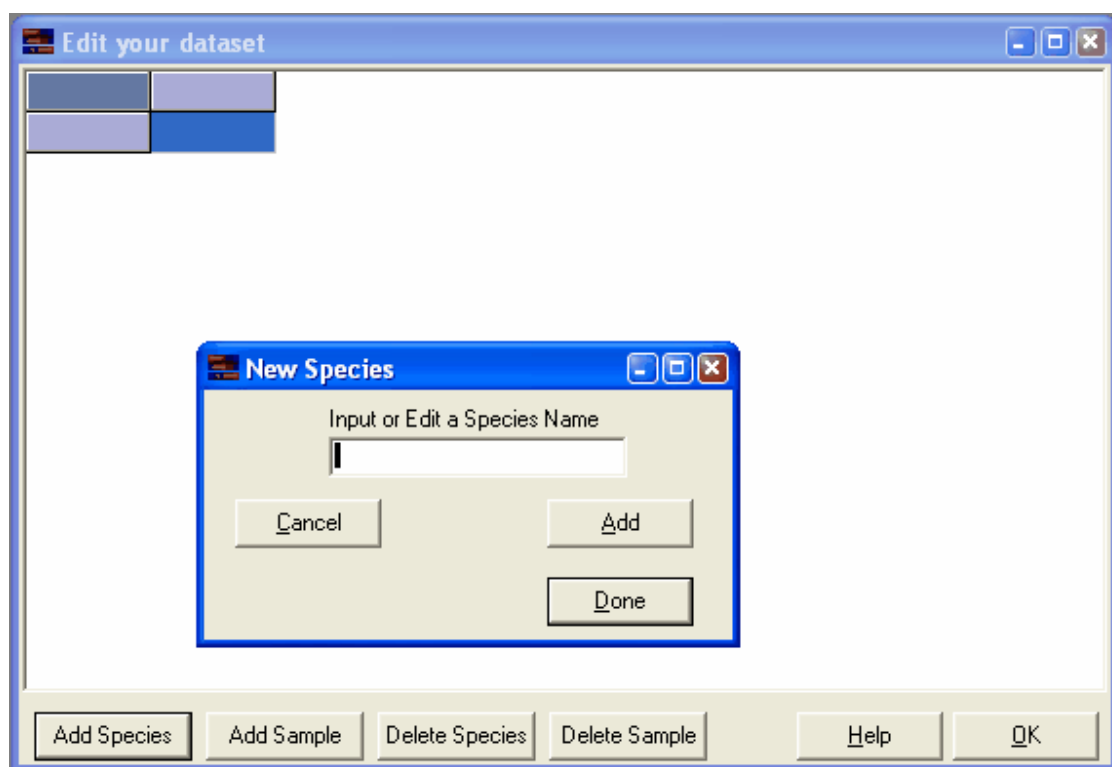
This button activates a standard Windows file dialogue with which you can select a file to open.



[Open an existing File](#)

7.4 Creating a new data set within SDR

Data sets can also be created and edited within *Species Diversity and Richness*. To create a new data set, select New from the File pull-down menu or click on the second button on the tool bar. You will be presented with the *Edit your dataset* box:



Add new species and new sample details using the buttons at the bottom of the screen, and input your data. To edit a species or sample name double-click in the first (title) cell of the row or column – a box appears in which you can alter the title or delete the row/column. To delete a row or column, click in a cell in it, and use the Delete Species / Delete Sample button.

NOTE: Commas are not allowed in the species name. For instance, Upright Hedge-parsley should not be entered as Hedge-parsley, Upright.

When you have finished entering data, click OK, and you will be prompted to choose a name for your new data file and save it.

By default, all samples are selected for inclusion in the analyses. If you wish to remove a sample from the analysis, but not delete it completely from the data set, use the Select Data function (4th button from left on the main program toolbar) – see [Selecting the Samples to Analyse](#).

Remember that for larger data sets you are advised to use a spreadsheet to create your dataset, as it will give you superior editing and data search facilities. See [Creating a data file using a spreadsheet](#).

7.5 Creating a data file using a spread sheet

A data set is stored as a comma-delimited text file with the extension .csv. It has the following form:

	Title1	Title2	Title3
Species1	21	1	5
Species2	15	5	0
Species3	0	7	0
Species4	1	9	0

Species5 0 0 8

We recommend that you organise large data sets using a spreadsheet program like Excel as this will give access to a wide range of sorting and editing procedures to ease your task. The normal arrangement of data is to have the samples (quadrats) as columns and the species as the rows. However, Excel has a maximum number of columns of 255, which can prove difficult if you have a data set with a very large number of sites/samples. If this is the case, it may be necessary to split the data set in to two or more subsets.

	A	B	C	D	E	F
1		Natres1	Natres2	Lakeout	Tygup	SDWdm
2	Polycelis felina (?)	0	0	1	0	0
3	Lymnaea palustris	0	0	0	0	0
4	Potamopyrgus jenkinsi	0	0	0	7	12
5	Valvata piscinalis	0	0	0	0	0
6	Succinea putris	0	0	0	0	1
7	Pisidium sp.	0	1	0	0	0
8	Tubificidae	4	2	3	7	9
9	Naididae	0	0	0	0	0
10	Lumbriculidae	1	0	10	2	4
11	Lumbriculidae	1	5	0	0	0
12	Eiseniella tetraedra	1	1	0	0	0
13	Gammarus pulex	1	0	0	32	0
14	Crangonyx pseudogracilis	0	0	6	0	0
15	Baetis rhodani	1	1	0	4	0

The table and image above show you how the data will look in Excel. The samples are arranged in columns. Each sample has a title field. Start the first sample in column 2. The data consists of the number of individuals observed in the sample. Put in zeros rather than leaving cells blank. The species names are entered from row 2 in column 1.

NOTE: Commas are not allowed in the species name. For instance, Upright Hedge-parsley should not be entered as Hedge-parsley, Upright.

When using Excel use the Save As function to save your data as a *.csv file. *Ensure that the work sheet you are saving only holds the tabulated data for analysis.* If your data set has been created using the convention that a blank cell means zero then use the find and replace function available in all common spreadsheets to search for blanks and replace them with 0 (zero).

Occasionally, errors occur because a blank space has been accidentally entered into a cell outside the data matrix. To prevent this happening, it is good practice, before saving your data set as a .csv file, to highlight the first 10 or so blank rows and columns below and to the right of the data matrix, and press 'Delete'. This will clear the cells of any accidentally-entered contents.

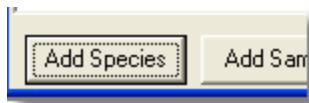
The data set above looks like this if viewed in a word processor program (such as Word, Word Perfect) or text viewer such as Notepad:

```
,Title1,Title2,Title3  
Species1,21,1,5  
Species2,15,5,0  
Species3,0,7,0  
Species4,1,9,0  
Species5,0,0,8
```

Note the leading comma on the first row, which will make the first cell blank.

7.6 Add species

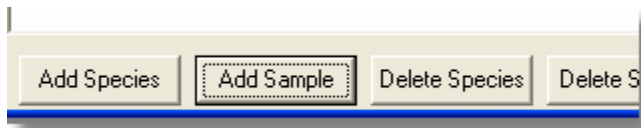
Add a new species to a data set using this button. The program will search the data set to make sure that it is unique.



NOTE: Commas are not allowed in the species name. For instance, Upright Hedge-parsley should not be entered as Hedge-parsley, Upright.

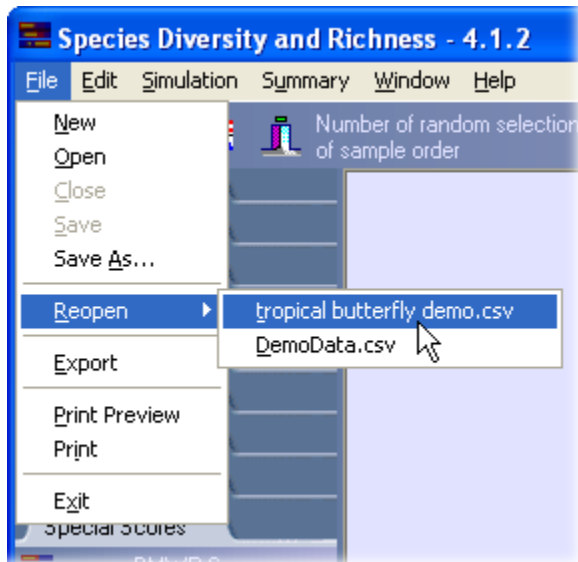
7.7 Add sample

Click this button to enter a name or title for the new sample. When finished, click OK, and the sample column will be added at the end of the data set.



7.8 Open an existing data file

Use either File: Open or the left hand button on the tool bar to select a data file for analysis. A list of previously-used files is available under Reopen on the File menu:

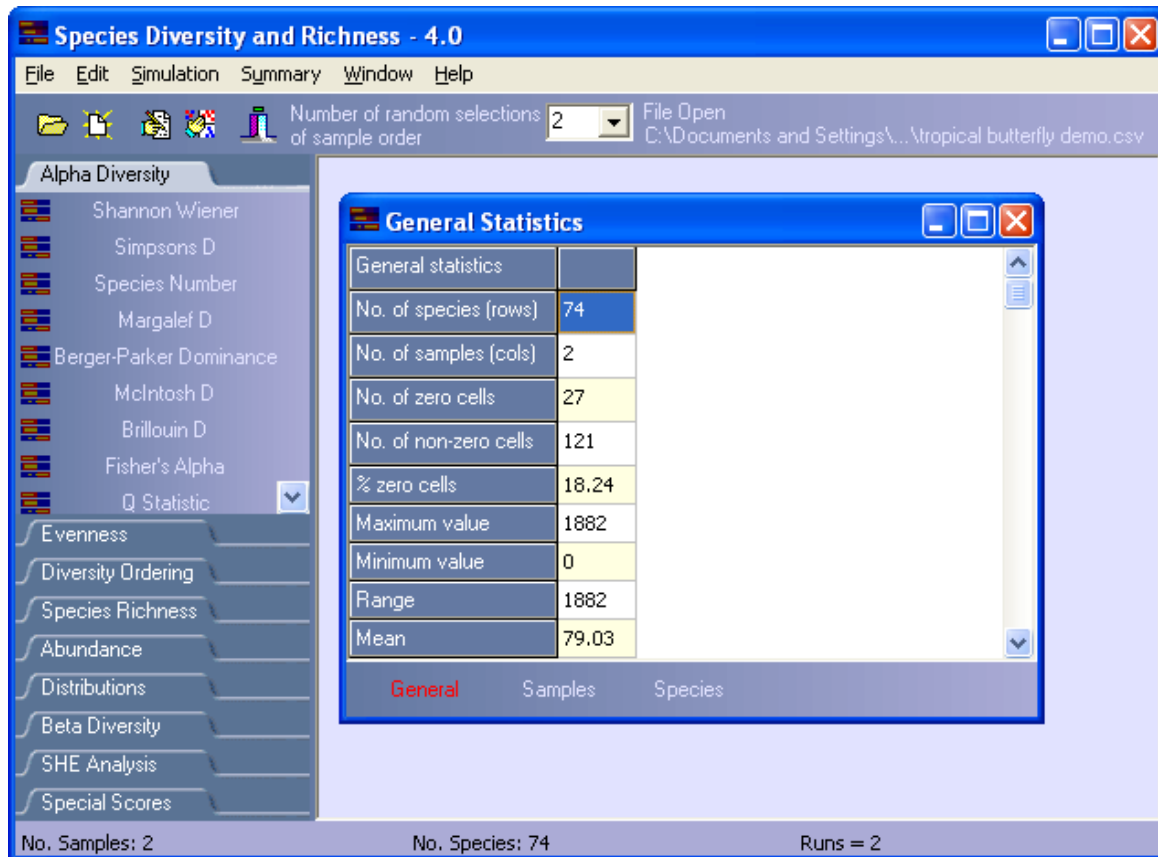


Species Diversity & Richness 4 will open either comma-separated text (*.csv) or Microsoft Excel spreadsheet files (*.xls). With Excel files, it is important that a number of points are observed:

1. If you are using an Excel file with multiple worksheets, the data will be imported from the worksheet that was open (visible) when the file was last saved.
2. The data should be present as a contiguous rectangular block, starting at Cell A1.
3. Cell A1 itself should be empty, with sample names present in Row 1 and variable (species) names present in Column 1.
4. The import procedure ignores formulae in cells and imports the visible values.
5. Ensure that all cells rightwards and downwards from cell B2 contain numerical data.
6. CAP can import directly from Excel whether Excel is open or not, provided the worksheet has been saved.

SDR requires all numbers in the data set to be either binary presence-absence data (i.e. 0 or 1), or positive integers; in opening a data set, the program will check the data and alert you to any non-standard numbers which would cause the program to fail. You can then edit the data set to correct these data points, either in the original spreadsheet program, if you used one, or using the [Edit Open Dataset button](#).

Once the file is opened, its name is displayed in the tool bar towards the top left of the main window, and the number of species it contains is given in the information bar at the base of the window. The General Statistics window will be displayed, showing all the available statistics on the data set.

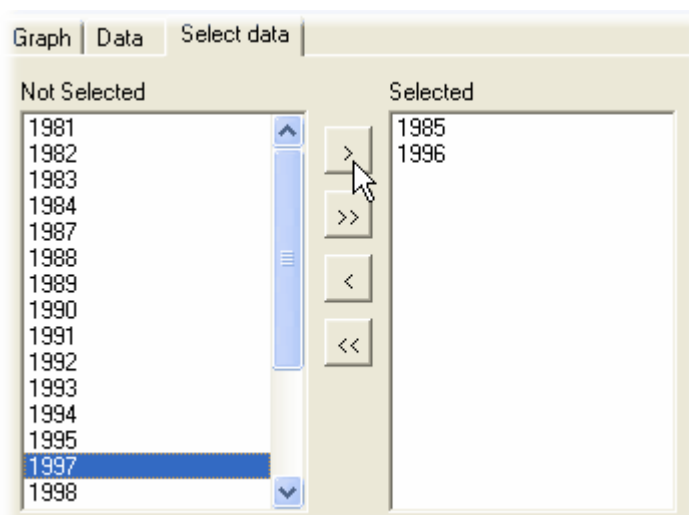


7.9 Selecting the Samples to Analyse

Not all the samples in the data set need be analysed. To select or deselect samples for analysis use the select data button (the 4th from left button on the tool bar).



This will open the select samples to analyse window. Use the buttons between the two boxes to deselect and select samples.



Once you have made your selection, the bottom bar will show the number of samples that will be used for the calculations.



7.10 Non-Integer Data

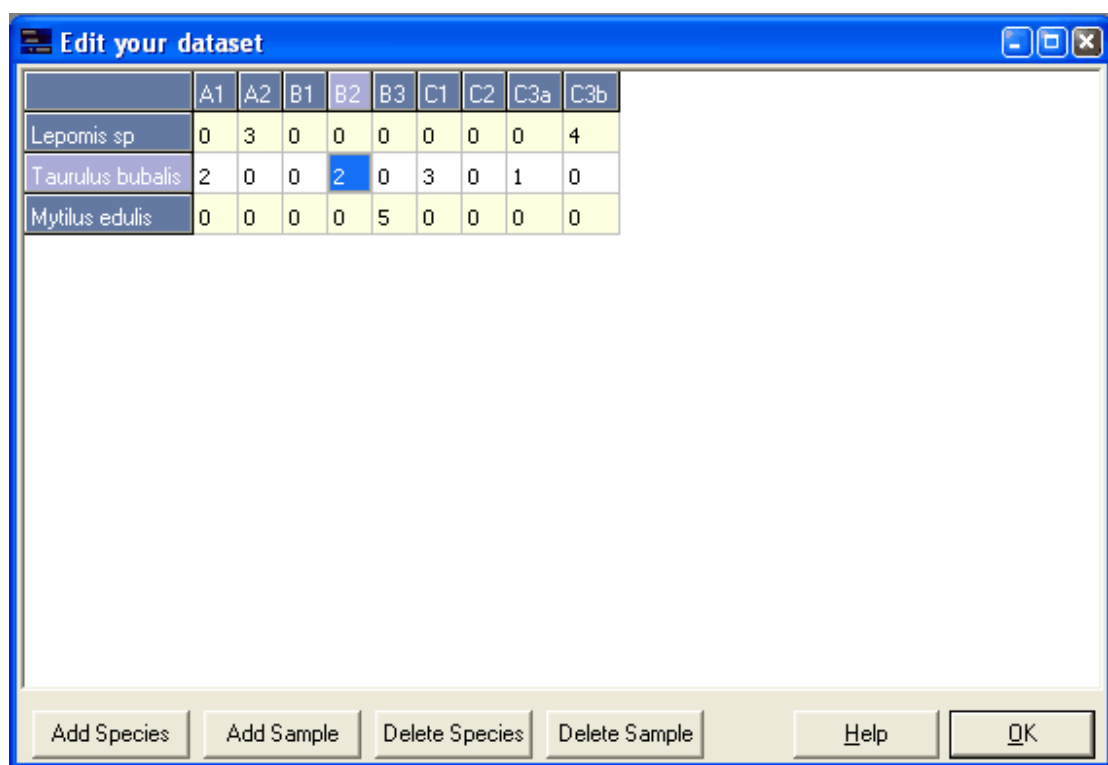
The program has been designed to analyse counts of animal and plant abundance and therefore SDR requires all numbers in the data set to be either binary presence-absence data (i.e. 0 or 1), or positive integers; in opening a data set, the program will check the data and alert you to any non-standard numbers which would cause the program to fail. You can then edit the data set to correct these data points, either in the original spreadsheet program, if you used one, or using the [Edit Open Dataset button](#).

There are cases where data are only available as individuals per unit area or volume, where the densities are expressed as real numbers. Diversity indices and ordering can be undertaken on these data. Simply convert the data to integers by multiplication and rounding in the spreadsheet while preparing the data set.

The richness estimators are not designed to be used on density data. For example, some methods need the number of species represented by one or two individuals in a sample. This cannot be obtained from numbers per unit area or similar measures. However, if you have a number of samples you can convert your data to presence/absence data and use the [Chao presence-absence index](#). This index uses the number of species recorded in 1, 2 or more samples to estimate total species richness.

7.11 Editing data

To edit an open data set click the Edit Open Dataset button (3rd from left in main program toolbar). This will open the *Edit Dataset* window (shown below). You can then change individual observations, add/delete species and samples, and select which samples are included in the analysis.



NOTE: Commas are not allowed in the species name. For instance, Upright Hedge-parsley should not be entered as Hedge-parsley, Upright.

To edit a particular observation, double-click in the cell you wish to change (the row and column headers change colour to indicate which row/column the cell is in), type in the new value, and either press the Enter key on your keyboard, or click into another cell with the cursor.

Add new species and new sample details using the buttons at the bottom of the screen, and input your data. To edit a species or sample name double-click in the first (title) cell of the row or column – a box appears in which you can alter the title or delete the row/column. To delete a row or column, click in a cell in it, and use the Delete Species / Delete Sample button.

An error message that can occur while using the program is 'division by zero'. This happens when one of your samples or species holds insufficient observations to undertake the procedure selected. You can either de-select the sample, add observations, or delete the sample or species from the data set.

7.12 Printing and exporting your results

The graphs and other results can be printed using the Print option from the File menu. The Print Preview option will show the page layout and allow image size, margins, and paper orientation to be changed. If you have Adobe Acrobat (note, not the free Adobe Acrobat Reader software) installed on your computer, you will be able to convert the chart directly to a .pdf file by selecting Acrobat Distiller or Acrobat PDF Writer from the list of available printers in the Print dialog box.

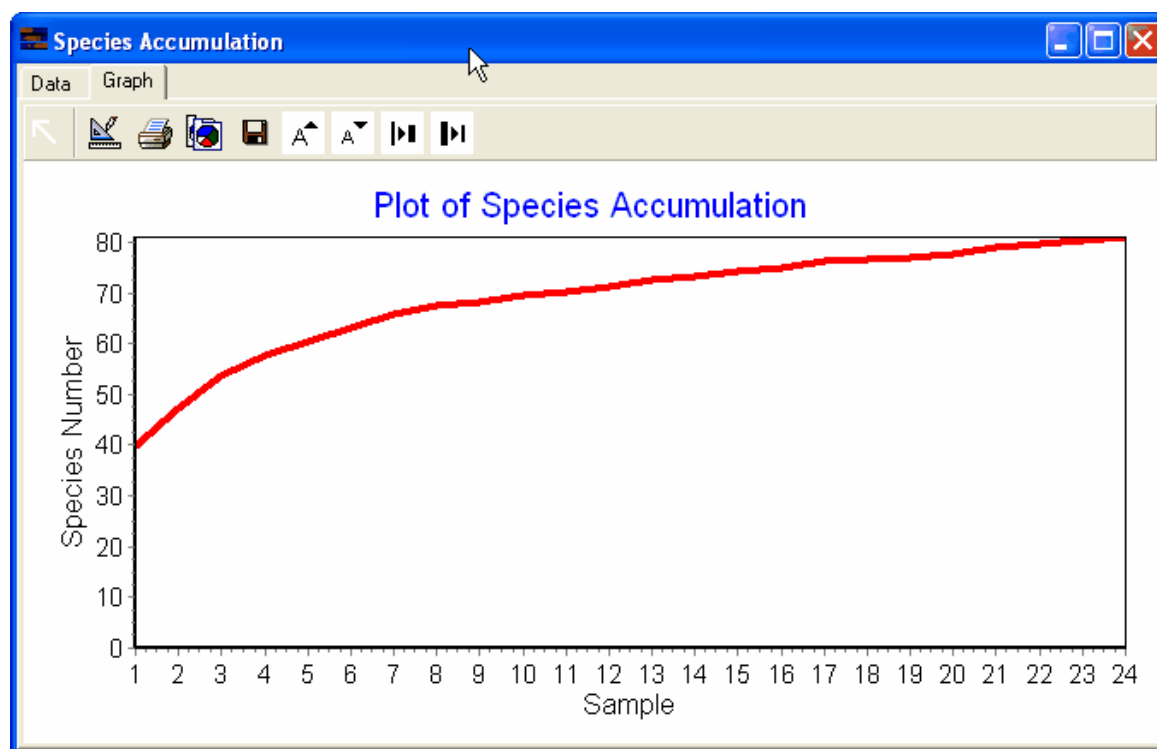
To copy the results to other applications, such as a spreadsheet or word processing program, use the Copy command from the Edit menu. Then in the other program,

choose Paste from the Edit menu. To copy, save or export graphs click on the Edit button above the graph, choose General and click on the Export button.

File formats

The chart can be saved in a number of different file formats: Enhanced Metafile (*.emf), Bitmap (*.bmp), JPEG (*.jpg), or Native (*.tee). Each file format has advantages and disadvantages.

- The advantage of Enhanced Metafile is that, if pasted into, for instance, a Word document, it can be resized by dragging, without losing resolution.
- Bitmaps are a lossless method of saving; the stored file will not lose any of the original's detail. Because of this, bitmaps tend to be much larger than compressed files such as Enhanced Metafiles or JPEGs.
- JPEGs are file formats which can be compressed to take up less space - useful if you wish to send one by email, put it on a website, or paste it in to a document. If they are compressed too heavily, they can lose resolution and detail, and spoil colours.
- The Native (*.tee) format saves all the chart attributes, and the data series, rather than the image itself. This means that you can save the chart, and open it again at a later date to edit it, using the free TeeReader software supplied on the installation CD. Tee files tend to be very small indeed; often less than 1KB.



The graph option buttons are described in order from left to right below.


Normal - (Arrow) Use this button to zoom in onto an area of your graph. Move the pointer to a point above and to the left of the area you want to enlarge, hold the left-hand mouse button down and drag down to the right and release the button. To reverse the enlargement, click on the left hand mouse button and move up and to the left.


Edit - (Set square and pencil) This button will offer a wide range of options to change the style of your graph. It is also used to export or copy your graph to file.

Print - Use this button to print the graph


Copy - Use this option to copy the graph to the clipboard.

Save - To save your graph. You can choose from several file types using the drop down list in the Save Dialog.

 - Increase the font size of the title and axis.

 - Decrease the font size of the title and axis.

 - Increase the plot line size.

 - Decrease the plot line size.

Many output windows offer a graphical display of the data that is accessed via a button at the top left of the window. These graphs can be copied to the clipboard for inclusion into other programs as bitmaps, metafiles or enhanced metafiles (see above). Metafiles are particularly useful if you wish to include a graph in a document created using a word processor such as Word, as they can be resized without loss of resolution.

Some output windows have particular output options that are described under the individual methods below.

7.13 Saving a data set

To save the current data set, use Save or Save As from the *File* menu.

7.14 Closing the program

Choosing the Exit button (last on the right on the lower toolbar), Exit from the File menu, or the cross symbol at the top right corner of the program window will close Species Diversity & Richness.

7.15 Maximum size of a variable

The maximum size of any number in *Species Diversity and Richness 4* is 2147483647. We think it unlikely that anyone can count more than that many individuals in a single lifetime.

7.16 Bottom bar

The bottom bar of the main screen gives:

The number of samples in the open data set,

The number of species in the open data set,

The number of randomisations of sample order for smoothing of species accumulation curves and species richness estimation is given as Runs = xxx

If you open a dataset which contains inappropriate data, a flashing red/white warning will also appear in the bottom toolbar.

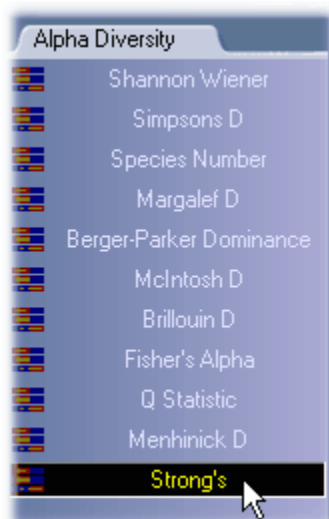
Part



8 Alpha diversity

The Alpha Diversity tab at the left of the program window displays a list of methods which can be run by clicking on their icon.

Comparative reviews of relative merits of the indices described below are provided by Taylor, (1978); Kempton (1979) and Magurran (2004). The program offers a choice of 11 measures of alpha diversity.



The index for each individual sample is tabulated. At the bottom of the table the index and jackknifed standard error for all the samples combined is also shown.

2001	2.013	0
2002	2.047	0
2003	2.232	0
2004	2.131	0
2005	1.812	0
All Sample Index	1.984	
Jackknife Std Error	0.09724	

Descriptions of these indices are available by following the jumps below.

[Shannon-Wiener](#)

[Simpson's D](#)

[Species Number](#)

[Margalef D](#)

[Berger-Parker Dominance](#)

[McIntosh D](#)

[Brillouin D](#)

[Fisher's alpha](#)

[Q statistic](#)

[Menhinick D](#)

[Strong's](#)

The output from all of the diversity estimators is presented in similar fashion. In the

first column is the name of the sample, and in the second the calculated value for the index. Remember that if you have chosen to subset the data then only the samples selected will be represented in the output. If the [Bootstrap](#) function is used, then the results will be displayed in the third and fourth columns. The change in diversity between samples can also be shown graphically, simply click on the Graph tab to display the chart.

To export or print the data see '[Printing and exporting your results](#)'.

To print the indices, click in the window to activate it, then choose File: Print on the top menu bar, or right-click in the results window and select Print from the pop-up menu.

8.1 Shannon-Wiener Index

Choose this diversity calculation from the Alpha Diversity tab.



This presents the Shannon-Wiener (also incorrectly known as Shannon-Weaver) diversity index for each sample. A simple plot of the way the index changes between samples is displayed by clicking on the Graph tab on the output window.

The function was originally devised to determine the amount of information in a code or signal, and is defined as:

$$H = - \sum_{i=1}^S p_i \log_e p_i$$

where p_i = the proportion of individuals in the i th species. Species Diversity & Richness calculates the index using the natural logarithm.

In terms of species abundance:

$$H = \log_e N - \frac{1}{N} \sum_{i=1}^w (p_i \log_e p_i) n_i$$

where n_i = the number of species with i individuals. The information measure is nits for base e and bits per individual for base 2 logarithms.

This ever-popular index is really not as good as its popularity would imply. However, it appears pleasantly sophisticated! To quote [Southwood and Henderson \(2000\)](#) "*..it is an insensitive measure of the character of the S:N relationship and is dominated by the abundant species*". The value of the Shannon-Wiener Index usually lies between 1.5 and 3.5 for ecological data and rarely exceeds 4.0. [May \(1975\)](#) notes that if species follow a log normal abundance distribution the sample must hold 100,000 species for H to be greater than 5.0. By itself, this index often gives little more insight

than [Species number](#).

To compare two indices see [Testing for significant differences between indices](#).

8.2 Simpson's Index

Simpson's index for each selected sample is displayed with this option. A simple plot of the way the index changes between samples is displayed by clicking on the Graph tab on the output window.

A diversity index proposed by Simpson (1949), to describe the probability that a second individual drawn from a population should be of the same species as the first. A similar type of index had a few years earlier been proposed by G. Yule to compare an author's characteristic vocabulary (frequency of different words in his writings). The statistic, C (or Y) is given by:

$$C = \sum_i^{s_{obs}} p_i^2$$

where, s_{obs} is the number of observations and, strictly,

$$p_i^2 = \frac{N_i(N_i - 1)}{N_T(N_T - 1)},$$

but is usually approximated as:

$$p_i^2 = \left(\frac{N_i}{N_T} \right)^2$$

where N_i is the number of individuals in the i th species and N_T the total individuals in the sample. The index is:

$$D = \frac{1}{C}$$

and the larger its value the greater the diversity.

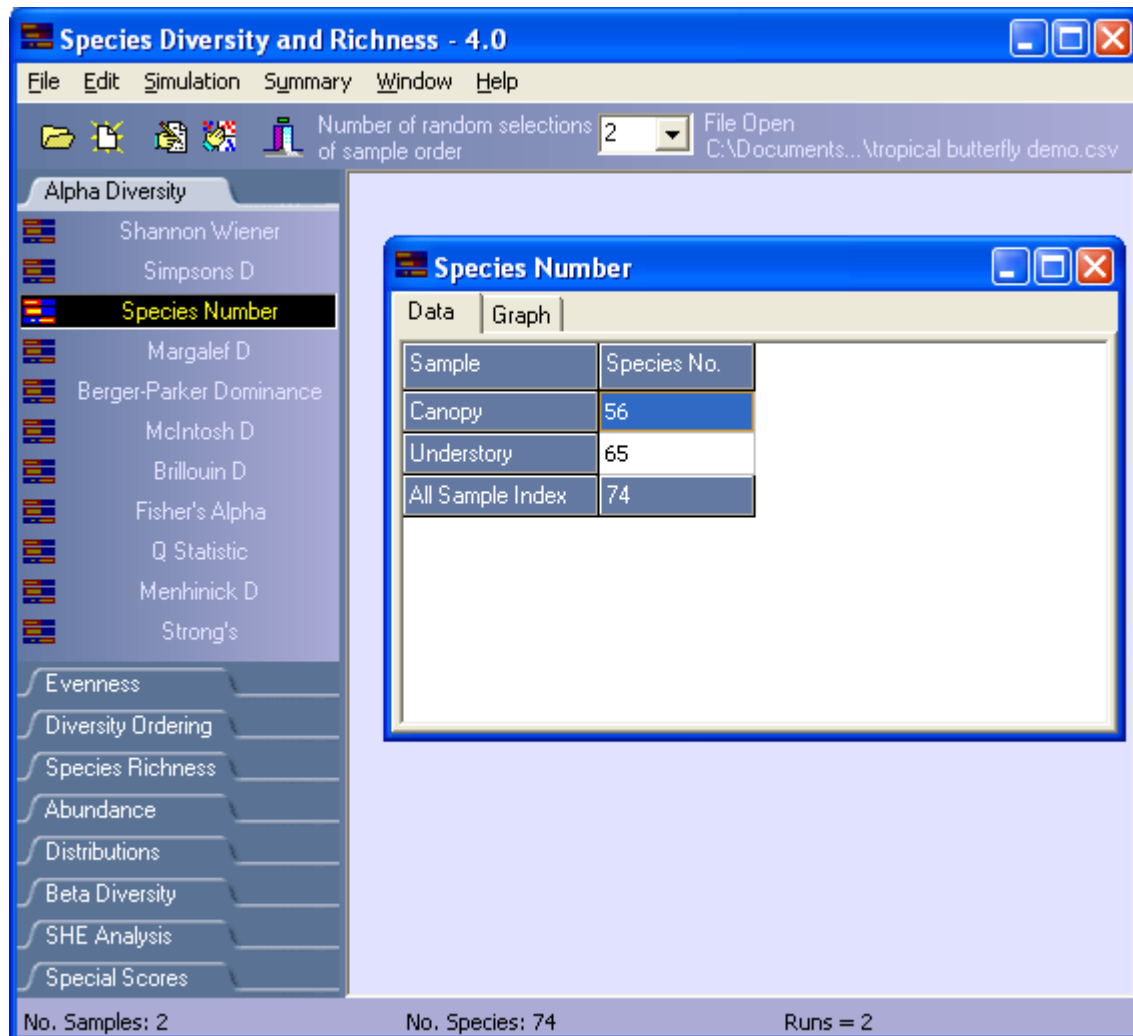
The statistic $1 - C$ gives a measure of the probability of the next encounter (by the collector or any animal moving at random) being with another species ([Hurlbert 1971](#)). [May \(1975\)](#) showed that this index is strongly influenced for values of $s_{obs} > 10$ by the underlying distribution. As [Magurran \(2004\)](#) states "Simpson's Index is heavily weighted towards the most abundant species in the sample, while being less sensitive to species richness."

However [Magurran \(2004\)](#) also states "The Simpson index is one of the most meaningful and robust diversity measures available. In essence it captures the variance of the species abundance distribution."

To compare two indices see [Testing for significant differences between indices](#).

8.3 Species Number

This window gives the total species number for each sample.



8.4 Margalef D

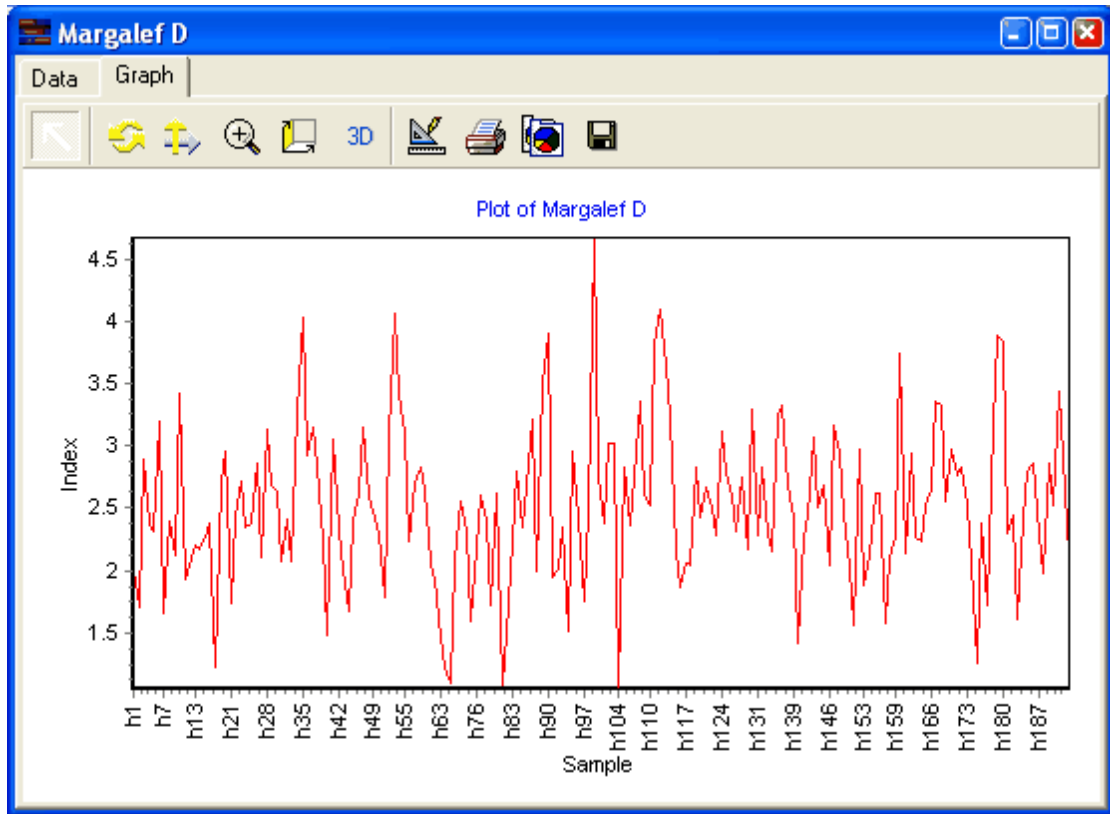
This window displays the index calculated separately for each chosen sample.

Margalef's D has been a favourite index for many years. It is calculated as the species number (S) minus 1 divided by the logarithm of the total number of individuals (N). This program uses the natural logarithm:

$$D = \frac{(S - 1)}{\ln N}$$

To compare two indices see [Testing for significant differences between indices.](#)

A simple plot of the way the index changes between samples is displayed by clicking on the Graph tab on the output window.



8.5 Berger-Parker Dominance

This window displays the index calculated separately for each chosen sample.

This surprisingly simple index was considered by [May \(1975\)](#) to be one of the best.

It is simple measure of the numerical importance of the most abundant species.

$$d = N_{\max} / N$$

where N_{\max} is the number of individuals in the most abundant species, and N is the total number of individuals in the sample.

The reciprocal of the index, $1/d$, is often used, so that an increase in the value of the index accompanies an increase in diversity and a reduction in dominance. We plot the dominance index d .

A simple plot of the way the index changes between samples is displayed by clicking on the Graph tab on the output window.

To compare two indices see [Testing for significant differences between indices.](#)

8.6 McIntosh D

Proposed by McIntosh (1967) as:

$$D = \frac{N - U}{N - \sqrt{N}}$$

where N is the total number of individuals in the sample and U is given by the expression :

$$U = \sqrt{\sum n_i^2}$$

where $n_{(i)}$ is the number of individuals in the i th species and the summation is undertaken over all the species. U is the Euclidean distance of the community from the origin when plotted in an S-dimensional hypervolume (we are not joking!).

This index has been little used.

A simple plot of the way the index changes between samples is displayed by clicking on the graph tab on the output window.

To compare two indices see [Testing for significant differences between indices.](#)

8.7 Brillouin D

The Brillouin index, HB, is calculated using:

$$HB = \frac{\ln N! - \sum_{i=1}^s \ln n_i!}{N}$$

where N is the total number of individuals in the sample, n_i is the number of individuals belonging to the i th species and s the species number.

The Brillouin index measures the diversity of a collection, as opposed to the Shannon index which measures a sample.

Pielou (1975) recommends this index in all situations where a collection is made, sampling was non-random or the full composition of the community is known. The value obtained rarely exceeds 4.5 and both the Brillouin and Shannon Indices tend to give similar comparative measures.

This information measure should be used in favour of the [Shannon index](#) when the species differ in their capture rates.

To compare two indices see [Testing for significant differences between indices.](#)

8.8 Fisher's alpha

This is a parametric index of diversity that assumes that the abundance of species follows the log series distribution:

$$\alpha x, \frac{\alpha x^2}{2}, \frac{\alpha x^3}{3}, \dots, \frac{\alpha x^n}{n}$$

where each term gives the number of species predicted to have 1,2,3,...n individuals in the sample. The index is the alpha parameter.

This is a useful index, which has been widely used. To test if a log series distribution is appropriate see [Fitting abundance distributions - Log series](#). It is estimated by an iterative procedure that may take an appreciable amount of time with large data sets.

A number of authors argue strongly in favour of this index (eg [Kempton, and Taylor, 1976](#))..

Hayek & Buzas (1997) believe it to be a useful index provided the ratio of the total number of individuals to the species number (N/S) exceeds 1.44. In many situations, alpha is approximately equal to the number of species represented by a single individual. The value of alpha is independent of sample size when the number of individuals in the sample exceeds 1000.

A simple plot of the way the index changes between samples is displayed by clicking on the Graph tab on the output window.

To compare two indices see [Testing for significant differences between indices](#).

8.9 Q statistic

This infrequently-used diversity measure was proposed by Kempton and Taylor (1976). It measures the interquartile slope of the cumulative abundance curve and is estimated by :

$$Q = \frac{\frac{1}{2}n_{R1} + \sum_{R1+1}^{R2-1} n_r + \frac{1}{2}n_{R2}}{\log(R2/R1)}$$

where

- nr = the total number of species with abundance R;
- S = the total number of species in the sample;
- R1 and R2 the 25% and 75% quartiles of the cumulative species curve;
- nR1 = the number of individuals in the class where R1 falls;
- nR2 = the number of individuals in the class where R2 falls.

This index has been little used, probably because of the difficulties of computation. It seems worth considering; see [Magurran \(2000\)](#).

To compare two indices see [Testing for significant differences between indices](#).

A simple plot of the way the index changes between samples is displayed by clicking

on the Graph tab on the output window.

8.10 Menhinick Index

Menhinick's index, D_{mn} (Whittaker, 1977), is calculated using:

$$D_{mn} = \frac{S}{\sqrt{N}}$$

where N is the total number of individuals in the sample and S the species number.

To compare two indices see [Testing for significant differences between indices.](#)

8.11 Strong's Index

Strong's dominance index, D_w ([Strong, 2002](#)), is calculated using:

$$D_w = \max_i \left[\left(\frac{b_i}{Q} \right) - \frac{i}{R} \right]$$

where

b_i is the sequential cumulative totaling of the i th species abundance values ranked from largest to smallest;

Q is the total number of individuals in the sample;

R is the number of species in the sample

and

\max_i is the largest calculated i th values

The meaning of the equation can be seen by studying the following figure reproduced from Strong (2002). The upper curve represents the [species accumulation curve](#) for the sample and the straight line a model in which each species has the same abundance.

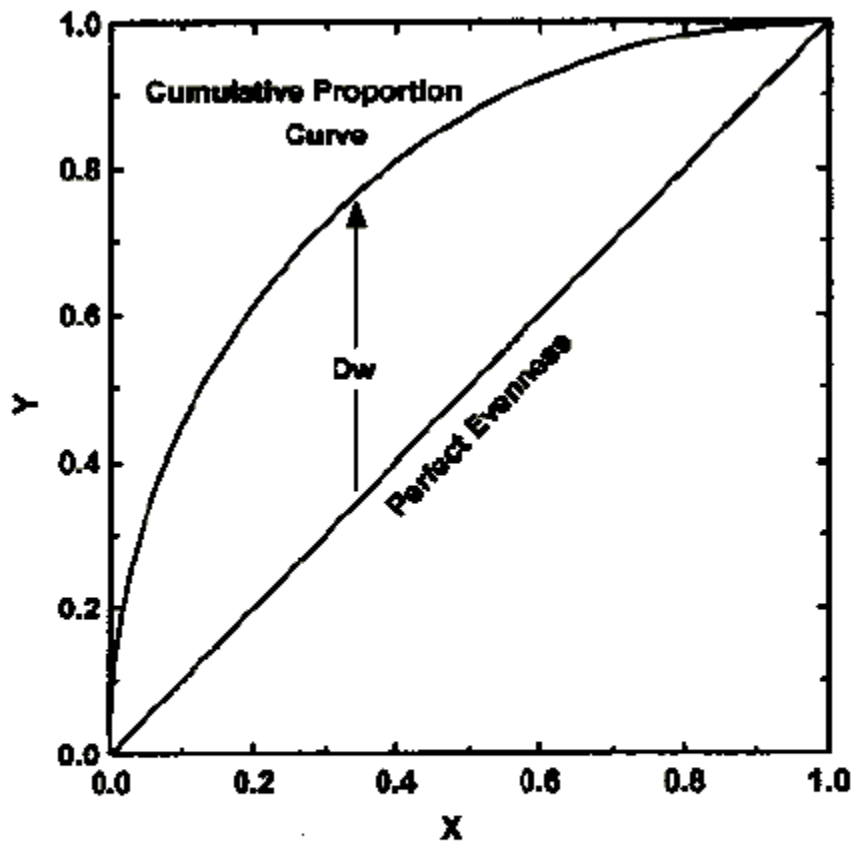


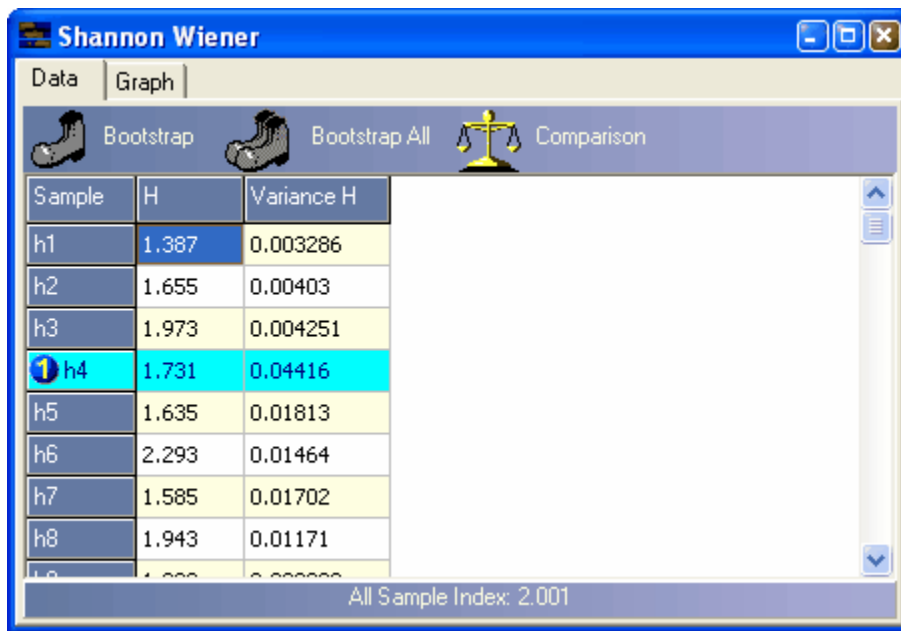
Figure 1. A cumulative proportion graph with an example of a dominance (D_W) measure. The x-axis (i/R) and y-axis (b/Q) of the diagram represent the right and left halves of the D_W equation. See Methods for definitions of individual parameters.

To compare two indices see [Testing for significant differences between indices.](#)

8.12 Bootstrap estimation of confidence intervals

After a diversity index method has been run, a standard bootstrap method for estimating upper and lower 95% confidence intervals can be undertaken. In the output window holding the estimated diversity index for each sample, first select a sample by clicking on its title. The selected row turns blue and a symbol indicates that you have selected the row (image below). Then select the bootstrap button at the top of the window.

Alternatively, use the Bootstrap All button to calculate the confidence intervals for all samples. Bootstrap confidence intervals are not automatically calculated because with large data sets they can take a long time to calculate.



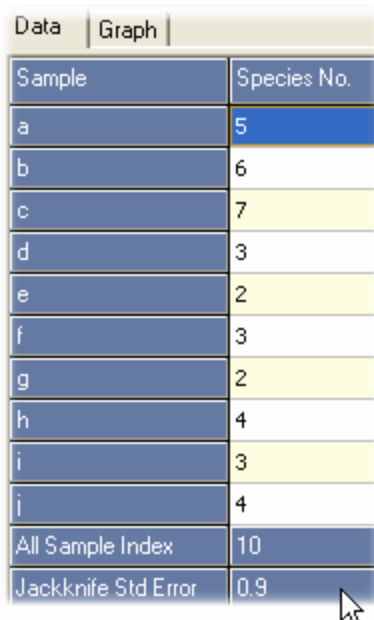
Sample	H	Variance H
h1	1.387	0.003286
h2	1.655	0.00403
h3	1.973	0.004251
h4	1.731	0.04416
h5	1.635	0.01813
h6	2.293	0.01464
h7	1.585	0.01702
h8	1.943	0.01171

All Sample Index: 2.001

8.13 Jackknifing diversity

Jackknifing is a technique for estimating the standard error associated with an estimate such as a diversity or evenness index.

When any of the appropriate alpha diversity or evenness measures is selected within *Species Diversity and Richness* jackknife estimate of the standard error for the combined (all sample) index will be given at the bottom of the output (see below).



Sample	Species No.
a	5
b	6
c	7
d	3
e	2
f	3
g	2
h	4
i	3
j	4
All Sample Index	10
Jackknife Std Error	0.9

For this Jackknife estimate to have meaning all of the samples included in the calculation must be replicates all from the same habitat. To select the samples for inclusion in the jackknifing procedure see [Selecting Samples to Analyse](#).

See [How jackknifing works](#) for details of the computation method.

Strictly speaking jackknifing cannot be used to calculate confidence intervals, just standard errors and bias. A [boot strap](#) must be used for confidence intervals.

8.13.1 How jackknifing works

The jackknife method uses less computational effort than a bootstrap analysis and is thus often quicker and easier to undertake. However, it cannot be used to estimate confidence intervals.

The general scheme is as follows:

1. Consider a situation where a parameter is to be estimated from n samples, e.g. a species diversity index from 10 kicknet samples from a stream.
2. Use all n samples to calculate the parameter of interest, E , the diversity index.
3. Now remove one of the samples at a time and recalculate the parameter of interest, E_i .
4. Calculate the pseudovalues for the parameter of interest :

$$\phi_i = nE - (n-1)E_i$$

where n = the total number of replicates, E = the estimate for all n samples, E_i = the estimate with sample i removed.

5. Estimate the mean and standard error of the parameter of interest from the n pseudovalues. The jackknife estimate of bias is E - the mean of the pseudovalues and the estimate of the standard error of the sample is simply the standard error of the pseudovalues:

$$SE = \sqrt{\frac{\sum (\phi_i - \bar{\phi})^2}{n(n-1)}}$$

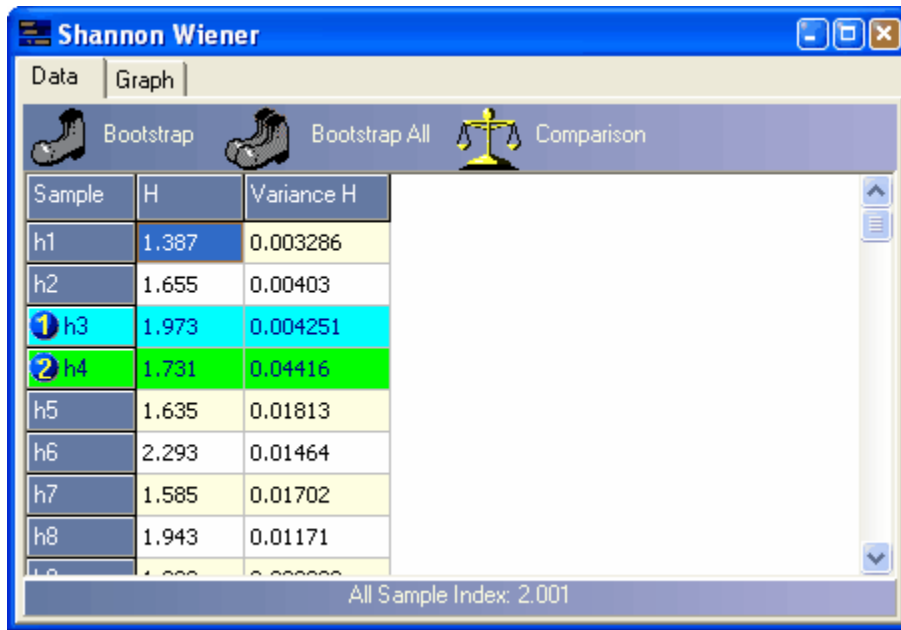
where

$\bar{\phi}$ is the mean of the pseudovalues.

8.14 Testing for significant differences

After a diversity index method has been run, a randomisation test for a significant difference in diversity between two samples can be undertaken. The method used is described by [Solow \(1993\)](#). A [tutorial](#) of this method is available.

In the diversity index output window, first select the two samples for comparison by left clicking with the mouse on their titles. Both rows will be selected as shown below. Which is sample 1 and which is sample 2 is determined by sample order.



Sample	H	Variance H
h1	1.387	0.003286
h2	1.655	0.00403
h3	1.973	0.004251
h4	1.731	0.04416
h5	1.635	0.01813
h6	2.293	0.01464
h7	1.585	0.01702
h8	1.943	0.01171

All Sample Index: 2.001

Then click the Comparison button at the top of the window.



This test re-samples 10,000 times from a distribution of species abundances produced by a summation of the two samples. Thus, for large data sets, the procedure may take some time. When the procedure is run, a timer gauge is shown to indicate the progress made. While the test is calculating, if, at any time, the Stop button is clicked, the results of the test up to the calculated number of samples will be displayed.

Summary of the method

1. The diversity of each of the samples is calculated and the difference between these indices (delta) calculated.
2. The two samples to be tested for a significant difference in their index are added together to form a single joint sample.
3. The individuals in this joint sample are then randomly assigned to two samples each of which has the same number of individuals as the actual two samples.
4. The diversity index for each of these generated samples is then calculated and the difference between these indices (delta) is stored.
5. 10,000 random assignments and calculation of delta are undertaken.
6. The observed value of delta is compared against the observed distribution of delta values generated at random to determine if the observed value for the difference between the indices of the two samples could have been generated by random chance.

7. If the observed value of delta is greater than that observed from 95% of the randomisations then a one-tailed test will find sample 1 to be significantly more diverse than sample 2.
8. If the absolute magnitude of the difference is greater than 95% of the absolute differences of the index generated at random then there is a significant difference between the indices.

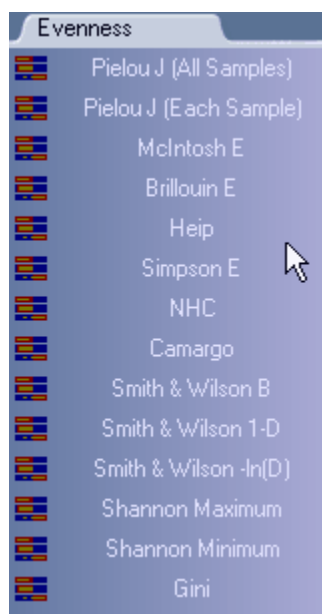
Part



IX

9 Evenness

The second tab, Evenness, in the left hand pane of the program opens a list of 14 equitability or evenness measures, as described below. See [Smith and Wilson's](#) view for a list of the best measures. Equitability or evenness refers to the pattern of distribution of the individuals between the species.



[Pielou J \(All Samples\)](#)
[Pielou J \(Each Sample\)](#)
[McIntosh E](#)
[Brillouin E](#)
[Heip](#)
[Simpson's E](#)
[NHC](#)
[Camargo](#)
[Smith & Wilson B](#)
[Smith & Wilson 1-D](#)
[Smith & Wilson -lnD](#)
[Shannon maximum](#)
[Shannon minimum](#)
[Gini](#)

The output from all of the evenness estimates is presented in similar fashion. In the first column is the name of the sample and in the second the calculated value for the index. Remember that if you have chosen to subset the data then only the samples selected will be represented in the output. If the [Bootstrap](#) function is used then the confidence intervals will be displayed in the third and fourth columns.

The change in diversity between samples can also be shown graphically - just click on the Graph tab in the upper left-hand corner of the output window.

To export or print the data see '[Printing and exporting your results](#)'.

To print the indices, click on the window to activate it. Then click on File: Print or right-click in the window and choose Print from the menu.

9.1 Smith and Wilson's view

Smith and Wilson (1996) considered the following evenness measures as most useful.

- [Simpson's evenness measure](#)
- [Camargo's index](#)
- [Smith and Wilson's B](#)
- [The modified Nee index \(NHC\)](#)

9.2 Pielou J (All samples)

This measure of Equitability compares the observed [Shannon-Wiener index](#) against the distribution of individuals between the observed species which would maximise diversity.

If H is the observed Shannon-Wiener index, the maximum value this could take is $\log(S)$, where S is the total number of species in the habitat. Therefore the index is:

$$J = H/\log(S).$$

There is a big problem with this index. You do not know S ! The program assumes that the total number of species observed over all the samples equals S . Therefore you should consider two points before using this index:

1. Have all the chosen samples come from the same habitat? The heterogeneity test button (see [Testing for Heterogeneity](#)) might be useful to help you to decide.
2. Has the sampling been sufficient to sample all species? Examination of the [species accumulation curve](#) may help you decide this.

9.3 Pielou J (Each Sample)

This measure of equitability compares the observed [Shannon-Wiener index](#) against the distribution of individuals between the observed species which would maximise diversity.

If H is the observed Shannon-Wiener index, the maximum value this could take is $\log(S)$, where S is the total number of species in the habitat. Therefore the index is:

$$J = H/\log(S).$$

There is a big problem with this index. You do not know S ! The program assumes that the total number of species, S , is the maximum observed in any one sample.

9.4 McIntosh E

This is an equitability measure based on the [McIntosh dominance index](#). McIntosh E is defined as ([Pielou, 1975](#)):

$$D = \frac{N - U}{N - \frac{N}{\sqrt{S}}}$$

where N is the total number of individuals in the sample and S is the total number of species in the sample.

9.5 Brillouin E

Evenness (E) for the [Brillouin diversity index](#) (HB) is calculated using:

$$E = \frac{HB}{HB_{MAX}}$$

where HB_{MAX} is calculated as:

$$HB_{MAX} = \frac{1}{N} \ln \frac{N!}{\{([N/S])^{S-r} \cdot ([N/S] + 1)^r\}}$$

with [N/S] = the integer of N/S, and r = N – S[N/S]

9.6 Heip

The Heip evenness measure ([Heip,1974](#)) is defined as:

$$E = \frac{(e^H - 1)}{(S - 1)}$$

where

H is the [Shannon diversity index](#) and S is species number.

Heip (1974) developed this index to remove the dependence on S that earlier indices possessed. Further, he felt that previous indices did not always give a low value when an ecologist would have thought evenness to be low. A property which Heip considered important was that this index remains constant when the numbers of all species is multiplied by a constant.

9.7 Simpson's E

This index is based on [Simpson's diversity index](#), D and is defined as:

$$E = \frac{1/D}{S}$$

where D is Simpson's diversity index and

S is the number of species.

[Krebs \(1989\)](#) noted that for continuous data or data with large numbers of records the maximum value for Simpson's D is 1/S.

9.8 NHC

NHC is an abbreviation of Nee, Harvey and Cotgreave's evenness measure.

$$E = -2 / \pi \arctan(b)$$

where b is the slope of the [log abundance - rank curve](#) fitted by linear regression. This is also termed the slope of the dominance-diversity curve.

9.9 Camargo

The Camargo evenness index ([Camargo, 1993](#)) is defined as:

$$E = 1 - \left[\sum_{i=1}^s \sum_{j=i+1}^s \left(\frac{p_i - p_j}{S} \right) \right]$$

where

p_i is the proportion of species i in the sample;
 p_j is the proportion of species j in the sample and
 S is the total number of species.

9.10 Smith & Wilson B

Smith and Wilson's evenness index B ([Smith & Wilson, 1996](#)) is defined as:

$$E = 1 - \frac{2}{\pi \arctan \left\{ \frac{\sum_{i=1}^s \left(\ln n_i - \sum_{j=1}^s \ln n_j / S \right)^2}{S} \right\}}$$

where

n_i is the number of individuals in species i ;
 n_j is the number of individuals in species j and
 S is the total species number.

This index is based on the variance in abundance. The variance is calculated using log abundances which means the index examines proportional differences between species. The variance obtained is multiplied by the factor $-2/\pi \arctan()$ to give an index with the range 0 to 1 with 0 representing minimum evenness.

9.11 Smith & Wilson 1-D

Smith and Wilson's evenness index 1-D ([Smith & Wilson, 1996](#)) is defined as:

$$E = \frac{1 - D}{1 - \frac{1}{S}}$$

where

D is [Simpson's diversity index](#) and
 S is the total number of species.

9.12 Smith & Wilson -lnD

Smith and Wilson's evenness index 1/D ([Smith & Wilson, 1996](#)) is defined as:

$$E = \frac{-\ln D}{\ln S}$$

where

D is [Simpson's diversity index](#) and
 S is the total number of species.

9.13 Shannon maximum

This is simply the maximum value the [Shannon-Wiener index](#) could produce for the given data set and is given by $\ln(S)$, where S is the total number of species.

9.14 Shannon minimum

This is simply the minimum value the [Shannon-Wiener index](#) could produce for the given data set and is given by :

$$S_{\min} = \ln(N) - \left[\frac{(N - S + 1) \ln(N - S + 1)}{N} \right]$$

where

N is the total number of individuals in the sample and
S is the total number of species.

9.15 Gini

The Gini coefficient is defined as (Gini, 1912):

$$E = \frac{2}{mS^2} \left(\sum_{i=1}^S (S+1-i)x_i \right) - \frac{1}{S}$$

where

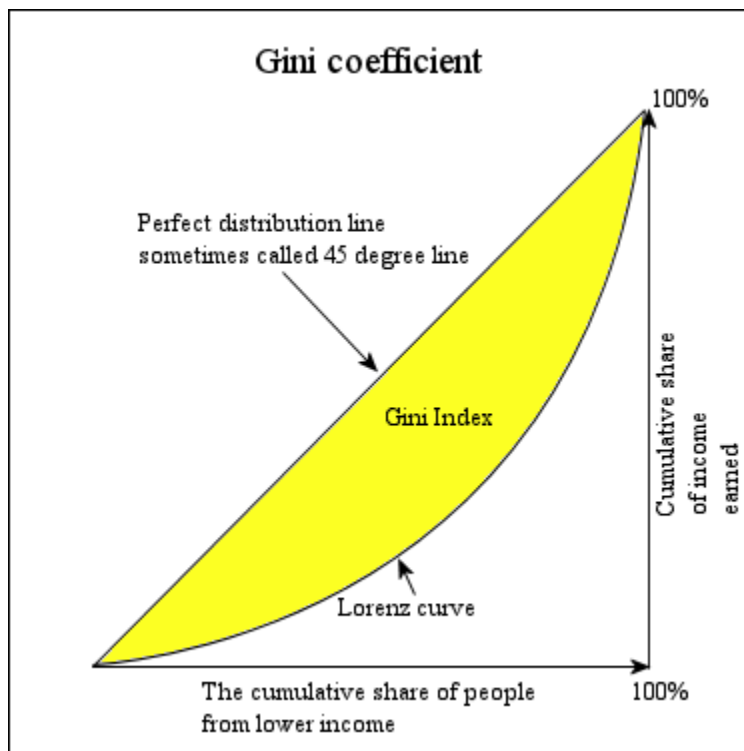
S is the number of species in the sample

x_i is the abundance of the i th species ranked from least to most abundant, $i = 1$ to S
and

m is the mean abundance of a species - the mean of the x_i values.

The Gini coefficient is a measure of inequality developed by the Italian statistician Corrado Gini and published in his 1912 paper "Variabilità e mutabilità". It is usually used to measure income inequality, but can be used to measure any form of uneven distribution. The Gini coefficient is a number between 0 and 1, where 0 corresponds with perfect inequality and 1 corresponds with perfect equality (where each species has the same abundance).

If every species has the same abundance then the species rank curve would be a straight line at 45 degrees (see below); because of inequalities in abundance the actual curve lies below this line (see below). The Gini coefficient is ratio of the areas below these curves.



Part



10 SHE Analysis

10.1 SHE analysis

SHE analysis examines the relationship between S ([species richness](#)), H (information – the [Shannon-Wiener diversity index](#)) and E (evenness as measured using the Shannon-Wiener evenness index, otherwise known as [Pielou J](#)) in the samples. It is therefore an approach to look at the contribution of species number and equitability to changes in diversity. SHE analysis follows the way these parameters changes with increasing sampling effort.

The output shows a spreadsheet and plot of S, H and E for all selected samples.

SHE analysis is useful for identifying ecotones (regions where different ecological communities intersect, such as the edge of lakes) (Hayek & Buzas, 1997). It is also thought to be a useful method for testing whether the data conform most closely to a log-normal, log-series or MacArthur's broken stick model. It is probably the most effective practical method for testing for 'goodness-of-fit' to these models.

The figure below is reproduced from Magurran 2004 and shows the way in which SHE analysis can be used to identify the fit to broken stick, log normal and log series species abundance distributions. Each model shows a different response in equitability and species richness with increased sampling effort.

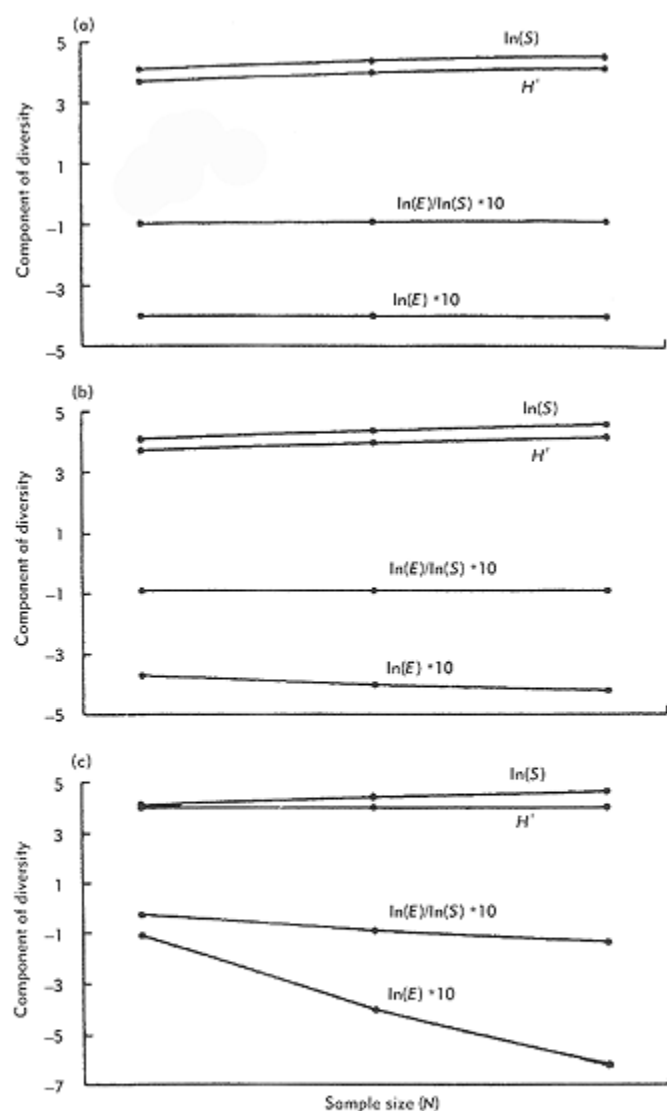


Figure 4.3 SHE analysis plots showing expected patterns for (a) broken stick, (b) log normal, and (c) log series distributions in relation to increasing N . Both $\ln(E)/\ln(S)$ and $\ln(E)$ are multiplied by 10. In the broken stick both S and H' are expected to increase and E to stay constant. The log normal is associated with an increase in S and H' but a decline in E . With the log series S will increase, H' will remain constant, and E will decrease. (Redrawn with permission from Hayek & Buzas 1997.)

Part



XI

11 Diversity ordering

Two diversity ordering methods, [Rényi](#) and [Right-tailed Sum](#), are listed when the diversity ordering button is selected. Simply click on the icon to run either method.

Different diversity indices may differ in the ranking they give to communities (Hurlbert, 1971; Tothmeresz, 1995). An example from Tothmeresz (1995) illustrates the point. Consider three artificial communities with the following sets of species abundances for each of which diversity has been calculated using both

Shannon-Wiener (H) and (2) Simpson's (D):

Community A : {33,29,28,5,5}, H = 1.3808, D = 0.7309,

Community B: {42,30,10,8,5,5}, H = 1.4574, D = 0.7194,

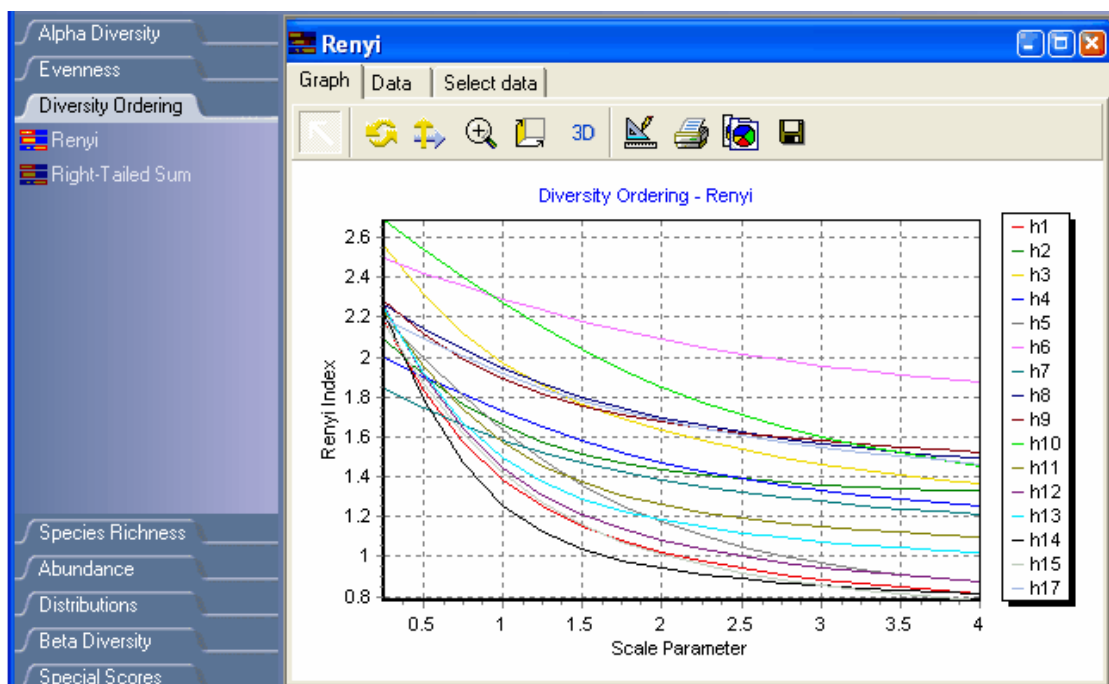
Community C: {32,21,16,12,9,6,4}, H = 0.639, D = 1.822,

Because $H(B) > H(A)$ it could be argued B is the most diverse, however, as $D(A) > D(B)$ the opposite conclusion could also be entertained. Communities such as A and B which cannot be ordered are termed non-comparable.

Such inconsistencies are an inevitable result of summarising both relative abundance and species number using a single number (Patil & Taillie 1979). Diversity profiles offer a solution to this problem by identifying those communities that are consistent in their relative diversity. This requires the use of a diversity index family, of which there are a number to choose from (Tothmeresz, 1995).

The output window

The output is a tabbed display. The first tab shows the graph of the indices for each sample. The graph can be changed, copied or saved using the Chart menu on the main window. The second tab shows a table of the plotted data. These data can be printed using File: Print, or the Print button on the Chart toolbar, and copied using the Edit menu. Diversity ordering results in a line for every sample, which for a large data set can result in a meaningless mess if they are all plotted together. The program selects only the first 16 samples to plot. To select the samples to be displayed on the graph use the Select Data tab.



[The Rényi family](#)

[Right-tailed sum](#)

11.1 Renyi family

When comparing the diversity of samples it is important to consider if the relative diversity changes with the diversity index used. If it does, then it is clear that any arguments based on relative magnitude of the index might not be robust. This method allows the relative magnitude across a range of indices to be compared.

Of the diversity ordering methods (See [Diversity Ordering](#)), perhaps the most generally useful is Rényi's family ([Rényi, 1961](#)) which is based on the concept of entropy and is defined as:

$$H_{\alpha} = \left(\log \sum_{i=1}^s p_i^{\alpha} \right) / (1 - \alpha)$$

where α is the order ($\alpha \geq 0$, $\alpha \neq 0$), p_i the proportional abundance of the i th species and \log the logarithm to a base of choice - often e .

[Hill \(1973\)](#) used an almost identical index N_{α} which is related to H_{α} by the equality

$$H_{\alpha} = \log(N_{\alpha})$$

He demonstrated that N_{α} for $\alpha = 0, 1, 2$ gives the total species number, Shannon-Wiener, H and Simpson's D respectively. Thus by varying α , or 'a' we may generate a range of diversity measures. To test for non-comparability of communities H_{α} is calculated for a range of α values and the results presented graphically. If a community is always greater it can be considered to be more diverse. If two communities cross over they are non-comparable.

11.2 Right-tailed Sum Diversity Ordering

The proportion of the total community made up of each species is arranged in descending order of abundance. Given a total of S species and a proportion of each species given by p_j , then I_i is calculated as a coordinate pair where i is the number of species which are excluded from the summation, termed the order:

$$I_i = \left(i, \sum_{j=i+1}^s p_j \right)$$

Therefore in a sample where many species have a similar abundance, the plot will descend slowly with increasing i , whereas in a sample heavily dominated by large numbers of just a few species, the plot will drop away steeply as i increases.

see [diversity ordering](#)

Part



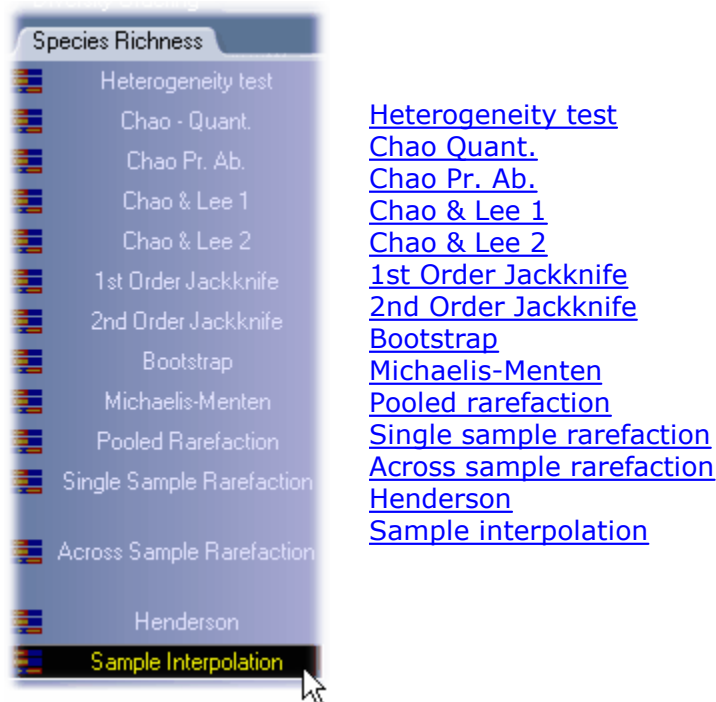
XII

12 Species richness estimators

Species Diversity and Richness offers a variety of methods for species richness estimation. Generally reliable methods are the Chao and Jackknife techniques. Interestingly, with the Chao method the results obtained with presence/ absence data are almost as good as those obtained with quantitative data. If the data set comprises presence/absence data then the only method suitable for you is the Chao Presence/absence method labeled as [Chao Pr. Ab.](#)

Species inventories for particular habitats or localities are frequently required for purposes such as conservation management. Because a complete census is rarely feasible the community must be sampled. An important problem that then arises is to estimate via sampling the total species (or other taxon) number, S_{max} , for the locality. This will give both a measure of the completeness of the inventory and also allow comparison with the species richness of other localities. An estimate of the maximum species number is also useful when assessing if the further information to be gained from continued sampling justifies the cost. At present no clear consensus as to the best approach is available.

The program also offers 4 rarefaction procedures (3 individual-based methods, plus a sample-based method called here Sample Interpolation) to make allowance for differences in sampling effort when comparing the species complement of different samples.



[Using the Michaelis-Menten method as a stopping rule](#)

[Using parametric models based on relative abundance](#)

The output from all of the species richness estimators are presented in similar fashion.

On the top bar there is a drop-down menu that allows the selection of the [number of random selections](#) of sample order (R). To let you see how the estimators improve with the addition of further samples estimators are calculated for 1,2, 3, 4n

samples, where n is the total number of samples in the data set. These calculations will be performed R times and the estimators presented are averages over the R runs. The samples selected for each run are chosen at random. If you would like to see the calculations performed on your actual data in the sample order of the original data set then set R to 1.

The change in an estimator with sample number can also be shown graphically, just click on the Graph button in the upper left-hand corner of the window holding the data.

To export or print the data see [Printing and exporting your results](#).

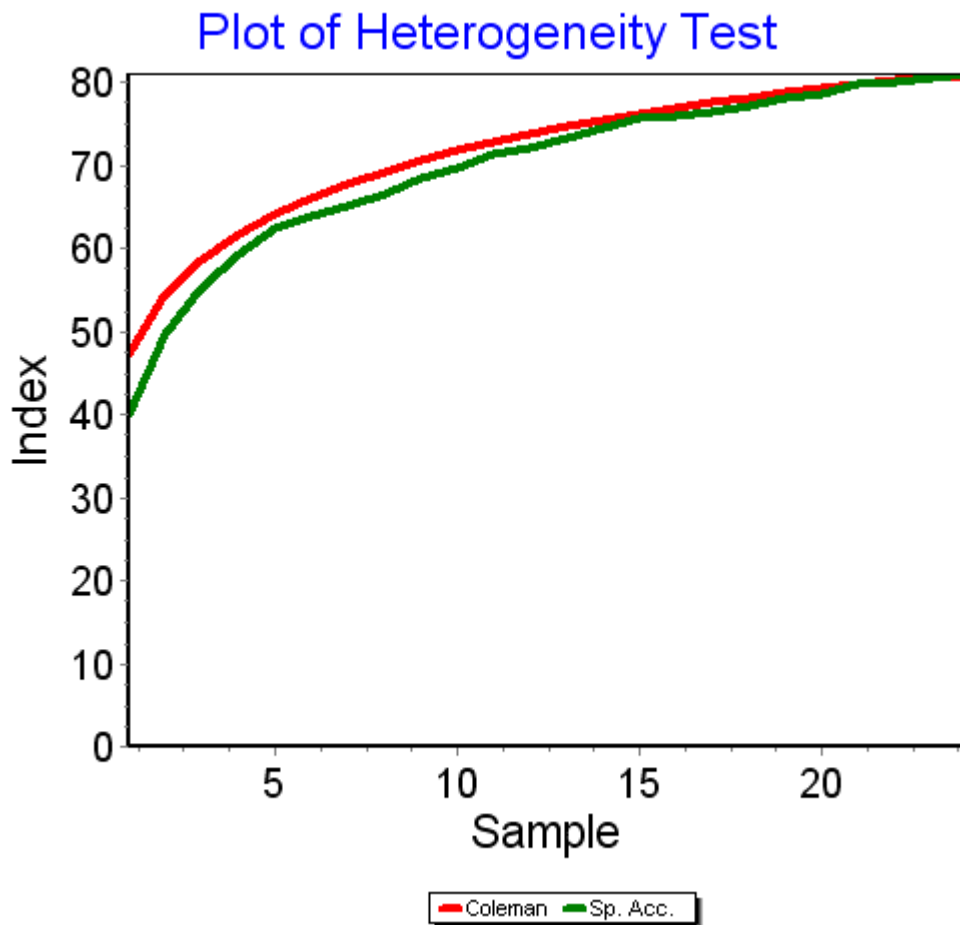
12.1 Heterogeneity test

As extrapolation to estimate the total species complement for the habitat, S_{\max} , is only possible if the species accumulation curve is derived from a homogeneous (stable) community, the first task is to look for heterogeneity. If species abundance rather than presence-absence data is available, [Colwell & Coddington \(1994\)](#) suggest comparing the mean randomised species accumulation curve with the curve expected if all the individuals caught over all the samples were randomly assigned to the samples. If the expected curve rises more steeply from the origin then heterogeneity is greater than could be explained by random sampling error alone. The expected curve calculated by *Species Diversity and Richness* is that described by [Coleman \(1981\) & Coleman et al. \(1982\)](#). The increase in species number over a series of samples, S_a , is calculated as:

$$S_a = S_{TOT} - \sum_{i=1}^{i=S_{TOT}} (1 - \alpha)^{n_i}$$

where S_{TOT} is the total species number recorded, n_i the total number of individuals belonging to species i , a the sample number and α the fraction of the total sampling effort undertaken by sample a .

The graph below was created from a very large data set on estuarine fish. It can be seen that the Coleman curve is above the observed species acquisition curve, suggesting some sample heterogeneity. This might be expected as the samples were collected over a 25 year period and there were considerable changes in the community over the period of study.



12.2 Chao Quantitative

This is the first of the series of estimators of species richness. The output presents a series of estimates using 1 sample, then 2 samples, then 3 samples etc. Thus to get sensible estimates you must ensure that all the data come from samples obtained from the same community.

If [Number of Random Selections](#) (R) is set > 1 then the sequence samples are selected R times at random from the complete set of samples and the mean estimate calculated. As this removes sample order effects it is generally to be recommended.

By looking at the progressive change in the estimates it is possible to assess if sufficient samples have been taken to stabilise the estimate.

A simple plot of the way the estimate changes with the number of samples used is shown by clicking on the Graph tab in the results window.

At present there are no clear indications as to which of these measures is superior and they all should be used with caution. The Chao estimators seem to be some of the best available.

Using the observed number of species represented by one, a, or two, b, individuals in the sample Chao (1984) derived the simple estimator:

$$\hat{S}_{\max} = S_{\text{obs}} + \left(\frac{a^2}{2b} \right) \quad \text{Equ(1)}$$

where S_{obs} is the actual number of species in the sample, a the number of species represented by a single individual, and b the number of species represented by two individuals.

[Chao \(1987\)](#) gives the variance of this estimate as:

$$\text{var}(\hat{S}_{\max}) = b \left[\left(\frac{a/b}{4} \right)^4 + \left(\frac{a}{b} \right)^3 + \left(\frac{a/b}{2} \right)^2 \right] \quad \text{Equ(2)}$$

Note that when all the species have been observed more than twice, the census is considered complete.

12.3 Chao Presence/absence

This is the second of the series of estimators of species richness. The output presents a series of estimates using 1, 2, 3 and so on samples. Thus to get sensible estimates you must ensure that all the data comes from samples obtained from the same community.

If [Number of Random Selections](#) (R) is set > 1 then the sequence samples are selected R times at random from the complete set of samples and the mean estimate calculated. As this removes sample order effects it is generally to be recommended.

By looking at the progressive change in the estimates it is possible to assess if sufficient samples have been taken to stabilise the estimate.

A simple plot of the way the estimate changes with the number of samples used is shown by clicking on the graph button in the upper left hand corner of the window.

These methods are discussed by [Colwell & Coddington \(1995\)](#). At present there are no clear indications as to which is superior and they all should be used with caution. Developed by [Chao \(1984\)](#) this estimator is

$$S = S(\text{obs}) + G/2M,$$

where G is the square of the number of species that only occur in one sample (L), M is the number of species that occur in only two samples and $S(\text{obs})$ the observed species number.

The variance is calculated as for the [Chao quantitative](#) estimator with L replacing a and M replacing b .

12.4 Chao & Lee 1

This is also known as an ACE estimator of richness (Abundance-based Coverage Estimator of species richness). The method was developed by Chao and Lee (1992)

It is assumed that

$$S_{\text{obs}} = S_{\text{rare}} + S_{\text{abund}}$$

The sample coverage estimate based on abundance data is defined as:

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

where

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

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

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

The estimate the coefficient of variation of the Fi's is

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

If [Number of Random Selections](#) (R) is set > 1 then the sequence samples are selected R times at random from the complete set of samples, and the mean estimate calculated. As this removes sample order effects it is generally to be recommended.

By looking at the progressive change in the estimates it is possible to assess if sufficient samples have been taken to stabilise the estimate.

A simple plot of the way the estimate changes with the number of samples used is shown by clicking on the Graph tab.

12.5 Chao & Lee 2

This is also known as the ICE (Incidence-based Coverage) Estimator of species richness (Lee and Chao 1994).

It is assumed that

$$S_{\text{obs}} = S_{\text{inf r}} + S_{\text{freq}}$$

The sample coverage estimate based on incidence data is

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

where

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

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

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

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

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

Note: The formula for ICE is undefined when all infrequent species are Uniques ($Q_1 = N_{infr}$, yielding $C = 0$).

The output presents a series of estimates using 1, 2, 3... and so on samples. Thus to get sensible estimates you must ensure that all the data derive from samples obtained from the same community.

If [Number of Random Selections](#) (R) is set > 1 then the sequence samples are selected R times at random from the complete set of samples and the mean estimate calculated. As this removes sample order effects it is generally to be recommended.

By looking at the progressive change in the estimates it is possible to assess if sufficient samples have been taken to stabilise the estimate.

A simple plot of the way the estimate changes with the number of samples used is shown by clicking on the Graph tab.

12.6 1st Order Jackknife

Independently, [Heltse & Forrester \(1983\)](#) and [Burnham & Overton \(1978\)](#) developed the first-order jackknife estimator:

$$\hat{S}_{\max} = S_{obs} + a(n-1/n)$$

where n is the number of samples and a the number of species only found in one sample.

Heltshel & Forrester (1983) give the variance of this estimate as:

$$\text{var}(\hat{S}_{\max}) = \frac{n-1}{n} \left(\sum_0^{s_{obs}} j^2 f_j - \frac{L^2}{n} \right)$$

where f_j is the number of samples holding j of the L species only found in one sample. Further jackknife estimators are discussed in Burnham & Overton (1979) and Smith & van Belle (1984). Experience suggests that this is a good general-purpose estimator of species richness.

The output presents a series of estimates using 1, 2, 3... and so on samples. Thus to get sensible estimates you must ensure that all the data come from samples obtained from the same community.

A simple plot of the way the estimate changes with the number of samples used is shown by clicking on the Graph tab.

If [Number of Random Selections](#) (R) is set > 1 then the sequence samples are selected R times at random from the complete set of samples and the mean estimate calculated. As this removes sample order effects it is generally to be recommended.

By looking at the progressive change in the estimates it is possible to assess if sufficient samples have been taken to stabilise the estimate.

12.7 2nd Order jackknife

[Burnham & Overton \(1978\)](#) developed the second-order jackknife estimator:

$$\hat{S}_{\max} = S_{obs} + \left[\frac{L(2n-3)}{n} - \frac{M(n-2)^2}{n(n-1)} \right]$$

where L is the number of species only found in one sample and M is the number of species only found in two samples.

12.8 Bootstrap

A bootstrap estimate of species richness can be calculated as follows (Smith & van Belle, 1984):

1. Randomly select with replacement n samples from the total available, and calculate:

$$S_{\max} = S_{obs} + \sum_{i=1}^{s_{obs}} (1-p_i)^n$$

where p_i is the proportion of the n that has species i present.

2. Repeat step 1 a large number of times - say 50 to 200, and calculate the mean estimate of S_{max} and the variance as follows:

$$\text{var}(S_{max}) = \sum (1 - p_i)^n [1 - (1 - p_i)^n] + \sum \sum [q_{i,j}^n - (1 - p_i)^n (1 - p_j)^n]$$

where q_{ij} is the proportion of the n bootstraps which hold both species i and j .

The output presents a series of estimates using 1, 2, 3... and so on samples. Thus to get sensible estimates you must ensure that all the data come from samples obtained from the same community.

If [Number of Random Selections](#) (R) is set > 1 then the sequence samples are selected R times at random from the complete set of samples and the mean estimate calculated. As this removes sample order effects it is generally to be recommended.

By looking at the progressive change in the estimates it is possible to assess if sufficient samples have been taken to stabilise the estimate.

A simple plot of the way the estimate changes with the number of samples used is shown by clicking on the Graph tab in the results window.

12.9 Michaelis-Menten

The asymptotic behaviour of the accumulation curve can also be modelled as the hyperbola:

$$S(n) = \frac{S_{max} n}{B + n}$$

where S_{max} and B are fitted constants. This is the Michaelis-Menten equation used in enzyme kinetics and thus there is an extensive literature discussing the estimation of its parameters, which unfortunately presents considerable statistical difficulties. One approach favoured by (Raaijmakers 1987) is to calculate S_{max} and B using their maximum likelihood estimators. While this method of estimating S_{max} has been criticised by (Lamas *et al.* 1991) it seems the best option available and is used by *Species Diversity and Richness*.

The output presents a series of estimates using 1, 2, 3... and so on samples. Thus to get sensible estimates you must ensure that all the data comes from samples obtained from the same community.

If [Number of Random Selections](#) (R) is set > 1 then the sequence samples are selected R times at random from the complete set of samples and the mean estimate calculated. As this removes sample order effects it is generally to be recommended.

By looking at the progressive change in the estimates it is possible to assess if sufficient samples have been taken to stabilise the estimate.

A simple plot of the way the estimate changes with the number of samples used is shown by clicking on the Graph tab in the results window.

12.10 Pooled rarefaction

Species Diversity and Richness offers a number of rarefaction methods. If you select Pooled Rarefaction the program undertakes individual-based rarefaction on the pooled data for all of the samples within the data set. First all of the samples are summed to form a grand sample. Then the average number of individuals in a single sample is calculated. The standard method for both the finite and infinite versions of the rarefaction curve of [Heck et al. \(1975\)](#) is then used to calculate the species number as the number of individuals increases. This procedure of calculating from the sum of all the samples only makes sense if it is assumed that all the samples selected for inclusion are from the same community.

Rarefaction is a procedure for analyzing the number of species (species richness) among collections, when all collections are scaled down to the same number of individuals. This scaling procedure was termed 'rarefaction' by Sanders (1968) who used an incorrect equation which was corrected by [Hurlbert \(1971\)](#). The number of species, S_n , that can be expected from a random sample of n individuals, drawn without replacement from N individuals distributed among S species, is given by

$$E(S_n) = \sum_{i=1}^S \left[1 - \frac{\binom{N - N_i}{n}}{\binom{N}{n}} \right]$$

where S is the total number of species found in the collection, and N_i is the number of individuals of the i th species.

The formula computes the expected number of species in a random sample of n individuals as the sum of the probabilities that each species will be included in the sample.

The variance of the estimate was given by [Heck et al \(1975\)](#) as:

$Var(S_n) =$

$$\sum_{i=1}^s \left\{ \left[\frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \left[1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \right\}$$

$$+ 2 \sum_{i < j} \left\{ \left[\frac{\binom{N-N_i-N_j}{n}}{\binom{N}{n}} \right] \right.$$

$$\left. - \left[\frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \left[\frac{\binom{N-N_j}{n}}{\binom{N}{n}} \right] \right\}$$

These procedures calculate the number of combinations of the data and thus require a considerable amount of computation for large data sets. In addition to the estimated species richness the output also includes the standard deviation of the estimates.

Sampling is assumed to be without replacement for the Finite version, and with replacement for the Infinite version. Generally it is the Finite version which is quoted by authors. The above equations are for sampling without replacement.

See also:

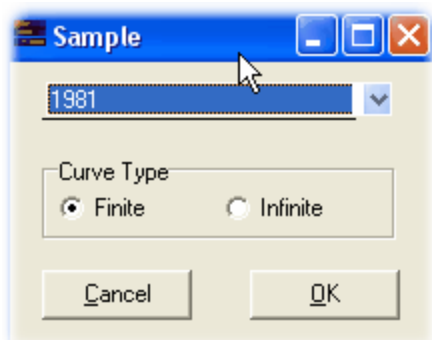
[Single sample rarefaction](#)

[Across sample rarefaction](#)

[Sample interpolation](#)

12.11 Single sample rarefaction

This method estimates how the species number in a selected sample changes with the number of individuals. When selected, a dialog box appears where you select which sample to apply the method to, and whether the Finite or Infinite version of the rarefaction curve is required.



Sampling is assumed to be without replacement for the Finite version, and with replacement for the Infinite version. Generally it is the Finite version which is quoted by authors.

The tabulated output presents the change in estimated species richness with the number of individuals and the standard error of the estimate. The Graph tab shows the output graphically.

Rarefaction is a procedure for analyzing the number of species (species richness) among collections, when all collections are scaled down to the same number of individuals. This scaling procedure was termed 'rarefaction' by Sanders (1968) who used an incorrect equation which was corrected by [Hurlbert \(1971\)](#). The number of species, S_n , that can be expected from a random sample of n individuals, drawn without replacement from N individuals distributed among S species, is given by

$$E(S_n) = \sum_{i=1}^S \left[1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right]$$

where S is the total number of species found in the collection, and N_i is the number of individuals of the i th species.

The formula computes the expected number of species in a random sample of n individuals as the sum of the probabilities that each species will be included in the sample.

The variance of the estimate was given by [Heck et al \(1975\)](#) as:

$$\begin{aligned} \text{Var}(S_n) = & \sum_{i=1}^S \left\{ \left[\frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \left[1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \right\} \\ & + 2 \sum_{i < j}^S \left\{ \left[\frac{\binom{N-N_i-N_j}{n}}{\binom{N}{n}} \right] \right. \\ & \left. - \left[\frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \left[\frac{\binom{N-N_j}{n}}{\binom{N}{n}} \right] \right\} \end{aligned}$$

The standard error is the square root of the variance.

These procedures calculate the number of combinations of the data and thus require a considerable amount of computation for large data sets. In addition to the estimated species richness the output also includes the standard deviation of the estimates.

Sampling is assumed to be without replacement for the Finite version, and with replacement for the Infinite version. Generally it is the Finite version which is quoted by authors.

See also:

[Pooled rarefaction](#)

[Across sample rarefaction](#)

[Sample interpolation](#)

12.12 Across sample rarefaction

This method estimates the species richness of every sample in your data set for a standard number of individuals. By standardising on a constant number of individuals it is possible to compare the samples for a standardised effort and thus determine if species richness varies between samples. This approach is particularly useful when it is clear from the [species accumulation curve](#) that species sampling for one or more samples is far from complete.

Species Diversity and Richness uses as the standard number of individuals the minimum number observed in any one of the samples in the data set. It is therefore essential to ensure that you do not apply this approach to a data set that includes samples comprising only one or a few individuals!

Rarefaction is a procedure for analyzing the number of species (species richness) among collections, when all collections are scaled down to the same number of individuals. This scaling procedure was termed 'rarefaction' by Sanders (1968) who used an incorrect equation which was corrected by [Hurlbert \(1971\)](#). The number of species, S_n , that can be expected from a random sample of n individuals, drawn without replacement from N individuals distributed among S species, is given by

$$E(S_n) = \sum_{i=1}^S \left[1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right]$$

where S is the total number of species found in the collection, and N_i is the number of individuals of the i th species.

The formula computes the expected number of species in a random sample of n individuals as the sum of the probabilities that each species will be included in the sample.

The variance of the estimate was given by [Heck et al \(1975\)](#) as:

$$\begin{aligned} \text{Var}(S_n) = & \sum_{i=1}^S \left\{ \left[\frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \left[1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \right\} \\ & + 2 \sum_{i < j}^S \left\{ \left[\frac{\binom{N-N_i-N_j}{n}}{\binom{N}{n}} \right] \right. \\ & \left. - \left[\frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \left[\frac{\binom{N-N_j}{n}}{\binom{N}{n}} \right] \right\} \end{aligned}$$

This method estimates the species richness of each sample if they all contained the same number of individuals. It is therefore a way of allowing for different sampling effort if the total number of individuals in each sample can be assumed to give a

measure of effort. All of the selected samples must come from the same habitat or community.

12.13 Sample interpolation

This method estimates the number of species that would be observed for different numbers of samples ranging from zero to the total number of samples in the data set. To use this method all of the samples selected must come from the same community or habitat.

Some authors would refer to this method as a type of rarefaction. In this program we have used rarefaction only for methods based on numbers of individuals rather than samples.

The calculations use the method of [Colwell et al](#) (2004) which calculates the species accumulation using a binomial mixture model.

The method is based on the counts of species observed in 1, 2 .. H samples (S_j) where the total number of samples is H.

This method assumes without replacement.

For interpolation, there is an unbiased estimator $E(h)$, the expected number after h samples, that is based on the counts s_j , appropriately weighted by combinatorial coefficients.

Now S_{obs} , the total species number observed, is the sum of the S_j values.

Colwell et al (2004) show that

$$E(h) = S_{obs} - \sum_{j=1}^{j=H} \alpha_{jh} s_j$$

where the combinatorial coefficients α_{jk} are defined by

$$\alpha_{jh} = \frac{(H-h)!(H-j)!}{(H-h-j)!H!}$$

or zero for $j + h > H$.

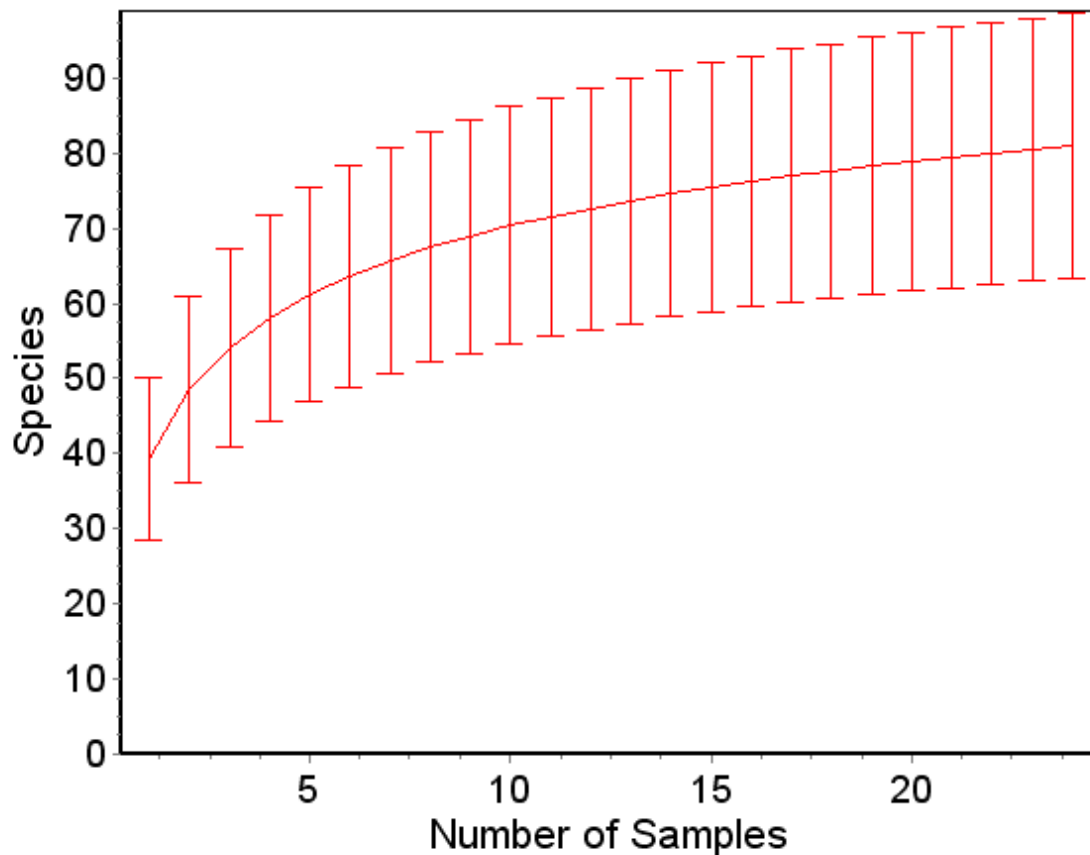
The standard errors are calculated using the estimated variance equation :

$$\text{var}(h) = \sum_{j=1}^H (1 - \alpha_{jh})^2 s_j$$

This is a highly conservative estimate of variance which assumes that the total number of unknown species that could eventually be caught is infinite. However, experience shows that conservative approaches to error are wise because of the dynamic and clumped nature of natural systems.

The upper and lower 95% errors are calculated as 1.96 times the standard deviations.

A typical example of the output is shown below:



12.14 Henderson

[Magurran & Henderson \(2003\)](#) argued that all communities can be divided into residents and tourists (migrants). While it can sometimes be assumed that a sampled habitat is closed to migrants, when sampling large scale systems over extended periods this cannot be the case. This method estimates the species richness of the resident species and the rate of arrival of migrants for a habitat open to migrants.

The species acquisition curve is fitted to a hyperbolic-linear model that takes the form:

$$S(n) = \left[\frac{S_{\max} n}{k_1 + n} \right] + k_2 n$$

where $S(n)$ is the number of species recorded after n samples,

S_{\max} is the total number of resident species,

k_1 is a parameter that determines the rate of acquisition of resident species it is the sampling effort needed to collect half the total number of resident species and

k_2 is a parameter that describes the rate of acquisition of migratory (non-resident)

species.

The hyperbolic-linear model is fitted using non-linear regression. For some sets of data the initial values chosen by the program may be inappropriate, leading to a failure to find a solution. This is most likely to occur if you are using data which do not fit this model.

This equation is best fitted to a species acquisition curve smoothed by averaging the randomised order of the samples.

The output gives the predicted increase in species number with sampling effort which is also presented graphically.

At the top of the output grid (see below) the estimated model parameters are given:

Resident Species: this is the estimated number of resident species living permanently in the habitat. In the example below the community is estimated to hold about 61 residents.

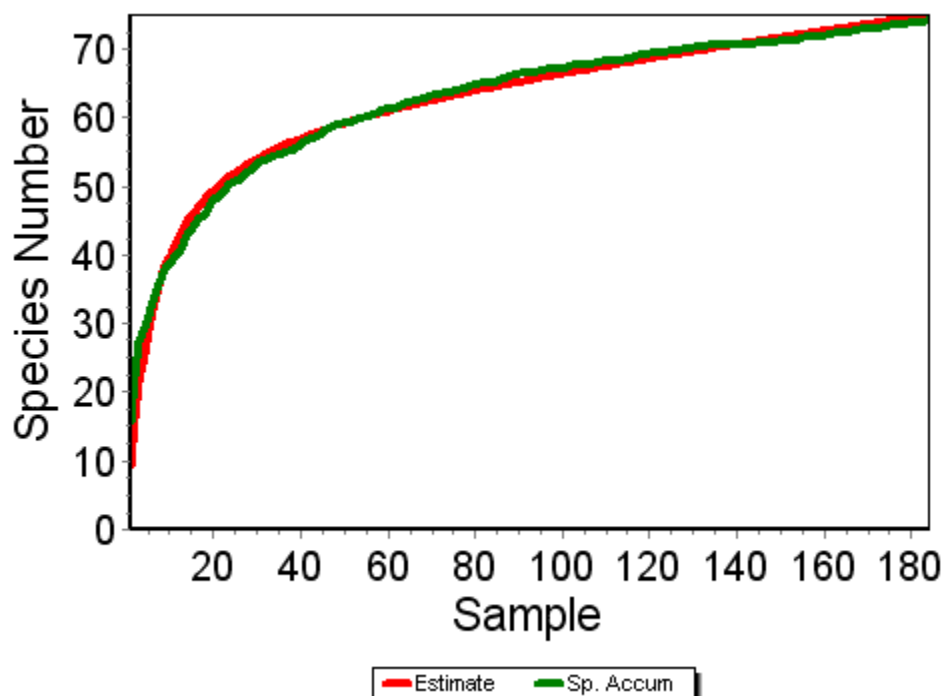
No Sample for 1/2 Sp. Max.: this gives the number of samples required to produce a species list comprising half the total number of resident species.

Migrant Sp./sample: This gives the rate of acquisition of migrants in migrants per sample. In the example below 0.085 migrants arrive per sample, or expressed another way, it requires on average 100 samples to capture 8.5 migrant species.

Estimate	Sp. Accum	Parameters	
9.331	15.9	Resident Species	61.11
16.23	22.6	No Sample for 1/2 Sp. Max	5.609
21.55	27.4	Migrant Sp./sample	0.08525
25.78	29.4		

Note that this model solves using a non-linear regression method that can fail to give a result if the initial parameter guesses are poor or if the data set is inappropriate.

The graph below shows the type of data that are appropriate and the predicted curve that was fitted by the program.



12.15 Number of Random Selections

When calculating species richness or fitting a species accumulation curve, the order in which the samples are listed can have a large impact on the result. To avoid this problem Species Diversity and Richness can randomise the sample order many times and produce an average for the number of species observed after 1, 2 3 .. n samples.

On the top bar there is a drop-down menu that allows the selection of the number of random selections of sample order (R).



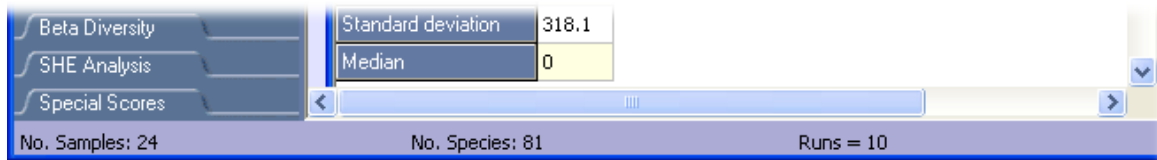
If you do not want the order of the samples randomised then set a value of 1.

For the species richness estimators, R determines the number of randomisations of the samples that will be undertaken. The order in which the samples are analysed can have a large effect on richness estimators. For example, consider a situation in which one sample was unusually species-rich and held many species represented by only one or two individuals. If this were the first sample in the series then it could result in exaggerated species richness estimates. However, if it were the last it would probably have little effect, as most of the species would already have been recorded in other samples. By randomising sample order and averaging the estimates obtained over all the randomisations we remove this sample order effect.

The number of randomisations should be chosen with care. A large data set with a large number of runs can take a few minutes to calculate. If your data set is large (above 50 species by 50 samples) try a small number of randomisations (about 10)

first. If the calculations are undertaken rapidly then increase the number and consider if it has made the estimates more consistent. There is little point in choosing a large value of R if you have a small number of samples.

The number of randomisations of sample order applied to the data is shown at the base of the program window. The image below shows runs = 10, which is a typical default when a data set is opened in *Species Diversity and Richness*.



12.16 Using the Michaelis-Menten method as a stopping rule

The output for this method gives both the estimated asymptote and the observed or average species number. Undertake the Michaelis-Menten calculations with '[Number of random selections of sample order](#)' (R) set to 1. As sampling proceeds the estimated asymptote which initially lies well above the observed species number moves closer to the observed species number. Sampling can be assumed to be complete when the asymptotic estimate equals or is less than the observed species number.

12.17 Using parametric models based on relative abundance

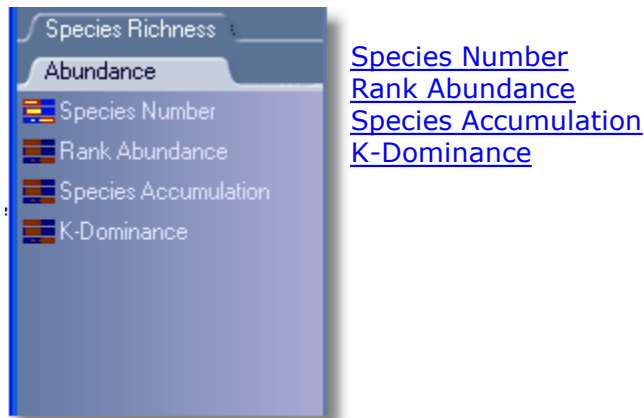
Given quantitative data, species abundance models describing the relative abundance of species within the community such as the log normal can be used to estimate the total species complement, S_{\max} . If the data fit a log normal distribution then the number of unsampled species is given by the missing part of the distribution to the left of the veil-line. Alternative distributions which have been suggested include the Poisson-log normal (Bulmer, 1974) and the log series (Williams, 1964) amongst others. It is unlikely that parametric methods yield reliable estimates for S_{\max} and they should probably be avoided. An estimate of total species richness can be obtained using the [Truncated log normal](#) option within *Species Diversity and Richness*.

Part



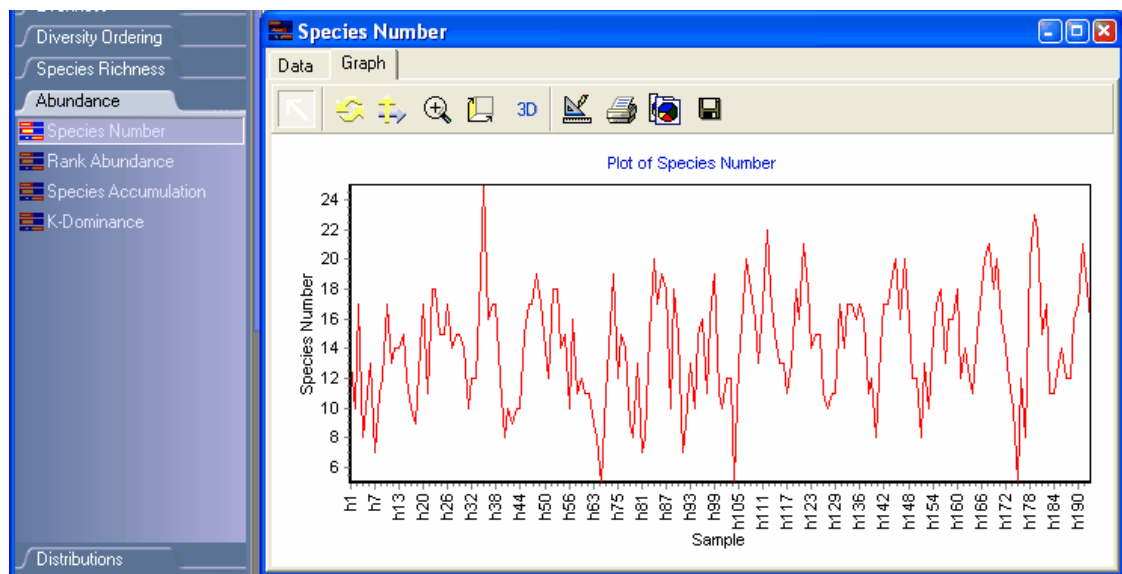
13 Abundance

Selecting the abundance tab offers four simple data display options which can be selected by clicking on their respective titles.



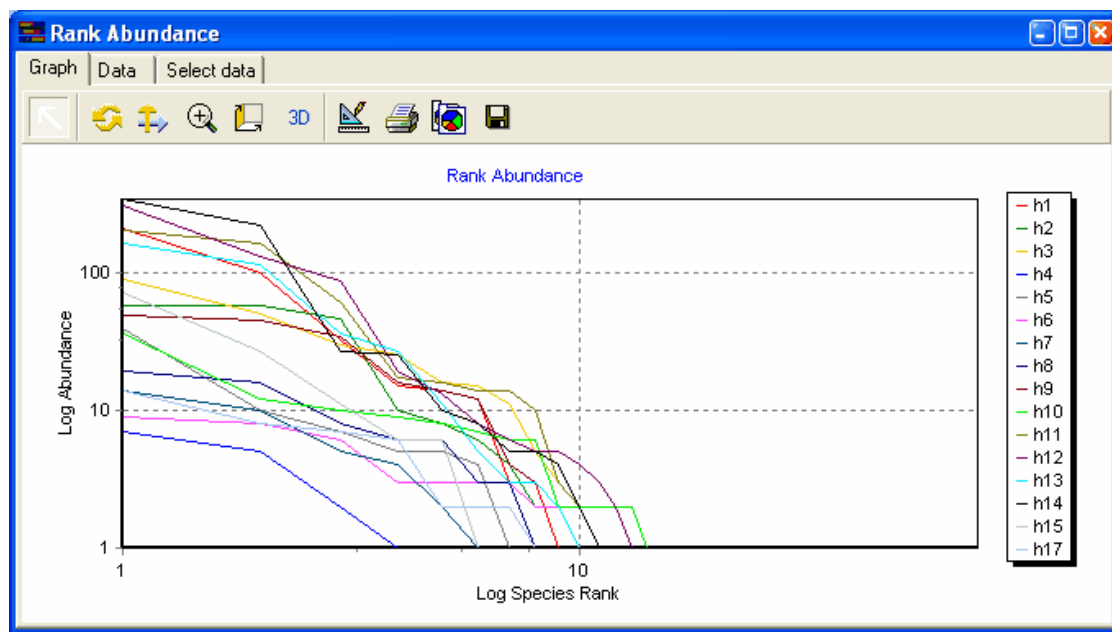
13.1 Species Number

This window gives the total species number for each sample. The plot is selected from the Graph tab.

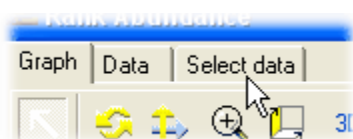


13.2 Rank Abundance

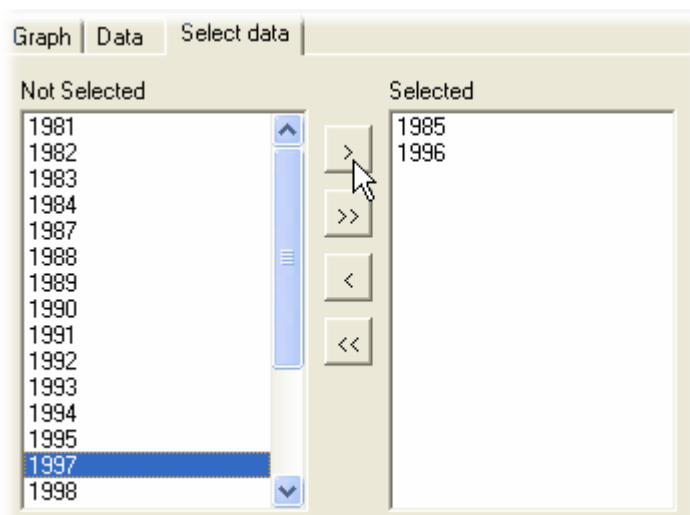
This button shows the graph of log species abundance against log rank order for each sample. It is often the first step in any analysis. Each sample is displayed separately and distinguished by line colour.



If the data set includes a large number of samples then only the first 16 can be unambiguously displayed at one time because of screen/printer colour and resolution limitations. It is therefore possible to choose which samples to include in the plot from the *Select data* tab.



Select using the selection buttons between the two boxes which hold the unselected and selected samples.

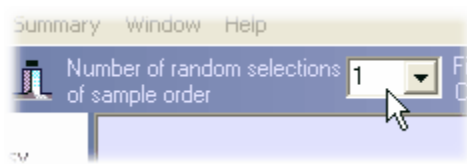


See [Rank abundance plot tutorial](#)

13.3 Species Accumulation Curves

The plot of the cumulative number of species, $S(n)$, collected against a measure of the sampling effort (n) is termed the species accumulation curve. The sampling effort can be measured in many different ways; some examples are the number of quadrats taken, total number of animals handled, hours of observation, or volume of water filtered. As effort increases, gradually more and more of the species living in a habitat will be caught, until eventually only the rarest species or occasional visitors remain unrecorded. When this occurs increased effort will not increase the recorded species number. The species accumulation curve will have reached an asymptote.

If '[Number of Random Selections of sample order](#)' (in the drop-down menu at the top of the program window) is set to 1 (see below) then Species Accumulation will give the simple species accumulation curve for your samples.



See [Species accumulation curve tutorial](#) for more information.

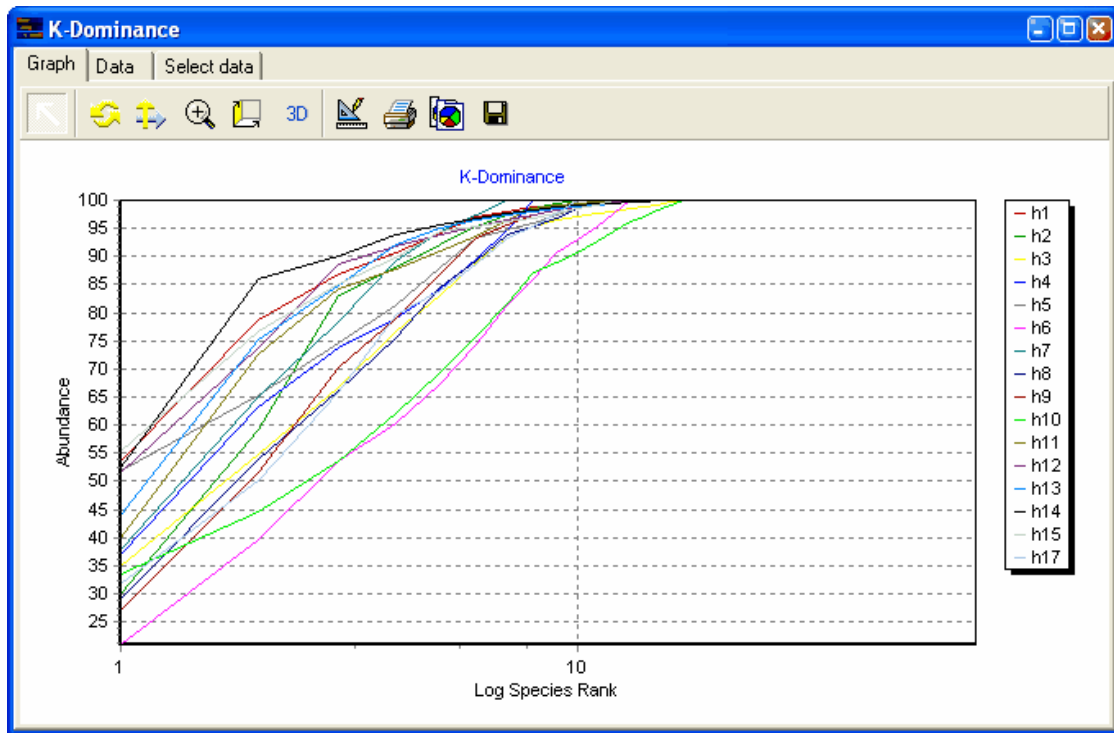
For larger numbers of randomisations the average species accumulation curve is plotted for randomly shuffled samples. For example, with 10 randomisations, the order of the available samples is shuffled 10 times and the average calculated. The effect of this procedure is to remove the effect of sample order on the species accumulation curve and hence produce a smoothed curve.

Using species accumulation curves

When a species accumulation curve approaches an asymptote the user knows that sampling effort has been sufficient to collect most of the species present, and also that the asymptotic value is a measure of the total species complement. We assume here that the method is capable of catching all the species - this is not often the case! Asymptotic behaviour will only be observed if the samples all come from a single, reasonably homogeneous, habitat.

13.4 K dominance

This is a plot of the percentage cumulative abundance plotted against log species rank ([Lambshead et al. 1983](#)). It is a graphical method used for comparing diversity between samples. Note that the lower line has the higher diversity and that if the lines for two samples cross then they will tend to rank differently for different diversity indices.



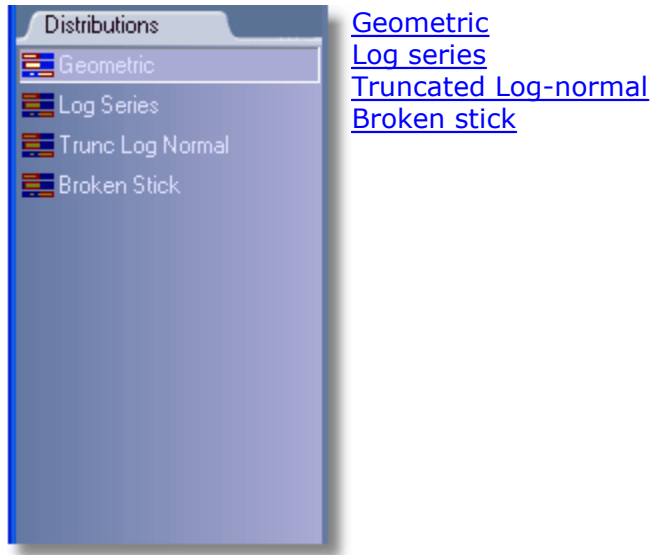
The output shows the plot and a table of the species in rank order for each sample.

Part



14 Fitting abundance distributions

A number of models have been proposed to describe the species - abundance relationship, of which the four models available for fitting in Species Diversity and Richness are by far the most important. All four methods use a similar tabbed output window. When any of these buttons is first activated the selected distribution is fitted to the data in sample 1. To select a different sample for analysis use the drop down menu at the bottom of the window.



14.1 Geometric

The geometric distribution is described by:

$$n_i = NC_k k (1 - k)^{i-1}$$

where

n_i = the number of individuals belonging to the i th species;

N = the total number of individuals;

C_k is a constant to ensure that the total sums to N .

The parameter k is estimated by iteration, after which C_k is calculated.

The output is displayed in a tabbed window.

Species	Observed No.	Expected No.
1	210	131.68
2	98	87.669
3	32	58.369
4	15	38.861
5	14	25.873
6	12	17.226
7	3	11.469
8	3	7.6355

Sample: h1

The O/E tab displays the observed and expected abundances of the species and the Param/Chi tab presents the estimated values for k and the results of a Chi-Squared test of the observed and expected observations. If the value of p is < 0.05 then the distributions are significantly different at the 5% level. The goodness of fit results are displayed in a rich text window. To move around the displayed text click on the text and then use the arrow keys on the keyboard. These results can be copied onto the clipboard by selecting Copy from the Edit menu in the normal Windows manner.

14.2 Log series

The log series distribution is described by:

$$, \frac{ax^2}{2}, \frac{ax^3}{3}, \dots, \frac{ax^n}{n}$$

where each term gives the number of species predicted to have 1,2,3,...,n individuals in the sample.

The parameter a is estimated by iteration, after which x is calculated.

The output is displayed in a tabbed window.

Abundance Class	Upper Bound	Observed	Expected
1	2.5	5	3.8482
2	8.5	2	3.0516
3	256.5	6	5.9137

The O/E tab displays the observed and expected abundances of the species. These are arranged in abundance classes and the upper column gives the upper bound of each class. The Chart tab shows a plot of the observed and expected frequency distributions arranged by class. The Observed are plotted as a histogram and the Expected as a green line. The Param/Chi tab presents the estimated values for α , x and the results of a Chi-Squared test of the observed and expected observations. If the value of p is < 0.05 then the distributions are significantly different at the 5% level. The goodness of fit results are displayed in a rich text window. To move around the displayed text click on the text and then use the arrow keys on the keyboard. These results can be copied and pasted in to other applications by selecting Copy from the Edit menu in the normal Windows manner.

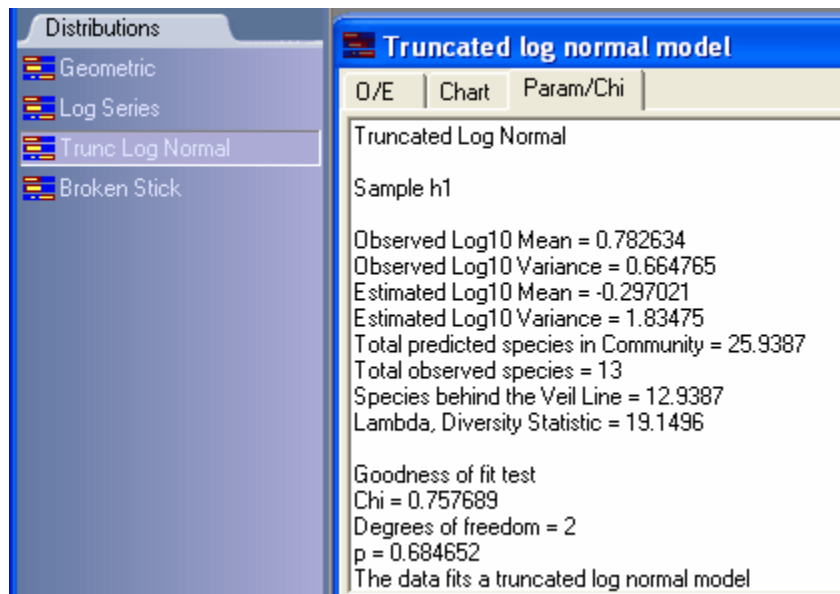
It has been suggested that [SHE analysis](#) is one of the best ways of deciding if a log series log normal or broken strick model gives the best fit to the observed data.

14.3 Truncated log-normal

Many have claimed that the majority of communities display a log normal distribution. However, rarer species will not be fully represented in a finite sample so that in practice we lose the left-hand tail of the distribution. The veil line represents the distance from the right-hand edge of the distribution at which species becomes too rare to occur in the sample.

The output is displayed in a tabbed window. The O/E tab displays the observed and expected abundances of the species. These are arranged in abundance classes and the upper column gives the upper bound of each class. The Chart tab shows a plot of the observed and expected frequency distributions arranged by class. The observed are plotted as a histogram and the expected as a green line. The Param/Chi tab presents the estimated values for the following parameters:

- the observed mean of the logged observations,
- the observed variance of the logged observations,
- the estimated mean of the log normal distribution,
- the estimated variance of the log normal distribution,
- the predicted total number of species in the community,
- the observed number of species,
- the species beyond the veil line, in other words the species number missing from the sample,
- lambda, the diversity statistic (the estimated species number divided by the standard deviation).



The results of a Chi-Squared test of the observed and expected observations are also given. If the value of p is < 0.05 then the distributions are significantly different at the 5% level. The goodness of fit results are displayed in a rich text window. To move around the displayed text click on the text and then use the arrow keys on the keyboard. These results can be copied and pasted in to other applications by selecting Copy from the Edit menu in the normal Windows manner.

It has been suggested that [SHE analysis](#) is one of the best ways of deciding if a log series, log normal or broken stick model gives the best fit to the observed data.

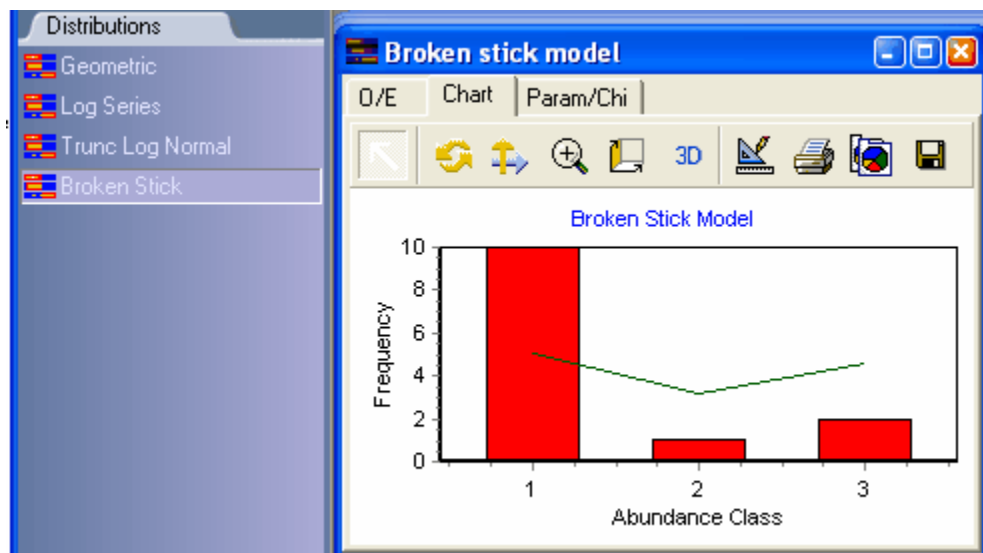
14.4 Broken stick

The broken stick model is calculated using the formula:

$$S(n) = \left[\frac{S(S-1)}{N} \right] \left(\frac{(1-n)}{N} \right)^{s-2}$$

where $S(n)$ is the number of species in the abundance class of species with n individuals.

The output is displayed in a tabbed window.



The O/E tab displays the observed and expected abundances of the species. These are arranged in abundance classes and the upper column gives the upper bound of each class. The Chart tab shows a plot of the observed and expected frequency distributions arranged by class. The observed are plotted as a histogram and the expected as a green line. The Param/Chi tab shows the results of a Chi-Squared test of the observed and expected observations. If the value of p is < 0.05 then the distributions are significantly different at the 5% level. The goodness of fit results are displayed in a rich text window. To move around the displayed text click on the text and then use the arrow keys on the keyboard. These results can be copied and pasted in to other applications by selecting Copy from the Edit menu in the normal Windows manner.

It has been suggested that [SHE analysis](#) is one of the best ways of deciding if a log series log normal or broken stick model gives the best fit to the observed data.

Part



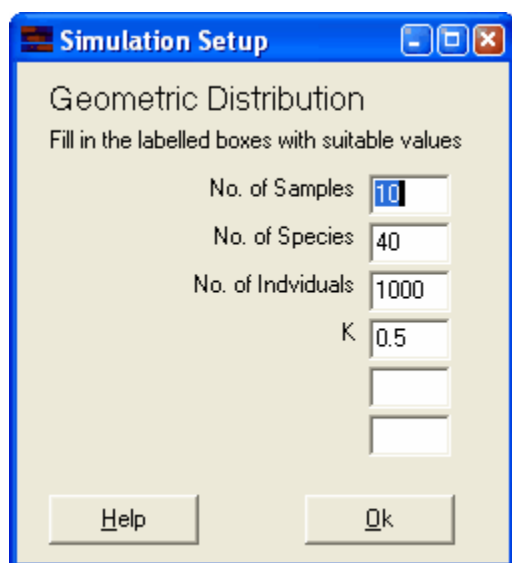
XV

15 Data Simulation

15.1 Data Simulation

Data sets can be simulated that represent samples taken from geometric, log series, truncated log normal or broken stick models of species abundance. The simulated data sets produced become the active data set. Therefore, if you have made changes to an existing data set open in SDR, it should be saved before starting a simulation.

To create a simulated data set simply choose the type of distribution you want from the Simulation drop-down menu on the top bar of the program, fill in the required parameters with reasonable values, and click OK.



When the parameter input window is first opened, *Species Diversity and Richness* will have automatically entered some parameters that will produce sensible numbers. It is possible to choose parameters that will produce nonsense data. A summary of the parameters required, and their meaning, is given below.

No. of samples: The number of independent data sets created – the columns in the data set.

No. of species: This is the number of species in the community.

No. of individuals: For a geometric distribution you need to state the number of individuals in each sample.

K: The parameter for the geometric distribution, it must lie between 0 and 1.

Alpha: The log series diversity parameter, typical values lie between 4 and 6.

No. of Species in sample: The number of species captured in each sample. This parameter is required to simulate the log series.

Mean: The mean species abundance in the population or hypothetical sampling area.

SD: The standard deviation of population abundance for the hypothetical population.

proportion – this is the proportion of the individuals in each hypothetical population that is captured, it must lie between 0 and 1.

Maximum Sp Abundance: The abundance of the commonest species in the hypothetical population.

Once the parameters have been entered, the simulations are produced by clicking OK. The resulting data set will be displayed in a standard data grid. Click OK, and the simulated data will be available for use by all the *Species Diversity and Richness* methods. The user can save the data as a .csv file by choosing File: Save As.

Part



XVI

16 Beta diversity

The beta diversity tab offers one option that is selected by clicking on the name.

[Beta diversity Indices](#)

Note that in all analyses, the results are presented in individual windows, this makes it possible to view the results of different analyses simultaneously.

16.1 Beta diversity indices

β diversity measures the increase in species diversity along transects and is particularly applicable to the study of environmental gradients. It measures two attributes, the number of distinct habitats within a region and the replacement of species by another between disjoint parts of the same habitat. All of the selected samples in the active data set will be used to calculate the indices. ***It is assumed that the samples are arranged in the data grid in their order of occurrence along the transect.***

The output can be printed by choosing File: Print, and copied by

- Choosing Edit: Copy
- Clicking in the output window and using Ctrl+C
- Right-clicking in the output window and selecting Copy

The six indices calculated, which are described below are those considered by Wilson and Schmida (1984). All six only require presence/absence data.

1. Whittaker's β_w

$$\beta_w = \left(\frac{S}{\alpha} \right) - 1$$

where S = the total number of species and α the average species richness of the samples. All samples must have the same size (or sampling effort).

2. Cody's β_c

$$\beta_c = \frac{g(H) + l(H)}{2}$$

where $g(H)$ is the number of species gained and $l(H)$ the number lost moving along the transect.

3, 4, & 5. Routledge's β_R , β_I and β_E

$$\beta_R = \frac{S^2}{2r + S} - 1$$

where S is the total species number for the transect and r the number of species pairs with overlapping distributions.
Assuming equal sample sizes,

$$\beta_T = \log(T) - \left[\left(\frac{1}{T} \right) \sum e_i \log(e_i) \right] - \left[\left(\frac{1}{T} \right) \sum \alpha_i \log(\alpha_i) \right]$$

where e_i is the number of samples along the transect in which species i is present and α_i the species richness of sample i and T is $\sum e_i$.

The third of Routledge's indices is simply

$$\beta_g = \exp(\beta_T) - 1$$

6. Wilson and Schmida's β_T

$$\beta_T = \frac{[g(H) + l(H)]}{2\alpha}$$

where the parameters are defined as for β_c and β_w .

Based on an assessment of the essential properties of a useful index: ability to detect change, additivity, independence of α and independence of sample size Wilson and Schmida (1984) concluded that β_w was best. Schmida and Wilson's own measure, 'T' came a close second in that study.

7. Harrison 1

This is a modification of Whittaker's measure and is given by the equation:

$$\beta = \left\{ \frac{\left[\left(\frac{S}{\bar{\alpha}} - 1 \right) \right]}{(N-1)} \right\} \times 100$$

where α the average species richness of the samples. This measure ranges from 0 (no species turn over) to 100 (every sample holds a unique set of species). This measure allows transects of different size to be compared.

8. Harrison 2

A second modification of Harrison given by the equation:

$$\beta = \left\{ \frac{\left[\left(\frac{S}{\alpha_{\max}} - 1 \right) \right]}{(N-1)} \right\} \times 100$$

where α the maximum species richness in any one sample. This modification is insensitive to species richness.

Part



17 Special indices

A number of methods have been developed to score the quality of freshwater habitats based upon the species or taxa present. *Species Diversity and Richness* offers a range of the methods most frequently used in the United Kingdom and Ireland. These methods will work for most other temperate parts of the world. They could also be used in the tropics, but may no longer be reliable as they do not consider groups such as prawns and crabs which can be abundant in these localities. The program offers 4 related techniques.

[BMWP scores](#)

[ASPT index](#)

[Irish water quality rating](#)

[LIFE scores](#)

17.1 BMWP score

The Biological Monitoring Working Party Score system is a method of assessing water quality using the families of insects and other aquatic invertebrates present. This scoring system is only relevant in the UK.

If this is selected you will be presented with a window offering a choice of BMWP score systems



[Original BMWP](#)

[Revised BMWP](#)

[Habitat Specific - Riffles](#)

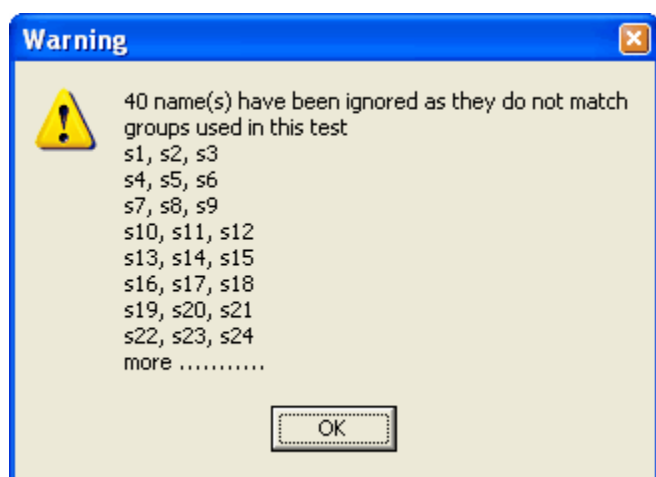
[Habitat Specific - Riffles and Pools](#)

[Habitat Specific - Pools](#)

To use this method you must give the program observation data arranged as samples by family. It is important that you spell the family names correctly because the program scans your list of names and scores those that correspond with a standard list.

Pisces has developed software to automate this process, see [T-group](#).

If your names do not correspond with those in the list, you will see a message similar to this, indicating that corrections must be made.



The data file bmwp.scr must be in the same directory as the data to be analysed. The bmwp.scr file is a text file containing the names and their associated scores.

The standard list is as follows:

Siphonuridae	10	Corduliidae	8	Dytiscidae	5
Heptageniidae	10	Libellulidae	8	Gyrinidae	5
Leptophlebiidae	10	Psychomyiidae	8	Hydrophilidae	5
Ephemerellidae	10	Philopotamidae	8	Clambudae	5
Potamanthidae	10	Caenidae	7	Helodidae	5
Ephemeridae	10	Nemouridae	7	Dryopidea	5
Taeniopterygidae	10	Rhyacophilidae	7	Elminthidae	5
Leuctridae	10	Polycentropodidae	7	Chrysomelidae	5
Capniidae	10	Limnephilidae	7	Curculionidae	5
Perlodidae	10	Neritidae	7	Hydropsychidae	5
Perlidae	10	Viviparidae	6	Tipulidae	5
Chloroperlidae	10	Ancylidae	6	Simuliidae	5
Aphelocheiridae	10	Hydroptilidae	6	Planariidae	5
Phryganeidae	10	Unionidae	6	Dendrocoelidae	5
Molannidae	10	Corophiidae	6	Baetidae	5
Beraeidae	10	Gammaridae	6	Sialidae	4
Odontoceridae	10	Platycnemididae	6	Piscicolidae	4
Leptoceridae	10	Coenagriidae	6	Valvatidae	4
Goeridae	10	Mesoveliidae	6	Hydrobiidae	3
Lepidostomatidae	10	Hydrometridae	5	Lymnaeidae	3
Brachycentridae	10	Gerridae	5	Physidae	3
Sericostomatidae	10	Nepidae	5	Planorbidae	3
Astacidae	10	Naucoridae	5	Sphaeriidae	3
Lestidae	10	Notonectidae	5	Glossiphoniidae	3
Agriidae	8	Pleidae	5	Hirudidae	3
Gomphidae	8	Corixidae	5	Erpobdellidae	3
Cordulegasteridae	8	Haliplidae	5	Asellidae	3
Aeshnidae	8	Hygrobiidae	5	Chironomidae	3
	8		5	Oligochaeta	3
	8		5		2
					1

17.1.1 Habitat specific BMWP

Recently revised BMWP scores for specific stream habitats have been designed.

The Habitat Specific Scores are based on the following substrate compositions
 Riffles: $\geq 70\%$ boulders and pebbles Pool: $\geq 70\%$ sand and silt Riffle/Pool:
 the remainder

This has resulted in a modified scoring system for the families present. The different scoring systems are shown below

	Original BMWP	Revised BMWP	Riffle/ Riffles	Pools	Pools
Planariidae	5	4.2	4.5	4.1	3.7
Dendrocoelidae	5	3.1	2.3	4.1	3.1
Neritidae	6	7.5	6.7	8.1	9.3
Viviparidae	6	6.3	2.1	4.7	7.1
Valvatidae	3	2.8	2.5	2.5	3.2
Hydrobiidae	3	3.9	4.1	3.9	3.7
Lymnaeidae	3	3	3.2	3.1	2.8
Physidae	3	1.8	0.9	1.5	2.8
Planorbidae	3	2.9	2.6	2.9	3.1
Ancylidae	6	5.6	5.5	5.5	6.2
Unionidae	6	5.2	4.7	4.8	5.5
Sphaeriidae	3	3.6	3.7	3.7	3.4
Oligochaeta	1	3.5	3.9	3.2	2.5
Piscicolidae	4	5	4.5	5.4	5.2
Glossiphoniidae	3	3.1	3	3.3	2.9
Hirudidae	3	0	0.3	0.3	
Erpobdellidae	3	2.8	2.8	2.8	2.6
Asellidae	3	2.1	1.5	2.4	2.7
Corophiidae	6	6.1	5.4	5.1	6.5
Gammaridae	6	4.5	4.7	4.3	4.3
Astacidae	8	9	8.8	9	11.2
Siphonuridae	10	11	11		
Baetidae	4	5.3	5.5	4.8	5.1
Heptageniidae	10	9.8	9.7	10.7	13
Leptophlebiidae	10	8.9	8.7	8.9	9.9
Ephemerellidae	10	7.7	7.6	8.1	9.3
Potamanthidae	10	7.6	7.6		
Ephemeridae	10	9.3	9	9.2	11
Caenidae	7	7.1	7.2	7.3	6.4
Taeniopterygidae	10	10.8	10.7	12.1	
Nemouridae	7	9.1	9.2	8.5	8.8
Leuctridae	10	9.9	9.8	10.4	11.2
Capniidae	10	10			10.1
Perlodidae	10	10.7	10.8	10.7	10.9
Perlidae		10		12.5	12.2
Chloroperlidae	10	12.4	12.5	12.1	
Platycnemidae	6	5.1	3.6	5.4	5.7
Coenagriidae	6	3.5	2.6	3.3	3.8
Lestidae	8	5.4		5.4	
Calopterygidae	8	6.4	6	6.1	7.6
Gomphidae	8				
Cordulegasteridae	8	8.6	9.5	6.5	7.6
Aeshnidae	8	6.1	7	6.9	5.7
Corduliidae	8				
Libellulidae	8	5			5
Mesoveliidae	5	4.7	4.9	4	5.1
Hydrometridae	5	5.3	5	6.2	4.9

Gerridae	5	4.7		4.5	5	4.7	
Nepidae		5	4.3		4.1	4.2	4.5
Naucoridae	5	4.3				4.3	
Aphelocheiridae	10	8.9		8.4	9.5	11.7	
Notonectidae	5	3.8		1.8	3.4	4.4	
Pleidae	5	3.9				3.9	
Corixidae	5	3.7		3.6	3.5	3.9	
Haliplidae	5	4		3.7	4.2	4.3	
Hygrobiiidae	5	2.6		5.6	-0.8	2.6	
Dytiscidae	5	4.8		5.2	4.3	4.2	
Gyrinidae	5	7.8		8.1	7.4	6.8	
Hydrophilidae	5	5.1		5.5	4.5	3.9	
Clambidae	5						
Scirtidae	5	6.5		6.9	6.2	5.8	
Dryopidae	5	6.5		6.5			
Elmidae		5	6.4		6.5	6.1	6.5
Chrysomelidae	5	4.2		4.9	1.1	4.1	
Curculionidae	5	4		4.7	3.1	2.9	
Sialidae	4	4.5		4.7	4.7	4.3	
Rhyacophilidae	7	8.3		8.2	8.6	9.6	
Philopotamidae	8	10.6		10.7	9.8		
Polycentropidae	7	8.6		8.6	8.4	8.7	
Psychomyiidae	8	6.9		6.4	7.4	8	
Hydropsychidae	5	6.6		6.6	6.5	7.2	
Hydroptilidae	6	6.7		6.7	6.8	6.5	
Phryganeidae	10	7		6.6	5.4	8	
Limnephilidae	7	6.9		7.1	6.5	6.6	
Molannidae	10	8.9		7.8	8.1	10	
Beraeidae	10	9		8.3	7.8	10	
Odontoceridae	10	10.9		10.8	11.4	11.7	
Leptoceridae	10	7.8		7.8	7.7	8.1	
Goeridae	10	9.9		9.8	9.6	12.4	
Lepidostomatidae	10	10.4		10.3	10.7	11.6	
Brachycentridae	10	9.4		9.3	9.7	11	
Sericostomatidae	10	9.2		9.1	9.3	10.3	
Tipulidae	5	5.5		5.6	5	5.1	
Chironomidae	2	3.7		4.1	3.4	2.8	
Simuliidae	5	5.8		5.9	5.1	5.5	

17.1.2 Revised BMWP

There has been a revision of the BMWP score system based on the the analysis of frequency of occurrence of the families recorded in approximately 17,000 British stream samples.

This has resulted in a modified scoring system for the families present. The different scoring systems are shown below

	Original BMWP	Revised BMWP	Riffle/ Riffles	Pools	Pools
Planariidae	5	4.2	4.5	4.1	3.7
Dendrocoelidae	5	3.1	2.3	4.1	3.1
Neritidae	6	7.5	6.7	8.1	9.3
Viviparidae	6	6.3	2.1	4.7	7.1
Valvatidae	3	2.8	2.5	2.5	3.2
Hydrobiidae	3	3.9	4.1	3.9	3.7
Lymnaeidae	3	3	3.2	3.1	2.8

Physidae	3	1.8		0.9	1.5	2.8
Planorbidae	3	2.9		2.6	2.9	3.1
Ancylidae	6	5.6		5.5	5.5	6.2
Unionidae	6	5.2		4.7	4.8	5.5
Sphaeriidae	3	3.6		3.7	3.7	3.4
Oligochaeta	1	3.5		3.9	3.2	2.5
Piscicolidae	4	5		4.5	5.4	5.2
Glossiphoniidae	3	3.1		3	3.3	2.9
Hirudidae	3	0		0.3	0.3	
Erpobdellidae	3	2.8		2.8	2.8	2.6
Asellidae	3	2.1		1.5	2.4	2.7
Corophiidae	6	6.1		5.4	5.1	6.5
Gammaridae	6	4.5		4.7	4.3	4.3
Astacidae	8	9		8.8	9	11.2
Siphonuridae	10	11		11		
Baetidae	4	5.3		5.5	4.8	5.1
Heptageniidae	10	9.8		9.7	10.7	13
Leptophlebiidae	10	8.9		8.7	8.9	9.9
Ephemerellidae	10	7.7		7.6	8.1	9.3
Potamanthidae	10	7.6		7.6		
Ephemeridae	10	9.3		9	9.2	11
Caenidae	7	7.1		7.2	7.3	6.4
Taeniopterygidae	10	10.8		10.7	12.1	
Nemouridae	7	9.1		9.2	8.5	8.8
Leuctridae	10	9.9		9.8	10.4	11.2
Capniidae	10	10				10.1
Perlodidae	10	10.7		10.8	10.7	10.9
Perlidae		10	12.5		12.5	12.2
Chloroperlidae	10	12.4		12.5	12.1	
Platycnemidae	6	5.1		3.6	5.4	5.7
Coenagriidae	6	3.5		2.6	3.3	3.8
Lestidae	8	5.4			5.4	
Calopterygidae	8	6.4		6	6.1	7.6
Gomphidae	8					
Cordulegasteridae	8	8.6		9.5	6.5	7.6
Aeshnidae	8	6.1		7	6.9	5.7
Corduliidae	8					
Libellulidae	8	5				5
Mesoveliidae	5	4.7		4.9	4	5.1
Hydrometridae	5	5.3		5	6.2	4.9
Gerridae	5	4.7		4.5	5	4.7
Nepidae		5	4.3		4.1	4.2
Naucoridae	5	4.3				4.3
Aphelocheiridae	10	8.9		8.4	9.5	11.7
Notonectidae	5	3.8		1.8	3.4	4.4
Pleidae	5	3.9				3.9
Corixidae	5	3.7		3.6	3.5	3.9
Haliplidae	5	4		3.7	4.2	4.3
Hygrobiidae	5	2.6		5.6	-0.8	2.6
Dytiscidae	5	4.8		5.2	4.3	4.2
Gyrinidae	5	7.8		8.1	7.4	6.8
Hydrophilidae	5	5.1		5.5	4.5	3.9
Clambidae	5					
Scirtidae	5	6.5		6.9	6.2	5.8
Dryopidae	5	6.5		6.5		
Elmidae		5	6.4		6.5	6.1
Chrysomelidae	5	4.2		4.9	1.1	4.1
Curculionidae	5	4		4.7	3.1	2.9
Sialidae	4	4.5		4.7	4.7	4.3
Rhyacophilidae	7	8.3		8.2	8.6	9.6
Philopotamidae	8	10.6		10.7	9.8	
Polycentropidae	7	8.6		8.6	8.4	8.7
Psychomyiidae	8	6.9		6.4	7.4	8

Hydropsychidae	5	6.6	6.6	6.5	7.2
Hydroptilidae	6	6.7	6.7	6.8	6.5
Phryganeidae	10	7	6.6	5.4	8
Limnephilidae	7	6.9	7.1	6.5	6.6
Molannidae	10	8.9	7.8	8.1	10
Beraeidae	10	9	8.3	7.8	10
Odontoceridae	10	10.9	10.8	11.4	11.7
Leptoceridae	10	7.8	7.8	7.7	8.1
Goeridae	10	9.9	9.8	9.6	12.4
Lepidostomatidae	10	10.4	10.3	10.7	11.6
Brachycentridae	10	9.4	9.3	9.7	11
Sericostomatidae	10	9.2	9.1	9.3	10.3
Tipulidae	5	5.5	5.6	5	5.1
Chironomidae	2	3.7	4.1	3.4	2.8
Simuliidae	5	5.8	5.9	5.1	5.5

17.1.3 Original BMWP

This version of the index uses the following family scores.

The data file bmwp.scr must be in the same directory as the data to be analysed. The bmwp.scr file is a text file containing the names and their associated scores.

The standard list is as follows:

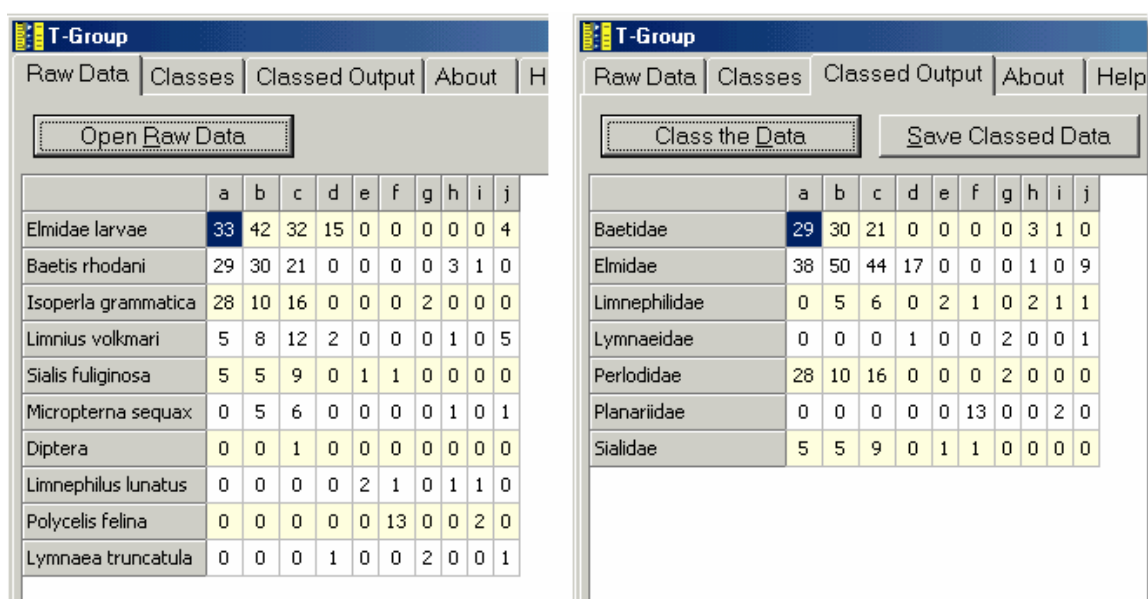
Siphonuridae	10	Corduliidae	8	Dytiscidae	5
Heptageniidae	10	Libellulidae	8	Gyrinidae	5
Leptophlebiidae	10	Psychomyiidae	8	Hydrophilidae	5
Ephemerellidae	10	Philopotamidae	8	Clambudae	5
Potamanthidae	10	Caenidae	7	Helodidae	5
Ephemeridae	10	Nemouridae	7	Dryopidea	5
Taeniopterygidae	10	Rhyacophilidae	7	Elminthidae	5
e	10	Polycentropodidae	7	Chrysomelidae	5
Leuctridae	10	ae	7	Curculionidae	5
Capniidae	10	Limnephilidae	7	Hydropsychidae	5
Perlodidae	10	Neritidae	6	Tipulidae	5
Perlidae	10	Viviparidae	6	Simuliidae	5
Chloroperlidae	10	Ancylidae	6	Planariidae	5
Aphelocheiridae	10	Hydroptilidae	6	Dendrocoelidae	5
Phryganeidae	10	Unionidae	6	Baetidae	5
Molannidae	10	Corophiidae	6	Sialidae	4
Beraeidae	10	Gammaridae	6	Piscicolidae	4
Odontoceridae	10	Platycnemididae	6	Valvatidae	4
Leptoceridae	10	e	6	Hydrobiidae	3
Goeridae	10	Coenagriidae	6	Lymnaeidae	3
Lepidostomatidae	10	Mesoveliidae	5	Physidae	3
ae	10	Hydrometridae	5	Planorbidae	3
Brachycentridae	10	Gerridae	5	Sphaeriidae	3
Sericostomatidae	10	Nepidae	5	Glossiphoniidae	3
e	8	Naucoridae	5	Hirudidae	3
Astacidae	8	Notonectidae	5	Erpobdellidae	3
Lestidae	8	Pleidae	5	Asellidae	3
Agriidae	8	Corixidae	5	Chironomidae	3
Gomphidae	8	Halplidae	5	Oligochaeta	3
Cordulegasteridae	8	Hygrobiidae	5		2
ae					1
Aeshnidae					

17.2 T-Group

To automate the tiresome task of grouping species data into families, T-Group is an absolute necessity. It is particularly useful for creating data for BMWP, ASPT and LIFE scores, and comes with the complete list of British Freshwater families.

An example of the program's output is shown below. T-Group can be downloaded from the Pisces website at:

<http://www.pisces-conservation.com/softutils.html>



17.3 ASPT index

The Average Score per Taxon (ASPT) is simply the [BMWP score](#) divided by the number of scoring taxa in the sample.

This can be calculated for any of the versions of BMWP score.

17.4 Irish water quality rating

To use this freshwater scoring system within *Species Diversity and Richness* undertake the following steps

- Set up a data set listing the all the taxa tabulated in Table 1 which are present in the samples as the rows.
- For each taxon score its abundance in the samples is entered as follows: 5 - well represented or dominant; 4 - common; 3 - present in small numbers; 2 - sparse or absent; 1 - usually absent.
- Once the data has been entered select Irish Q-rating from the special scores button and the value for each sample will be presented.

NB *Species Diversity and Richness* includes an example data set for this scoring system called **IrishQ_test.csv** which can be used and edited to understand how the data is structured.

A biological Assessment scheme suited to Irish rivers was devised by Toner in 1970 and later described by Flanagan and Toner (1972). This scheme, called the Quality Rating System, has been in use since 1971 when the first national survey of river water quality was carried out by An Foras Forbartha (Flanagan and Toner, 1972). To use this method you must give the program observation data arranged as samples by family (see [T-Group](#)) or other appropriate taxon as specified in Table 1 below. It is important that you spell the names correctly as the program looks through the list of names and scores those that correspond with a standard list. *Species Diversity and Richness* includes an example data set for this scoring system called **IrishQ_test.csv** which can be used and edited to understand how the data is structured.

The Quality Rating System relates the relative abundance of five key groups of macro-invertebrates to water quality. The scheme uses five basic water quality classes: Q1 - bad quality, Q2 - Poor quality, Q3 - doubtful quality, Q4 - fair quality, Q5 - good quality. In addition, the intermediate classes Q1-2, Q2-3, Q3-4 and Q4-5 are used to give an effective nine-point scale. *Species Diversity and Richness* reports the Q value as 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5 and 5. The Q value of 1.5 corresponds to the intermediate class Q1-2.

A summary of the scheme is given in Table 2. This table relates the relative-abundance of key macro-invertebrate groups to Q-Value. Note that a distinction is made between eroding and depositing substrata.

While the Quality Rating System is based primarily on macro-invertebrate communities, other factors are also taken into account in assessing Q-Value, for examples the presence and abundance of macrophytes and algae, and the occurrence of slime growths. An overall assessment is also made of the individual sampling site's ecological type and its general suitability for macro-invertebrate communities. These factors may result in a modification to the basic Q-Value derived solely from the decision tables shown in Table 2.

The full assessment of the site, from sampling to ascribing a Q-Value can be completed in the field within 20 to 30 minutes by an experienced biologist. Where possible, preference to be given to riffled sites with turbulent flow conditions and hand-net sampling to be carried out according to ISO Standard 7828-1905. In deeper waters, dredge or grab sampling may be used, but with a few exceptions this is not necessary due to the shallow nature of most Irish rivers. Kick sampling is carried out for 2 to 5 minutes and usually supplemented by a similar period hand picking animals from stones to ensure that those with efficient hold-fast mechanisms are represented in the sample. The sample is decanted and cleared of debris and then transferred to a large white tray on the riverbank for examination. A field inventory of the invertebrates and their relative abundance is then made *in situ*. A general site description is recorded on the field sheet in addition to the macro-invertebrate list. This includes measurements of dissolved oxygen (DO) and temperature, and descriptions of site substratum, water velocity, clarity and river flow conditions. The presence and abundance of macrophytes, algal taxa and 'sewage fungus' are also recorded. Reference faunal samples are retained, but routine laboratory sorting of samples and taxonomic identification of invertebrates to species level is not possible due to constraints of time and resources. Comparison of species lists made in the field with fully sorted and identified laboratory lists has shown that good accuracy can be achieved with field examination of samples, albeit at lower taxonomic detail. Field abundance values are, of necessity, semi-quantitative with each macro-invertebrate taxon being classified into one of 5 categories: 1-5, 6-20, 21-50, 51-100 and > 100

individuals. Again broad agreement has been found between field and laboratory assessments of abundance.

In terms of a programme aimed at assessing water quality in Irish rivers, the Quality Rating System has proved an efficient and reliable technique. The Biological Survey of River Quality in Ireland has been in operation since 1971. There has been very little modification of the scheme in that period. This has allowed long-term trends in water quality to be assessed. Due to the limited resources involved in the biological survey (the survey is carried out by three biologists) it takes four years to complete the national survey - with seriously polluted rivers getting sampled on an annual basis. Because there has not been a comprehensive physico-chemical river survey programme in the past, the biological survey of river quality has assumed an important role in the national reporting of water quality in Irish Rivers.

Table 1, Summary of the 'Irish Quality Rating System: Macro-invertebrate group A, B, C, D and E.

Group A - Sensitive	Group B - Less	Group C - Relatively	Group E - Tolerant	Group D - Most
Forms	Sensitive Forms	Tolerant Forms	Forms	Tolerant Forms
Isoperla	Leuctra	Caenis	Hirudinae	Tubificidae
Perla	Protonemura	Baetis rhodani	Valvatidae	Chironomus
Chloroperla	Amphinemura	Limnephilidae	Hydrobiidae	
Ecdyonurus	Ephemerella	Hydroptilidae	Lymnea	
Rithrogena	Ephemera	Glossosomatidae	Physidae	
Heptagenia	Baetidae	Rhyacophilidae	Planorbidae	
Siphonuridae	Pschomyidae	Philopotamidae	Sphaeriidae	
	Sericostomatidae	Polycentropidae	Asellidae	
	Odontoceridae	Hydropsychidae	Chironomidae	
	Lepidostomatidae	Coenagridae		
	Goeridea	Hemiptera		
	Molannidae	Tricladida		
	Beraeidae	Coleoptera		
	Odonata	Hydracarina		
	Aphelocheirus	Gammaridae		
	Rheotanytarsus	Sialidae		
		Tipulidae		
		Simuliidae		
		Ancylidae		
		Neritidae		
		Viviparidae		
		Astacidae		

Table 2. Decision tables for assigning Q-values based on the relative abundance of these groups at eroding and depositing sites.

Quality Rating	Eroding Sites				
	A	B	C	D	E
Q5	++++	+++	++	+	+-
Q4	++	++++	+++	++	+-
Q3	-	+-	++++	+++	++
Q2	-	-	+-	++++	+++
Q1	-	-	-	+-	++++
Quality Rating	Depositing Sites				
	A	B	C	D	E
Q5	+-	++++	+++	++	+-
Q4	-	++	++++	++	+-
Q3	-	+-	++	+++	++
Q2	-	-	+-	+++	+++
Q1	-	-	-	-	++++

Key:

++++	Well represented or dominant
+++	May be Common
++	May be present in small numbers
+-	Sparse or absent
-	Usually absent

Species Diversity and Richness selects the most appropriate Q rating by comparing the average score for each of the 5 groups of taxa defined in Table 1 against standard tables to find which quality rating which is most similar for eroding and depositing substrates. The similarity measure used is the Euclidean distance.

The standard tables of average scores used by *Species Diversity and Richness* to calculate the Irish quality rating are as follows:

Depositing sites

Rating	Group A	Group B	Group C	Group D	Group E
	0	0	0	0	5
Q1	0	0	0.5	1.5	4
	0	0	1	3	3
	0	0.5	1	3	2
Q1-2	0	1	1	3	1
	0	1	3	2	1
Q2	0	1	5	1	1
	0.5	3	4	1	1
	1	5	3	1	1
Q2-3					
Q3	Group A	Group B	Group C	Group D	Group E
	0	0	0	1	5
	0	0	0.5	3	4
Q3-4	0	0	1	5	3
	0	0.5	3	4	2
Q4	0	1	5	3	1
	0	3	4	2	1
	1	5	3	1	1
Q4-5	3	4	2	1	1
	5	3	1	1	1
Q5					

Eroding sites

Rating

Q1

Q1-2

Q2

Q2-3

Q3

Q3-4

Q4

Q4-5

Q5

17.5 LIFE scores

Lotic-invertebrate Index for Flow Evaluation (LIFE) is a method for linking benthic macroinvertebrate data to prevailing flow regimes. This is an index designed for British Waters and is described in Extence & Balbi (1999). It may be calculated either at the family level or at the species level, and *Species Diversity and Richness* offers both options. Some taxa for which it is impossible to classify their flow preferences or where taxonomic issues make them unreliable do not score. The Pisces utility [T-Group](#) is useful for combining species data into higher taxonomic groups.

The index is calculated by assigning each taxa to one of 6 groups ranging from a group primarily associated with rapid flows to a group holding forms frequently associated with drying or drought impacted sites. Each taxa is then placed in a second category relating to its abundance. There are 5 abundance categories 1-9, 10-99, ..., 10000+. These two values are then used to look up the score in a table. The final index is calculated as follows:

$$\text{LIFE} = \text{Sum}(\text{fs})/n$$

where sum(fs) is the sum of the individual taxon flow scores for the whole sample, and n is the number of taxa used to calculate the sum(fs).

Higher flows should result in higher LIFE scores

Part



18 References

18.1 References

- Bulmer, M. G. 1974 On fitting the Poisson log normal distribution to species abundance data. *Biometrics* **30**, 101-110.
- Burnham, K. P. & Overton, W. S. 1978 Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 623-633.
- Burnham, K. P. & Overton, W. S. 1979 Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60**, 927-936.
- Camargo, J. A. 1993. Must dominance increase with the number of subordinate species in competitive interactions? *J. Theoretical Biology*, **161**, 537-542
- Chao, A. 1984 Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11**, 265-270.
- Chao, A. 1987 Estimating the population size for capture-recapture data with unequal matchability. *Biometrics* **43**, 783-791.
- Chao, A. & Lee, S.-M. 1992 Estimating the number of classes via sample coverage. *Journal of the America Statistical Association* **87**, 210-217.
- Coleman, B. D. 1981 On random placement and species-area relations. *Mathematical Biosciences* **54**, 191-215.
- Coleman, M. D., Mares, M. D., Willig, M. R. & Hsieh, Y.-H. 1982 Randomness, area, and species richness. *Ecology* **63**, 1121-1133.
- Colwell, R. K. & Coddington, J. A. 1994 Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society (Series B)* **345**, 101-118.
- Colwell, R. K. Chang, X, M & Chang, J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, **98**, 2717-2727.
- Extence, C. A., D. M. Balbi, et al. (1999). "River flow indexing using British Benthic macroinvertebrates: a framework for setting hydroecological objectives." *Regulated Rivers: Research and Management* **15**: 543-574.
- Hayek L-A.C & Buzas M.A. (1997) *Surveying Natural Populations*, Columbia University Press, New York, 563pp.
- Heck, K.L., Jr., van Belle, G. & Simberloff, D. 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* **56**, 1459-1461.
- Heip, C. 1974. A new index measuring evenness. *J. Mar. Biol. Ass. UK.*, **54**, 555-557
- Heltsh, J. & Forrester, N. E. 1983 Estimating species richness using the jackknife procedure. *Biometrics* **39**, 1-11.
- Hill, M. O. 1973 Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**, 427-432.
- Holdridge, L. R., Grenke, W. C., Hatheway, W. H., Liang, T. & Tosi, J. A. 1971 *Forest environments in tropical life zones*. Oxford: Pergamon Press.
- Hurlbert, S. H. 1971 The non-concept of species diversity: a critique and alternative parameters. *Ecology* **52**, 577-86.
- James, F. C. & Rathbun, S. 1981 Rarefaction, relative abundance, and diversity of avian communities. *Auk* **98**, 785-800.
- Kempton, R. A. and Taylor, L. R. (1976) Models and statistics for species diversity. *Nature*, **262**, 818-820.
- Kempton, R. A. 1979 Structure of species abundance and measurement of diversity. *Biometrics* **35**, 307-322.
- Krebs, C. J. 1989. *Ecological Methodology*. Harper Collins, New York.
- Lamas, G., Robbins, R. K. & Harvey, D. J. 1991 A preliminary survey of the butterfly fauna of Pakitza, Parque Nacional del Manu, Peru, with an estimate of its species richness. *Publ. Mus. Hist. nat. UNMSM* **40**, 1-19.
- Lambshead PJD, Platt HM & KM Shaw. 1983. Detection of differences among

- assemblages of marine benthic species based on an assessment of dominance and diversity. *J. Nat. Hist.* 17: 859-874.
- Magurran, A. E. 1988 *Ecological diversity and its measurement*. London: Chapman and Hall.
- Magurran, A. E. 2004 *Measuring Ecological Diversity*. Blackwell Publishing, Oxford. 256pp.
- Magurran, A. E. & Henderson, P. A. 2003. Explaining the excess of rare species in natural species abundance distributions. *Nature*, **422**, 714-716.
- May, R. M. 1975 Patterns of species abundance and diversity. In *Ecology and Evolution of Communities* (ed. M. L. D. Cody, J. M.). Cambridge, Mass.: Harvard University Press.
- McIntosh, R. P. 1967 An index of diversity and the relation of certain concepts to diversity. *Ecology* **48**, 1115-1126.
- Miller, R. J. & Wiegert, R. G. 1989 Documenting completeness, species-area relations, and the species-abundance distribution of a regional flora. *Ecology* **70**, 16-22.
- Palmer, M. W. 1991 Estimating species richness: The second-order jackknife reconsidered. *Ecology* **72**, 1512-1513.
- Patil, G. P. & Taillie, C. 1979 An overview of diversity. In *Ecological Diversity in Theory and Practice* (ed. J. F. Grassle, G. P. Patil, W. Smith & C. Taillie), pp. 3-27. Fairland, MD: International Cooperative Publishing House.
- Pielou, E. C. 1975. *Ecological Diversity*. New York, Wiley InterScience.
- Raaijmakers, J. G. W. 1987 Statistical analysis of the Michaelis-Menten equation. *Biometrics* **43**, 793-803.
- Rényi, A. 1961 On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability.*, vol. 1 (ed. J. Neyman), pp. 547-561. Berkley, CA: University of California Press.
- Simpson, E. H. 1949 Measurement of diversity. *Nature* **163**, 688.
- Smith, B. & Wilson, J. B. 1996. A consumer's guide to evenness measures. *Oikos* **76**, 70-82.
- Smith, E. P. & van Belle, G. 1984 Non parametric estimation of species richness. *Biometrics* **40**, 119-129.
- Strong, W. L., 2002 Assessing species abundance unevenness within and between plant communities. *Community Ecology*, **3**, 237-246.
- Soberon, M. J. & Llorente, B. J. 1993 The use of species accumulation functions for the prediction of species richness. *Conservation Biology* **7**, 480-488.
- Solow, A. R. 1993. A simple test for change in community structure. *J. Anim. Ecol.*, **62(1)**, 191-193.
- Southwood, T. R. E. & Henderson, P. A. (2000) *Ecological Methods* 3rd Edition. Blackwell Science 575pp
- Taylor, L. R. 1978 Bates, Williams, Hutchinson- a variety of diversities. In *Diversity in insect faunas: 9th symposium of the Royal Entomological Society* (ed. L. A. Mound & N. Warloff), pp. 1-18. Oxford: Blackwell.
- Tipper, J. C. 1979 Rarefaction and rarefaction - the use and abuse of a method in paleoecology. *Paleobiology* **5**, 423-434.
- Tothmeresz, B. 1995 Comparison of different methods for diversity ordering. *Journal of Vegetation Science* **6**, 283-290.
- Williams, C. B. 1964 *Patterns in the balance of nature*. London: Academic Press.
- Wilson, M. V. & Shmida, A. (1984). Measuring beta diversity with presence/absence data. *J. Ecol.* **72**, 1055-1064.

Part



19 Contact Pisces

19.1 Contact Pisces

For most active windows, context sensitive help can be obtained by pressing F1, clicking on the help button or selecting the help drop down menu. or clicking on the right-hand mouse button and choosing help from the pop-up menu.
If pressing F1, make sure that the window that you are seeking help for is the active one.

If you have problems using the program or entering data which you cannot solve then contact Pisces Conservation Ltd by e-mailing pisces@pisces-conservation.com or by phone to UK (0)1590 674000 during office hours (09.00 to 17.00).

PISCES Conservation Ltd,
IRC House, The Square
Pennington, Lymington
Hants, SO41 8GN,
UK

Telephone	44 (0) 1590 674000
Fax	44 (0) 1590 675599

For more information, details of our other software, and answers to a range of technical queries, visit our web site at <http://www.pisces-conservation.com>

For details about our consultancy and research work, visit <http://www.irchouse.demon.co.uk>

Index

- A -

abundance 62, 72, 80, 81, 87, 89, 92, 93, 94, 95
accumulation 87, 89
ACE 72
add 32
adjustment 77, 80
alpha 40, 42, 43, 44, 45, 46, 47, 49, 50
altering 35
ASPT 104
authors 116

- B -

Berger 40
best choice index 55
beta 101
beta indices 55
bibliography 116
bitmap 36
BMWP 104, 106, 107, 109
bootstrap 48
bootstrap confidence intervals 48
brillouin 40, 45, 56
Bristol channel 11
broken stick 92, 95
butterfly 11

- C -

Carmargo 57
CD 7
change 35
chao 69
citation 9
cite 9
close 38
collection 45
columns 29
compare 50
comparing 20, 23, 65, 66
comparison 66, 80
comparison of richness 78, 81
comparisons 66
complement 69

complete 45
confidence 48
contact 119
contact pisces 119
copy 36
Cotgreave 57
coverage 72, 73
create data file 30
cross 34
cumulative 89
curve 87

- D -

data 28, 29, 35
data file 30
data sets 11
data type 30
dataset 29
demo 11, 28
demonstartions 11
Demonstration 11, 12, 15, 17, 20
demonstration data 28
deselect 34
Dispersion 23
distribution 92, 93, 94, 95
distributions 92
diversity 41, 45, 47, 49, 50, 65, 101
diversity ordering 12, 66
dominance 44, 47, 87
dummy data 98

- E -

edit 32
editing 35
effort 77, 80
effort adjustment 78, 81
End 38
equitability 55, 56, 62
equitability J 55
estimator 71, 72, 73, 74, 75, 101
evenness 54, 55, 56, 57, 58, 59
Excel 28, 30
export 36, 38

- F -

file 32, 38
file menu 29

Fisher 40, 93
fit 62, 95
fitting 92, 93
freshwater 104, 106, 107, 109, 110, 114

- G -

general 27
General statistics 23
geometric 92
getting data 32
Gini 59
graphs 36
grid 28

- H -

Harvey 57
Heip 56
help 119
heterogeneity 69
hinkley 11
how to 7, 27

- I -

ICE 73
Incidence 73
increase in species with effort 89
index 41, 42, 43, 44, 45, 46, 47, 49, 50, 55, 56, 57, 58, 59, 101, 104, 106, 107, 109, 110
individuals 38
information 41, 62
Input 28
install 7
instructions 27
integer 35
interpolation 81
intervals 48
introduction 2
Irish 104

- J -

Jackknife 69
Jackknifing 49, 50

- K -

kurtosis 23

- L -

Lee 69
LIFE 104
log 93, 94
log normal 92
log series 92
lotus 30

- M -

maximum 23, 29, 38, 58, 76
mcintosh 40, 55
mean 23
median 23
metafile 36
Michaelis Menten 69
michaelis-menten 85
minimum 23, 58

- N -

Nee 57
new 29
NHC 57
non-integer 35
non-random 45
normal 94
normality 94
number 40, 43, 71, 72, 73, 74, 75, 87, 101
number of samples 85

- O -

observation size 38
opening 32
operating system 5
order 87
ordering 65
output 36

- P -

papers 116
parametric models 85
Parker 40
paste 36
pisces conservation 119
print 36

printing 36

- Q -

Q 40, 46

- R -

RAM 5

random 84

randomisation test 50

randomise 84

randomize 84

range 23

rank 17, 87, 89

rank abundance 11, 59

rarefaction 69, 77, 78, 80, 81

rarefraction 77, 78, 80

rate of migration 82

real 35

reference 9

references 116

refs 116

regression 57

relative 87

relative abundance 85

renyi 12, 65, 66

Requirements 5

Richness 62, 69, 71, 72, 73, 74, 75, 76, 101

right-tailed 65

rows 29

runs 84

- S -

sample heterogeneity 70

sample number 85

Sample statistics 23

samples 29

save 38

score 104, 106, 107, 109, 110

select 34

selection 34

Shannon 40, 41, 55, 56, 58, 62

SHE 101

sheet 30

Significance 50

Simpson 40, 42, 56, 58

simulate 98

simulated data 98

simulation 98

size 29

skew 23

Slope 57

Smith and Wilson 57, 58

smooth 84

Solow 50

sort 23

special 114

species 29, 43, 87

species accumulation 11, 15, 47

species distribution 92

species effort curves 89

species in common 23

species not in common 23

species number 76, 87

Species statistics 23

spread 30

spreadsheet 30

starting 29, 32

statistical 50

stats 50

stopping rule 85

storage 5

support 119

- T -

tech support 119

test data 98

Testing for differences 20

tick 34

tourist species 82

transect 101

tropical 11

turnover 101

tutorial 11, 12, 15, 17, 20

type 35

- V -

variance 23

versions 2

- W -

Weaver 41

whats new 2

Wiener 41, 55

Endnotes 2... (after index)

Back Cover