



Copyright 2007, PISCES Conservation Ltd



ECOM II

Understanding relationships in nature

by Pisces Conservation Ltd

ECOM is designed to help you identify and understand the relationships that often exist between the physical and chemical variables present in the environment and the occurrence and abundance of species.

It offers a wide range of tried and tested techniques to handle and analyse the complex data that ecological sampling often generates.

Using the latest data handling techniques, ECOM II 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.

ECOM II

Copyright 2007, PISCES Conservation Ltd

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

Piscas Conservation Ltd.

Table of Contents

Foreword	I
Part I Introduction	2
Part II Improvements to Version 2.1 of ECOM	4
Part III General instructions	6
Part IV Quick guide to running a data set	9
Part V Installation and requirements	12
1 System requirements	12
2 Methods offered by ECOM	12
3 Installation	12
4 User preferences	12
Part VI Importing from Excel	15
Part VII Demonstration data sets	17
Part VIII File menu	19
Part IX Opening a data set	21
Part X Creating and editing a data set	24
1 Preparing large data sets in a spreadsheet program	24
2 Data entry from within ECOM	25
3 Editing existing data	27
4 Saving edited data or creating a new data file	27
5 Large data files	27
6 Maximum size of the data set and computation speed	27
Part XI Transforming and adjusting your data	29
1 Data transformations	29
2 Data relativisations	30
3 Dealing with zeros	31
4 Transposing data	32
Part XII The output screens	34
1 Summary of data	34
2 Statistics radio buttons	35
3 Sample Correspondence Check	36

4 Environmental variables from biological features	37
5 Data set statistics	37
Part XIII Groups	40
Part XIV Canonical Correspondence Analysis	46
1 What is CCA?	46
2 Selecting Environmental variables	47
3 Linear combinations of environmental variables	48
4 Dummy Environmental variables	48
5 Transforming environmental variables	49
6 Multicollinearity	49
7 Circular environmental variables	50
8 Species Scores	51
9 Variance & Eigenvalues	51
10 Sample Scores	53
11 Ordination Plot	54
12 Rank Plot	55
13 Correlation of Env. Vars	56
14 Binary Environments	56
15 Biplot Scores	57
16 Monte Carlo Simulations	58
17 Environment plot	59
18 Canonical Coef.	59
Part XV Multiple Regression	62
1 Scatter Plot	64
Part XVI Stepwise Multiple Linear Regression	66
1 Variable Selection	66
2 Foward Stepwise Linear Regression	67
3 Backward Stepwise Linear Regression	69
Part XVII Variable Selection	72
1 Sample/Species Data	72
2 Working Samples/Species Data	73
3 Environmental Data	74
4 Working Environmental Data	74
5 Redundancy Analysis	75
6 Correspondence Analysis	75
Part XVIII Printing and saving results	78
1 Exporting charts	78

2	Printing charts	79
3	Preparing charts for output	79
4	Exporting text and grid output	83
5	Graph display options	84
6	Tips for improving graphical output	85
7	Zooming and panning on graphs	85
Part XIX	Printing text and grid output	88
Part XX	Programs for ecologists by PISCES Conservation Ltd	91
1	CAP	92
2	ECOM	93
3	Species Diversity and Richness IV	95
4	Axis	97
5	Fuzzy Grouping	98
6	Simply Tagging	100
7	Population Estimation by Removal Sampling	101
8	Density from Distance	102
9	Growth II (Simply Growth)	103
10	Dynamica	105
11	Hedgerow Assistant	106
12	Simply Probit	107
Part XXI	Lecture Resources	109
1	Seashore Ecology	109
2	The Ecology of Freshwaters	110
3	The Galapagos Islands	111
4	An Introduction to Multivariate Statistics	112
5	An Introduction to Species Diversity and Richness	112
Part XXII	E-books	115
1	Entomology	115
2	Aquatic Biology	117
3	Botany	118
4	Classics of Science and Exploration	119
5	Pisces Thesis Series	120
Part XXIII	Worked Example	122
1	Preliminary inspection and data transformation	122
2	Examine and Select Environmental Variables	124
3	Interpreting the results	126
	Problems with Species-Environment Correlation	128
Part XXIV	Obtaining help	131

1 Common errors and problems	131
2 References	133
3 Demonstration version	133
4 Citation	133
5 Contacting PISCES for help	133
 Index	 135

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 LIMITED NOT 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.

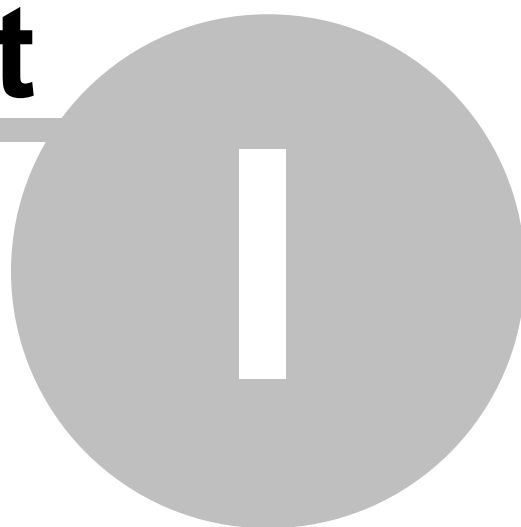
NO LIABILITY FOR CONSEQUENTIAL DAMAGES - TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW PISCES AND ITS SUPPLIERS SHALL NOT BE LIABLE FOR ANY OTHER DAMAGES WHATSOEVER (INCLUDING, WITHOUT LIMITATION, DAMAGES FOR LOSS OF BUSINESS PROFITS, BUSINESS INTERRUPTION, LOSS OF BUSINESS INFORMATION, OR OTHER 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
Tel +44 (0)1590 674000

Part



1 Introduction

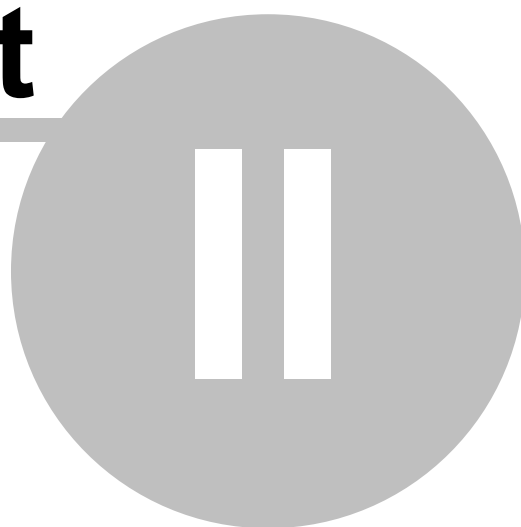
ECOM offers a range of analytical techniques for visualising and ordering the relationships in multivariate data commonly used by community ecologists including Canonical Correspondence Analysis (CCA), Redundancy Analysis (RA) and Multiple Regression (MR). Researchers in other fields such as palaeontology, archaeology and the social sciences also use many of these methods. At the core of all the methods is the exploration and quantification of the effects of the explanatory or environmental variables on the observed dependent variables. ECOM is focused on the analysis of ecological data so the dependent variables are termed biological variables and would typically be an array of species abundances in different samples. The environmental data set comprises observations of the explanatory variables for these samples and would typically hold information on physical variables such as temperature, salinity, altitude, nutrient status etc. However, it may be that the independent variables are themselves biological variables. For example, the insect community might be determined by a variables that measures the density of the tree canopy.

Data can be organized using Windows programs such as Microsoft Excel and the output from ECOM is displayed, exported and printed using standard Windows techniques. The result is a program that is easily learned and used by both students and professional ecologists. It is particularly useful for ecological teaching because it allows students to quickly enter data, try different transformations and explore a range of methods within a familiar Windows setting.

The user supplies two input data sets holding the biological and physical data for the samples (sites) respectively. These data are arranged as two-dimensional arrays. In ecology, it is usual for the samples, which are normally collected from set localities and may be called, for example, quadrats or stations, to form the columns. The variables for each sample are the rows and, in the case of the biological data, comprise the numbers of each species or other taxon observed. The environmental data may be either continuous variables such as pH, temperature or current speed, or binary variables such as soil or treatment type which are scored as either a 0 or 1. ECOM complements CAP (Community Analysis Package), also produced by PISCES Conservation, which offers a range of techniques that only require biological data to undertake the ordination or classification.

ECOM uses the same data structure as CAP and Species Diversity and Richness II, which calculates a wide range of diversity and species richness measures. Together, ECOM, CAP and Species Diversity and Richness II offer an extensive range of methods for the analysis of ecological communities.

Part

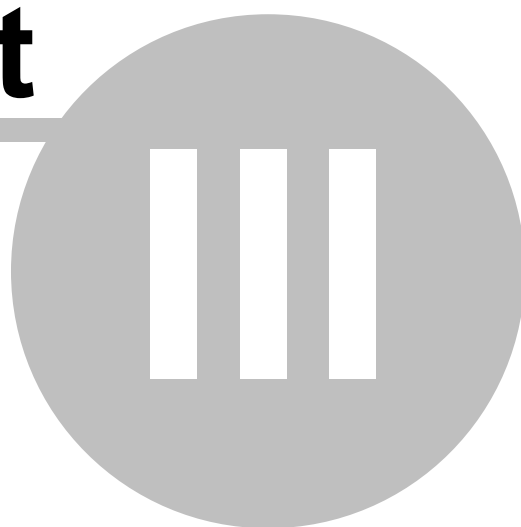


2 Improvements to Version 2.1 of ECOM

A number of significant improvements were incorporated into Version 2. Most importantly, new array handling technology was built into the code to enable data matrices of virtually any size to be handled. Calculation speed was also improved and even very large data sets can be handled with ease. The range of methods available was also increased, with the introduction of Forwards and Backwards Stepwise regression.

In version 2.1, the ability to group samples has been added, and on a technical note, the way the program stores settings has been altered so that it can work over a network of PCs, and to ensure compatibility with Windows Vista.

Part



3 General instructions

Start ECOM in the normal way, either by clicking on the desktop icon or by selecting the program via the Start button.

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

File: To create, open, save, export and print data.

Biological File: To select the file to open for the biological data.

Environmental File: To select the file to open for the environmental data.

Edit: To cut copy and paste to and from the active window and set the decimal points and precision of the display. Use this menu to assign samples to groups.

Ordination: To select the ordination methods

Variable Selection: To run a Redundancy or Correspondence analysis for selection of the most influential variables.

Monte Carlo: To run a Monte Carlo randomization on the data.

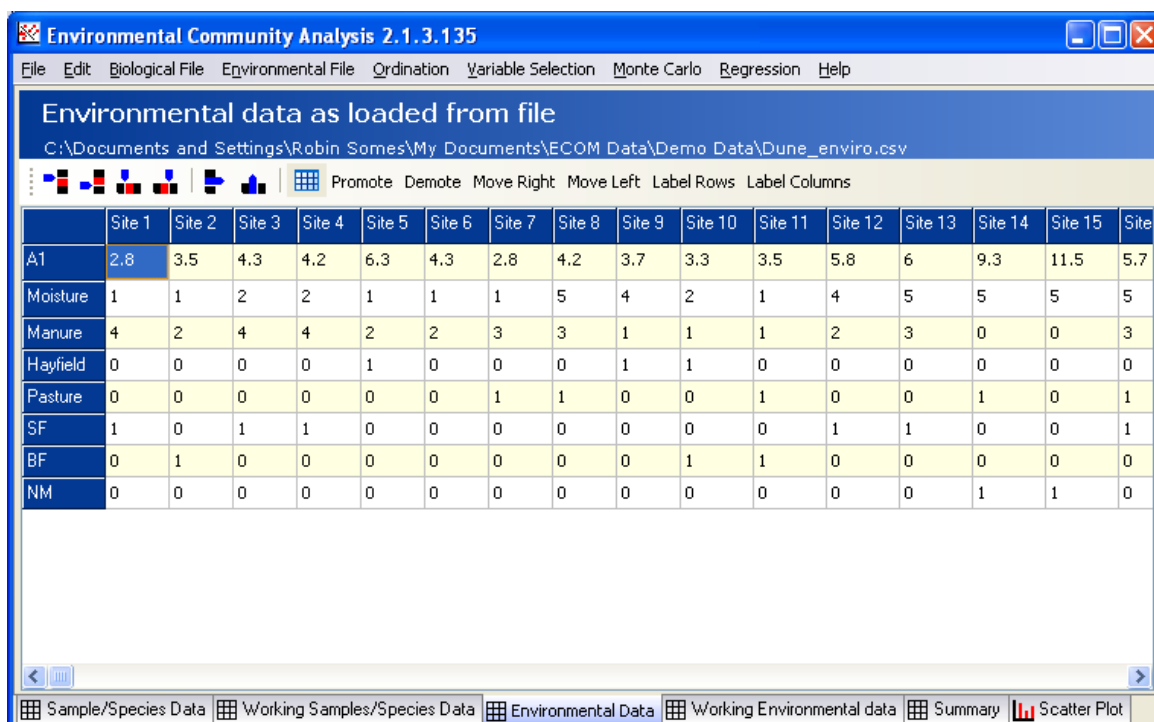
Regression: To select the multiple regression method.

Help: to enter the Help system and select the about box.



Note: the Monte Carlo menu is disabled until you perform an Ordination analysis.

When the program is started, you will be presented with a blank working area. Assuming that you have data already prepared, from the Biological File menu choose Open Biological and from Environmental File menu choose Open Environmental to load existing data sets. Once a data set is loaded or a new data set created, the window will be filled by a form with tabs labelled Sample/Species Data, Working Sample/Species Data, Environmental Data, Working Environmental Data, Summary and Scatter Plot.



The Raw Data forms, Sample/Species Data and Environmental Data show the data loaded from disc. The Working Data grids show the data sets upon which the calculations will be performed, Summary gives statistics for the data sets and Scatter Plot lets you select variables to plot as 2 and 3 dimensional graphs. You can switch between the different forms by clicking with the mouse on the tabs. Initially, the Raw Data and Working Data will display the same values. Once a transformation or some other adjustment has been applied to the data (for details see below) the

Working Data grids will display the adjusted data.

[Obtaining help](#)

[Opening a data set](#)

[Demonstration data sets](#)

[Creating and editing a data set](#)

[Maximum size of the data set and computation speed](#)

[Printing and exporting your results](#)

[Graph display options](#)

[Zooming and panning on graphs](#)

Part



IV

4 Quick guide to running a data set

1. Prepare the biological data set using a spreadsheet such as Microsoft Excel. The data are arranged with samples (quadrats, or stations) as columns and species (variables) as the rows. Give a title to each sample and species. Values can be real or integer and place 0 (a zero) in all localities where a species was not recorded. This can be automated by searching for blanks in Excel and making them 0. Make sure you use a zero and not a capital '0'!
2. Save this array as a .csv (comma delimited text) file from the Save As menu in Excel. Take care if using Excel that the work sheet does not hold any other data or this will be included in the csv file. ECOM can open Excel (xls) files, but the native format used is a csv file.
3. Prepare and save the environmental (physical) data set in a similar fashion to the biological data. Remember to give it a different name from that used for the biological data set.
4. Close the spreadsheet data file before opening the data in ECOM - otherwise you will get an I/O error 32 - access denied.
5. Run ECOM from the Start menu or by clicking on the executable file.
6. From the Biological File drop-down menu select **Open Biological** and choose your biological data file.
7. From the Environmental File drop-down menu select **Open Environmental** and choose the data file holding the environmental (physical) data. These are the independent variables.
8. The data will now open into the grids that can be seen by clicking on the tabs labeled **Sample/Species Data** and **Environmental Data** respectively. Before choosing a method it is wise to check your data for columns that sum to zero (samples with no species) and rows that sum to zero (species not found in any sample). First click on the **Working Sample/Species Data** tab. To remove zero sum rows or columns, select **Handling zeroes** - Delete 0 rows or Delete 0 columns in the Type of adjustment panel below the data. Once an adjustment has been made remember to click on the Submit button to create a data set for analysis.
9. Undertake any transformations or relativisations you may wish in the working data tabs for either the biological or environmental data. Note that any changes undertaken on the working data will not change the raw data.
10. From the drop down menus select the method you wish to use and, using the tabbed windows, examine the output.
11. Use **File|Print** to print output and **Edit|Copy** to copy a graph to a file or the clipboard.

Environmental Community Analysis 2.01

File Edit Biological File Environmental File Ordination Variable Selection Monte Carlo Regression Help

Sample - species data as used by the program

	h1	h2	h3	h4	h5	h6	h7
Hooknose (Pogge)	0	0	1	0	0	0	0
Shad Twaite	0	0	0	0	0	0	0
Sand eel Common	0	0	0	0	0	0	2
Eel Common	1	6	30	2	0	0	0
Goby Transparent	0	0	0	0	1	1	0
Sand smelt	0	0	0	0	0	0	0
Trigger Fish	0	0	0	0	0	0	0
Blenny	0	0	0	0	0	0	0
Drum	0	0	0	0	0	0	0
Rockling	0	0	0	0	0	0	0
Rockling North	0	0	0	0	0	3	0

Type of Adjustment

- ☒ Transform
- ☐ Relative
- ☐ Handling zeros
- ☐ Transpose

Reload Raw Data

- Log 10
- Log 10 (x+1)
- Log e
- Log e (x+1)
- Square Root

Selected Variable

-All-

Buttons: Reload Raw Data, Submit

Bottom Panel:

- Sample/Species Data
- Working Samples/Species Data
- Environmental Data
- Working Environmental data
- Summary
- Scatter Plot

Annotations:

- The variables (species) form the rows.
- The samples are arranged in columns.
- Choose a type of adjustment to your working data here. This will not permanently change your raw data.
- Select a particular transformation from the list here.
- You can choose which variables to transform here.
- Use this button to revert to your raw data.
- Click on submit to implement the desired transformation.

Part



V

5 Installation and requirements

[System requirements](#)
[Methods offered by ECOM](#)
[Installation](#)

5.1 System requirements

ECOM runs on Windows XP and Vista operating systems. It is likely that the program will run perfectly on older versions of Windows, but unfortunately we are no longer able to guarantee this.

Pentium PC running Windows XP or Vista
Minimum of 256 MB of RAM.
8 MB of hard disk space for installation

ECOM does not limit the size of your data set, however your hardware will. To use ECOM with very large data sets (1000 or more species or samples) you will need a fast modern machine with 256 MB or more of RAM memory.

5.2 Methods offered by ECOM

ECOM offers a range of related techniques for analysing the structure of communities and the distribution of individual species in relation to environmental variables. These are multiple regression, Redundancy Analysis (RA) and Canonical Correspondence Analysis (CCA). For more information about these techniques see [Canonical Correspondence Analysis](#), [Multiple Regression Forward](#) and [Backward Stepwise Regression](#) and [Redundancy Analysis](#). The program also offers a range of simple statistics and a scatter plotting option.

There are a number of different ways in which the output from Redundancy Analysis and Canonical Correspondence Analysis can be scaled for plotting. The output from RA can be chosen to scale to produce either a distance or correlation biplot, while for CCA it can be scaled with sites at the species centroids or vice versa.

5.3 Installation

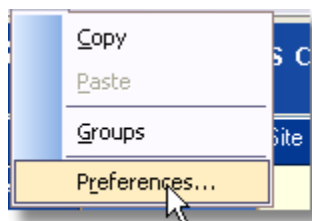
1. Place the ECOM CD in your CD drive: the installation process should begin automatically - follow the on-screen instructions.
2. If the CD does not auto-play, browse the CD in Windows Explorer or My Documents and double-click file named Setup.exe in the root directory.

When installation is complete there will be a ECOM group on your Start button/Programs menu, and a folder (directory) on your hard disk which holds the ECOM program files. The [demonstration data files](#) will be placed in **My Documents / Ecom Data / Demo Data**. An uninstall program will also be created, in case you wish to remove the program.

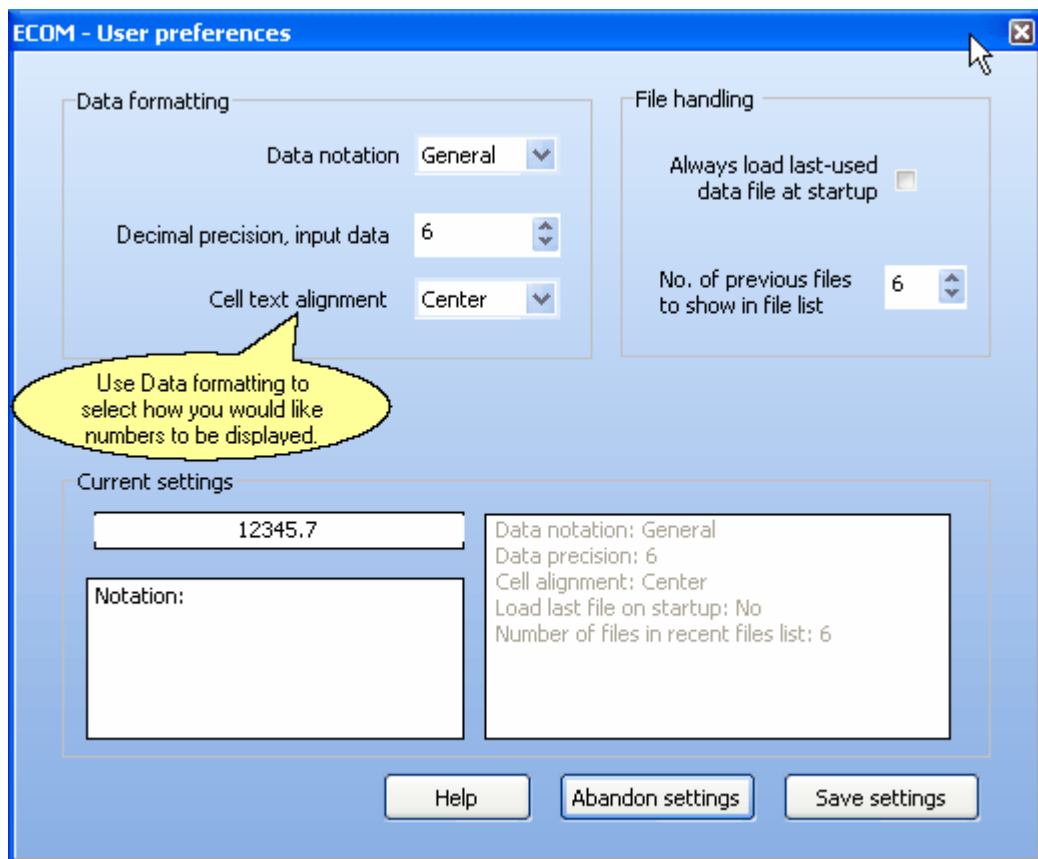
5.4 User preferences

ECOM 2.1 now gives you control over the display and handling of your data.

From the Edit menu, select Preferences.



This will open the User preferences screen -



Settings for data notation, precision and alignment can be set in the data formatting section. When you make a change to the data notation or decimal precision, these changes will take effect on the next time the program is started up. The options to reload the last file you were using whenever the program starts up and also to set the number of files visible in the file lists can be set in the File handling group. Note: if you close all files using the File | Close All menu options before exiting the program then no file names will be saved and the program will open with a blank screen next time. If, however you exit with data files still visible then their names will be saved for the next opening of the program.

Part



VI

6 Importing from Excel

ECOM offers the ability to load data directly from a Microsoft Excel spreadsheet.

It is important, however, that a number of points are observed:

1. If your Excel spreadsheet has multiple worksheets, the data will be imported from the worksheet that was visible when the workbook 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 within the contiguous block, rightwards and downwards from cell B2 contain numerical data.
6. ECOM can import directly from Excel whether Excel is open or not, provided the worksheet has been saved.

Part



VII

7 Demonstration data sets

Three demonstration data sets are placed on your hard disc during installation; by default they are stored in **My Documents / Ecom Data / Demo Data**. Each consists of a pair of files which must both be opened to run the analyses.

Demonstration 1

Dune_species.csv and Dune_enviro.csv hold the dune meadow data used by Jongman et al.

Demonstration 2

Test_species.csv and Test_enviro.csv hold the demonstration data used by Legendre & Legendre

Demonstration 3

Hinkley_species.csv and Hinkley_enviro.csv hold data on fish and climatic conditions at Hinkley Point in the Severn Estuary.

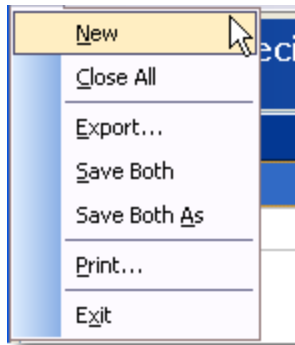
These demonstration data sets allow the user to test the program and, by opening these files in Excel or another spreadsheet, see how the data are organized. All demonstration files can be deleted without affecting the program.

Part



8 File menu

The File menu contains options as follows:



New : [Create a new data](#) grid within ECOM into which you can enter your data.

Close all : To close both the biological and the environmental datasets. If you have made changes to either of the raw data grids, you will be prompted to save them before closing.

Export : To [export](#) the currently active grid.

Save Both : To save both the biological and the environmental datasets under their current names, whether or not changes have been made to them.

Save Both As : To save both the biological and the environmental datasets under new names, with the option of saving as a csv file or an xls file.

Print : gives access to a wide range of [printing options](#).

Exit : closes the program.

Part

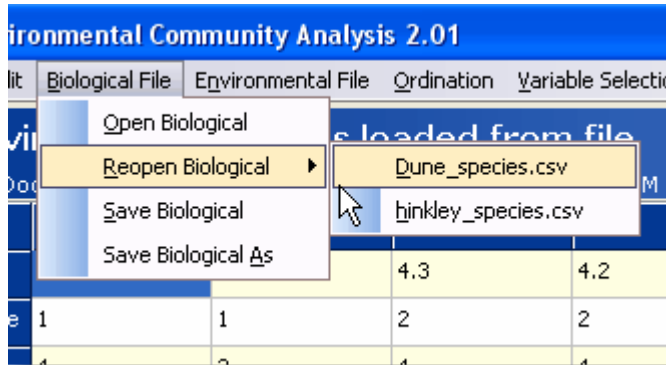


IX

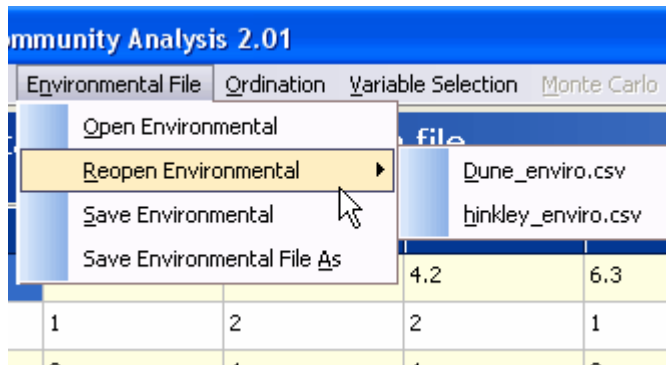
9 Opening a data set

To use ECOM you must open two data sets. One will hold the biological data and the second the environmental (physical) variables.

Use **Biological File|Open Biological** or **Biological File|Reopen Biological**



and **Environmental File|Open Environmental** or **Environmental File|Reopen Environmental**



to open the file dialogs to select the data files for analysis. The default file type for ECOM files is *.CSV - which is the standard file extension for comma delimited files.

Once a file has been selected the screen will appear as follows:

Environmental Community Analysis 2.01

File Edit Biological File Environmental File Ordination Variable Selection Monte Carlo Regression Help

Sample - species data as loaded from file
C:\Documents and Settings\Peter Henderson\Desktop\ECOM 2007\Data\Dune_species.csv

	site 1	site 2	site 3	site 4	site 5	site 6	site 7	site 8
Achmil	1	3	0	0	2	2	2	0
Agresto	0	0	4	8	0	0	0	4
Airpra	0	0	0	0	0	0	0	0
Alogen	0	2	7	2	0	0	0	5
Antodo	0	0	0	0	4	3	2	0
Belper	0	3	2	2	2	0	0	0
Brohor	0	4	0	3	2	0	2	0
Chealb	0	0	0	0	0	0	0	0
Cirarv	0	0	0	2	0	0	0	0
Elepal	0	0	0	0	0	0	0	4
Elyrep	4	4	4	4	4	0	0	0
Emprig	0	0	0	0	0	0	0	0
Hyprad	0	0	0	0	0	0	0	0
Junart	0	0	0	0	0	0	0	4
Junbuf	0	0	0	0	0	0	2	0
Leoaut	0	5	2	2	3	3	3	3
Lolper	7	5	6	5	2	6	6	4
Plalan	0	0	0	0	5	5	5	0
Poapra	4	4	5	4	2	3	4	4
Poatri	2	7	6	5	6	4	5	4

Sample/Species Data Working Samples/Species Data Environmental Data Working Environmental data Summary Scatter Plot

Initial screen showing raw data for the Dune_species.csv file supplied with ECOM. Species names are shown in the first column and sample code names, site 1 onwards are the column titles. At the bottom are the 6 tabs used to select the Raw Data grids, the Working Data grids, data set summary statistics and scatter plot options.

Part



10 Creating and editing a data set

You are recommended to prepare large data sets using a standard spreadsheet program such as Excel. Small data sets can be created directly within ECOM, but editing facilities are limited. Both the biological and physical data are organised in the same way and stored as a CSV, comma delimited file.

[Preparing large data sets in a spreadsheet program](#)

[Data entry from within ECOM](#)

[Editing existing data](#)

[Saving edited data or creating a new data file](#)

[Large data files](#)

[Maximum size of the data and computation speed](#)

10.1 Preparing large data sets in a spreadsheet program

You do not need to create data sets within ECOM; in fact the best way to enter large data sets is to use a spreadsheet such as Excel or Lotus 1-2-3, which will give access to a wide range of sorting and editing procedures to ease your task.

A data set should have the following format:

	Sample 1	Sample 2	Sample 3
Variable 1	21	1	5
Variable 2	15	5	0
Variable 3	0	7	0

The normal arrangement of community data within ECOM is to have the samples (quadrates) 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, the data can be arranged in Excel with the variables forming the columns, as the Transpose option within ECOM can be used to switch columns and rows. Numbers can be either integer or real; some methods may require integers, in most such cases the program will run with real data which will be automatically rounded.

The above table and the image below 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 (variable names) are input from row 2 in column 1.

	A	B	C	D	E	F	G	H	I	J
1		site 1	site 2	site 3	site 4	site 5	site 6	site 7	site 8	site 9
2	Achmil	1	3	0	0	2	2	2	0	
3	Agrsto	0	0	4	8	0	0	0	4	
4	Airpra	0	0	0	0	0	0	0	0	
5	Alogen	0	2	7	2	0	0	0	5	
6	Antodo	0	0	0	0	4	3	2	0	
7	Belper	0	3	2	2	2	0	0	0	
8	Brohor	0	4	0	3	2	0	2	0	
9	Chealb	0	0	0	0	0	0	0	0	
10	Cirarv	0	0	0	2	0	0	0	0	
11	Elepal	0	0	0	0	0	0	0	4	
12	Elyrep	4	4	4	4	4	0	0	0	
13	Emprig	0	0	0	0	0	0	0	0	
14	Hyprad	0	0	0	0	0	0	0	0	
15	Junart	0	0	0	0	0	0	0	4	
16	Junbuf	0	0	0	0	0	0	2	0	
17	Logout	0	5	2	2	3	2	2	2	

When using Excel use the **Save As** function to save your data as a *.csv file. This will result in a data file in the format used by ECOM. Alternatively, you can save as an Excel file which can be [imported into ECOM](#). 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 blank cells and replace them with 0 (zero).

Occasionally, errors occur because a blank space or a character 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 csv file can also be opened and edited in a text editor (e.g. Notepad) or word processor (Microsoft Word), in which case it will appear like this:

,Title1,Title2,Title3

Variable 1,21,1,5

Variable 2,15,5,0

Variable 3,0,7,0

Variable 4,1,9,0

Variable 5,0,0,8

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

see also:

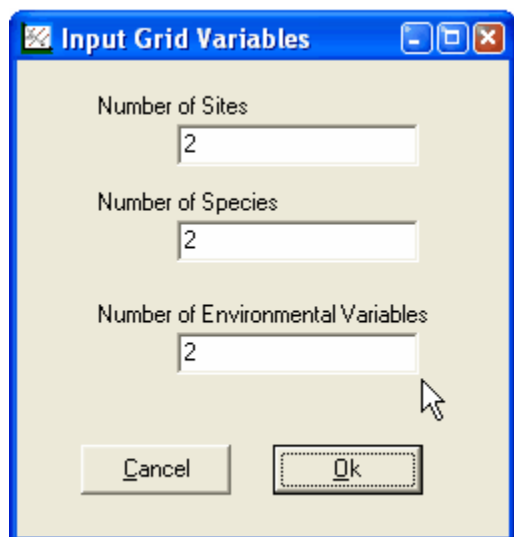
[Demonstration data sets](#)

[Opening a data set](#)

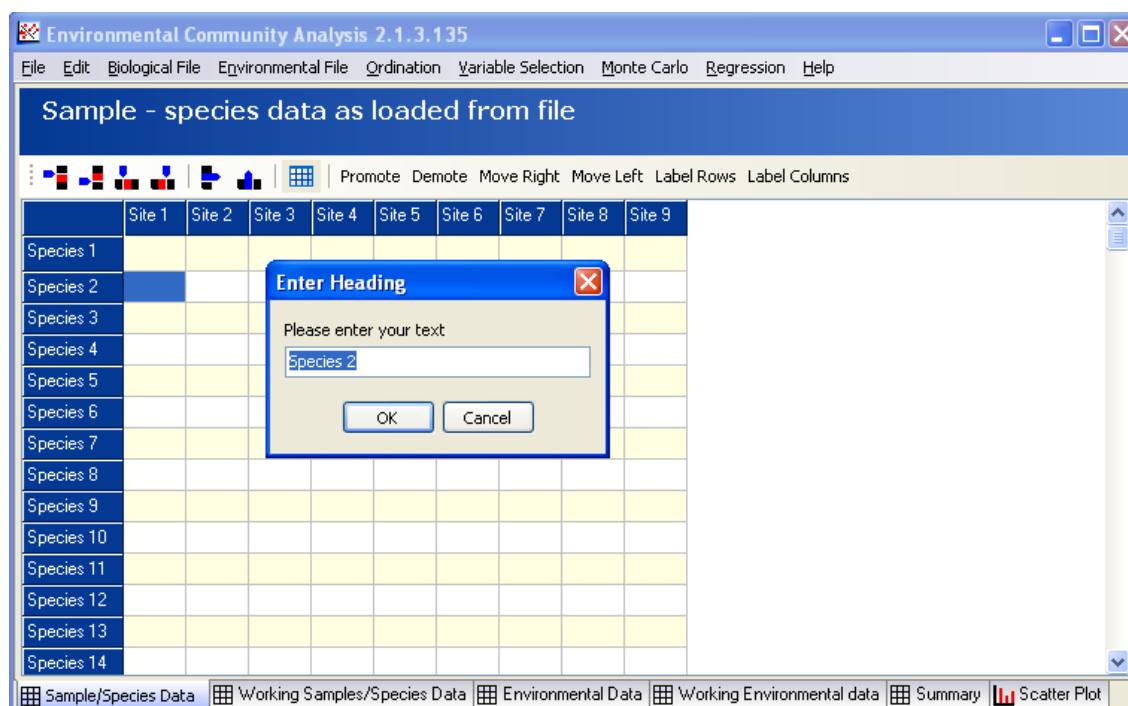
[Data entry from within ECOM](#)

10.2 Data entry from within ECOM

Data sets can be created and edited within ECOM. To create a new data set select **File| New** from the pull-down menus. A dialog window called Input Grid Variables will open with which you specify the number of sites, species and environmental variables. After you have entered the numbers, click on OK and the data grids will be created for data entry.



After the new file is opened, if you entered numbers as shown above, you will be presented with a 3 column by 3 row grid. The blue grid cells at the head of the columns starting from column 2 are used to hold sample names and similarly the first column starting from row 2 is used to hold species names. Double-click on each of these cells to show a dialog into which you can enter species names, site names and environmental variable names.



Leave the upper left cell free.

To enter data into the remaining cells in the grid, click on a cell and type a number. Numbers can be either integer or real, but some methods may require integers; in most such cases the program will run with real data which will be automatically rounded.

The return key moves you sequentially through the grid. The buttons on the toolbar above the data grid can be used to:

- Add Row Above
- Add Row Below
- Add Column to Left
- Add Column to Right

Delete Row
Delete Column
Create a grid of X columns by Y rows
Promote / Demote - move a row up to become the Header row of the grid, or move the header to become an ordinary row.
Label Rows / Columns - with standard Species / Site labels

10.3 Editing existing data

The raw data grid can be edited by using the mouse to click on a cell to select it and typing in a new value. Changes made to the raw data are transferred to the working data by first selecting the working data tab and then selecting Revert to raw and clicking on the **Submit** button. These changes will not alter a saved file until **File|Save** is used. ECOM will use the working grid thus created for all subsequent calculations. Note that output windows will not show calculations using the edited data until the methods have been re-run by selecting them in the normal fashion. The working data grid can also be edited. Changes made to the working grid will not be transferred to the raw data grid. Transformed or otherwise changed working data can be saved as a new data set using **File|Save as**.

10.4 Saving edited data or creating a new data file

When **File|Exit** is selected, if a data set has been altered in any way, but not saved, you will be asked if you would like to save the data. A data set can be copied and saved under a different name by selecting **File|Save as**.

10.5 Large data files

If you are planning to use large datasets with ECOM then you may find the PISCES LIST COMBINER very useful indeed. Excel has a limit of 255 columns per worksheet. LIST COMBINER will take all your worksheets and combine them all into one large csv file (summing and sorting the values for each variable along the way) ready to load directly into ECOM. See www.piscs-conservation.com/softutils.html for more details.

10.6 Maximum size of the data set and computation speed

In ECOM 2.1 new array handling technology enables the analysis of very large data sets.

Theoretically, the size of the input data matrix is unlimited although, in practice, the memory resources of the PC is usually the limiting factor. In tests we have run ECOM with a data array of 5000 species x 5000 samples without difficulty.

Our new approach to data handling means that on a reasonably modern PC, most of the computation, even on very large data sets, can be completed within a few seconds. All computations with data sets of up to 100 samples by 100 species are usually almost instantaneous.

If you do run into speed problems with very large data sets, it will generally be improved if there are no other Windows applications running at the time.

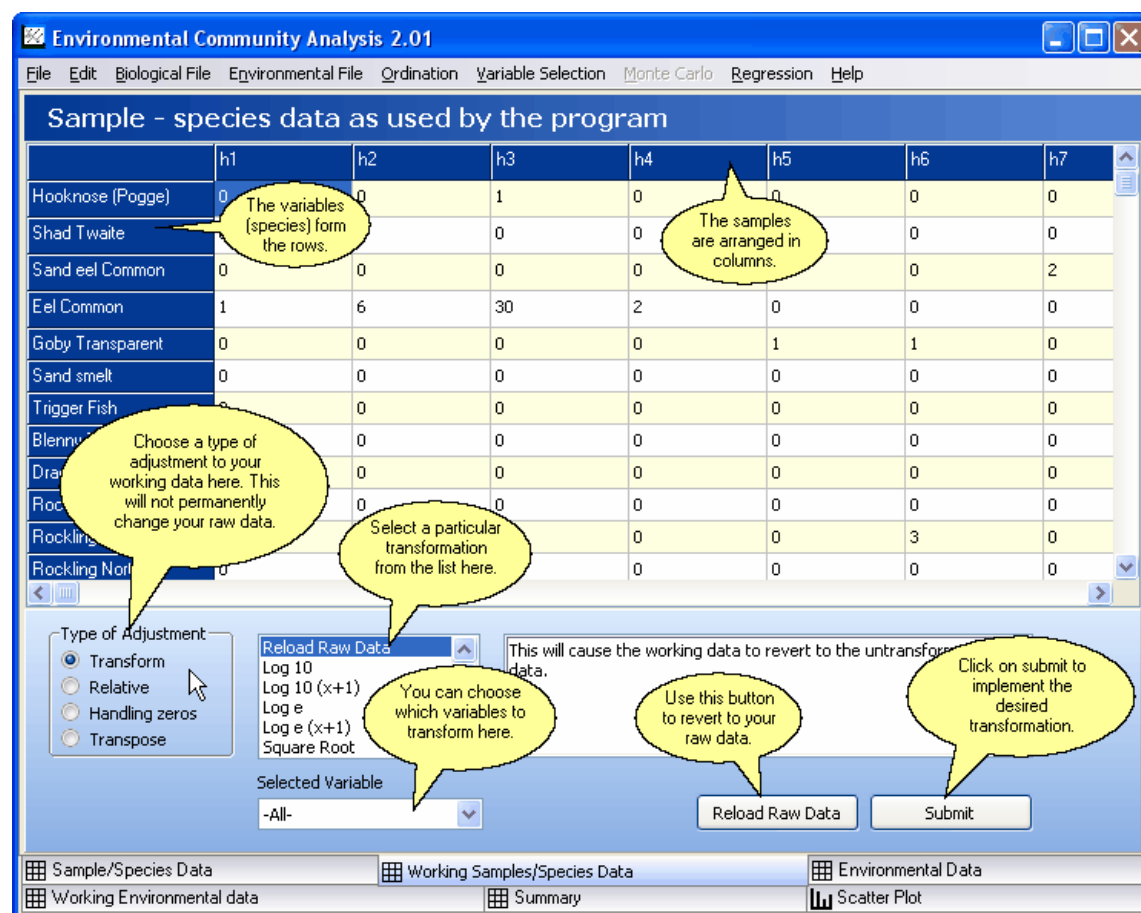
Part



XI

11 Transforming and adjusting your data

Your working data can be adjusted in a wide variety of ways from the **Working Sample/Species Data** tab. Use the Selected variable menu at the bottom of the window to select which variables are to be transformed or adjusted.



For more details see

[Data transformations](#)

[Relativisations](#)

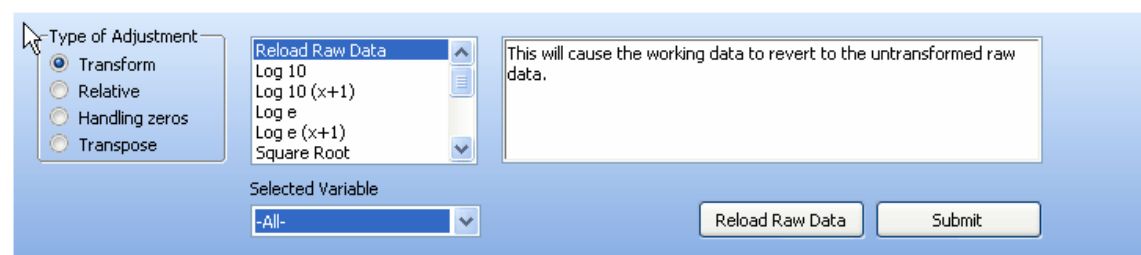
[Dealing with zeros](#)

[Transposing data](#)

[Transposing environmental variables](#)

11.1 Data transformations

The values in the working data grids can be transformed using a variety of functions. From the working data window select Transform in the Type of Adjustment panel.



The possible relative measures available within ECOM are given below. Select the transformation to be made and click on Submit to make the change. The transformation options within ECOM are

itemised below.

Reload Raw Data - This will cause the working data to revert to the raw data, so that you can start your investigations again.

Log 10 - Each value is transformed to the log to base 10. This cannot be done for numbers ≤ 0 .

Log 10 (x+1) - Each value is transformed by adding 1 and then calculating the log to base 10. This is used when the data contains zero values.

Log e - Each value is transformed to the log to base e (natural logs). This cannot be done for numbers ≤ 0 .

Loge (x+1) - Each value is transformed by adding 1 and then calculating the log to base e. This is used when the data contains zero values.

Square Root - the square root of each number is calculated. This cannot be done for negative numbers.

Arcsin - The Arcsin of each value is calculated. A transformation often used for percentage data.

Arcsin root - The Arcsin of the square root of each number is calculated.

Power - Each value, x, is transformed to x^a , where a is chosen by the user.

Add constant - A constant value, chosen by the user, is added to each value.

Subtract constant - A constant value, chosen by the user is subtracted from each value.

Multiply by constant - Each value is multiplied by a constant value chosen by the user.

Divide by constant - Each value is divided by a constant value chosen by the user.

See also: [Transforming environmental variables](#)

11.2 Data relativisations

The values in each row or column of the working data can be transformed so that their magnitudes are expressed relative to a variety of statistical measures. On the Working Data page select Relative in the Type of Adjustment panel.

The possible relative measures available within ECOM are given below. In each case where by row or column is not stated you can select which will be used. Select the adjustment to be made and click on Submit to make the change.

By Maximum value - For each row or column the maximum value is found and all values are divided by the maximum.

By Mean - For each row or column the mean value is found and all values are subtracted from the mean.

By StdD - For each row or column the standard deviation value is found and all values are divided by the standard deviation.

Binary using Mean - For each row or column the mean is found and all values above the mean are given the value 1 and all values below the mean zero.

Binary using Median - For each row or column the median is found and all values above the median are given the value 1 and all values below the mean zero.

General by Row - This allows you to define a general relativisation to be applied to each row. Each value is divided by

$$\sum_{i=1}^{i=x} x_i^a$$

where i is the column and a is a user selected parameter (the default value is 1) entered into the Enter value box displayed in the lower panel.

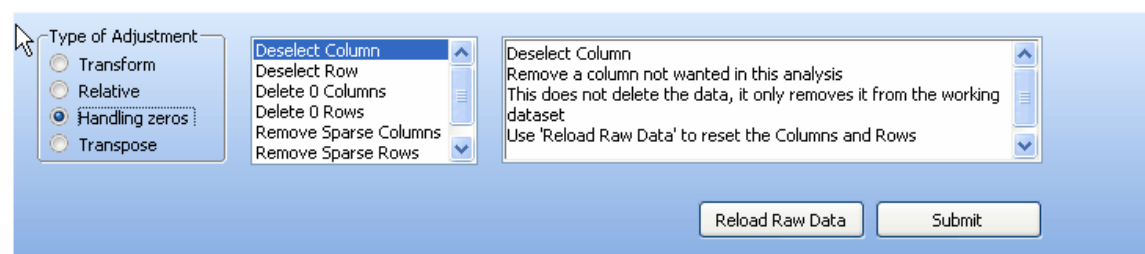
General by Column - This allows you to define a general relativisation to be applied to each column. Each value is divided by

$$\sum_{j=1}^{j=x} x \frac{a}{j}$$

where j is the row and a is a user selected parameter (the default value is 1) entered into the Enter value box displayed in the lower panel.

11.3 Dealing with zeros

Rows or columns in the working data holding zero values can be removed. On the Working Data page select **Handling zeros** in the **Type of Adjustment** panel.



Select the adjustment to be made and click on **Submit** to make the change. The possible options are as follows.

Deselect Rows: Removes row not wanted in this analysis. This does not delete the data, it only removes it from the working dataset

Deselect Columns: Removes column not wanted in this analysis. This does not delete the data, it only removes it from the working dataset

Delete zero columns - Every column in the data set that only contains zeros is removed.

Delete zero rows - Every row in the data set that only contains zeros is removed.

Remove sparse columns - Every column in the data set which contains $< x$ non zero elements is removed. The value of x is entered by the user in the At least x non zero value text box.

Remove sparse rows - Every row in the data set which contains $< x$ non zero elements is removed. The value of x is entered by the user in the At least x non zero value text box.

Beals - Beals smoothing is used on sparse data sets where many of the samples hold a small proportion of the total species list. As the transformation only uses the presence/absence of species, it should not be applied to good quality quantitative data as you will be discarding much of your information. It should only be performed on data arranged with the samples as columns. Described and discussed by Beals (1984) and McCune (1994) respectively, each element, e_{ij} , in the data array is replaced by:

$$e_{ij} = \frac{1}{N_i \sum_{k=1}^{k=\max S} \frac{Both_{jk}}{Samples_k}}$$

where i is the column (sample) number, j is the row (species) number, N_i is the number of species in column (sample) i , $Both_{jk}$ is the number of samples holding both species j and k , $Samples_k$ is

the number of samples holding species k and maxS is the total number of columns (samples).

11.4 Transposing data

Use this option to switch the rows and columns of the data sets - obviously if one data set is transposed, the other must be as well. Like all the other adjustments the transposition is applied to the working data set. Select **Transpose** in the **Type of Adjustment** radio box and click on the **Create working** data button.

NOTE: If you have allocated your samples to groups, you will find that your groups will be lost if you transpose your data. The solution is to save your working data using File | Export, then reload the data into ECOM; you can then re-assign samples to groups.

Part

XII

12 The output screens

The output from ECOM procedures are displayed in a tabbed notebook format. The contents of the different tabs are described below.

Tabs present after data are loaded

[Sample/Species Data](#)
[Working Samples/Species Data](#)
[Environmental Data](#)
[Working Environmental Data](#)
[Summary](#)
[Scatter Plot](#)

Tabs present after Redundancy Analysis

[Eigenvalues](#)
[Species Scores](#)
[Sample Scores](#)
[Ordination Plot](#)
[Correlation of Env.Vars](#)
[Biplot Scores](#)
[Multicollinearity](#)
[Monte Carlo](#)
[Scatter Plot](#)
[Variance](#)

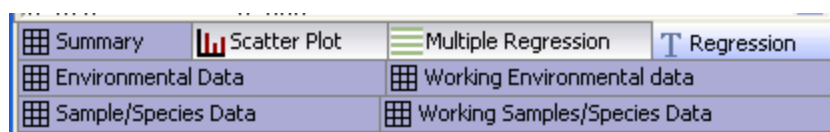
Tabs present after Canonical Correspondence Analysis

[Eigenvalues](#)
[Species Scores](#)
[Sample Scores](#)
[Ordination Plot](#)
[Rank Plot](#)
[Cor. Environmental Var](#)
[Binary Environments](#)
[Biplot Scores](#)
[Canonical Coef](#)
[Multicollinearity](#)
[Monte Carlo](#)

Tab present after Multiple Regression Analysis

[Multiple Regression](#)

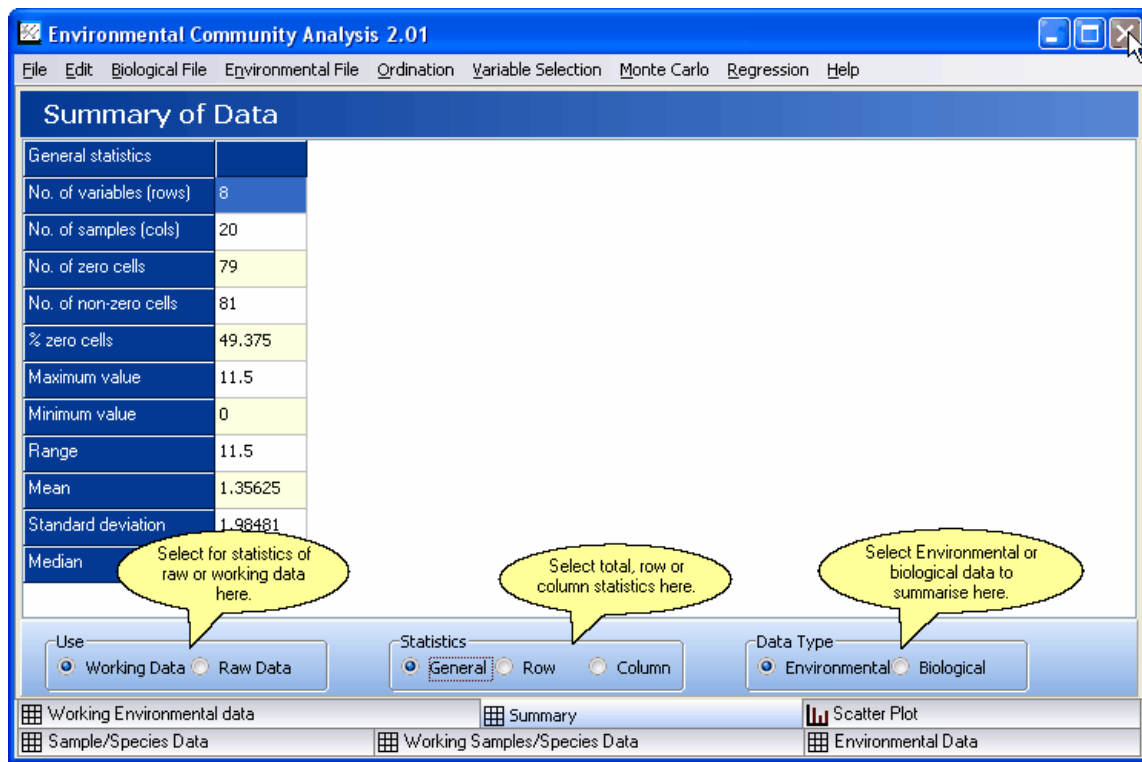
Each tab contains one of four symbols, to help you quickly identify what they contain.



1. A grid, signifying raw or working data, or basic statistics relating to the data.
2. A histogram, signifying graphical output.
3. A series of green & white lines, signifying calculated output.
4. A letter 'T', signifying text output.

12.1 Summary of data

This screen gives summary statistics for both the raw and the working data sets. When initially activated the screen shows general statistics for the working environmental data set. This can be changed using the options in the panel below the grid.



When first activated the data grid will display the following general statistics for the working data.

No. of variables (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.

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

% of zero cells - This is the number of zero entries divided by the number of cells in the data matrix.

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.

By using the radio buttons you can get a wider range of statistical facts about both the environmental and biological variables and the sites. Remember that these calculations can be undertaken on either the working or raw data.

To print the output select **File|Print**

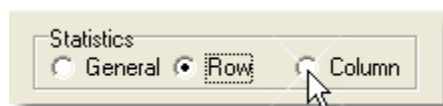
For more information on the statistics presented see:

[General dataset statistics](#)

[Statistics for columns and rows](#)

12.2 Statistics radio buttons

Use the Statistics radio buttons to learn more about your data, either in columns or rows.



Values for each row or column are shown numbered in the order they appear in the original data matrix. The statistics calculated are as follows.

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.

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.

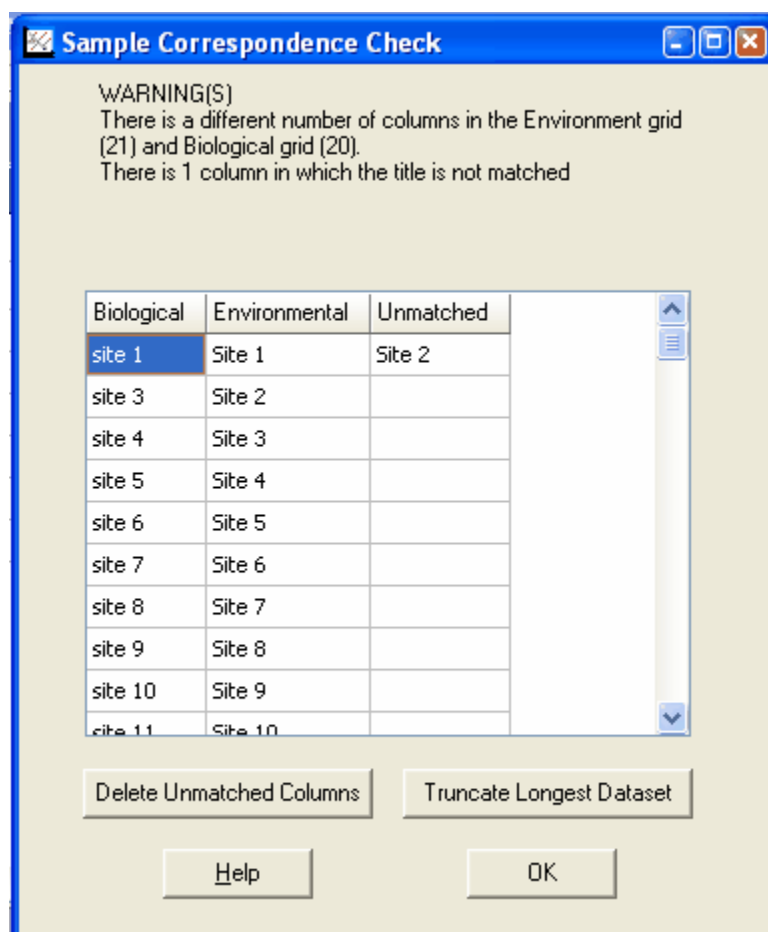
Variance - This is the variance of all the values in each row or column of the data matrix.

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.

12.3 Sample Correspondence Check

This form, which appears before any analysis is undertaken, is shown if the result of a comparison of sample titles in the Biological and Environmental data arrays shows them to be unmatched. The first column shows the order of titles in the Working Biological Data grid and the second the order of samples titles in the Working Environmental Data grid. Here, we deleted data for Site 2 in the Biological Data in order to create the disparity shown here. The third column indicates whether the two match. The comparison is not case sensitive.



Delete Unmatched Columns button - Use this to remove samples that are not present in both data sets.

Truncate Longest Dataset button - Use this to truncate both data sets to the number of samples in the smaller data set.

OK button - Click to continue with analysis

12.4 Environmental variables from biological features

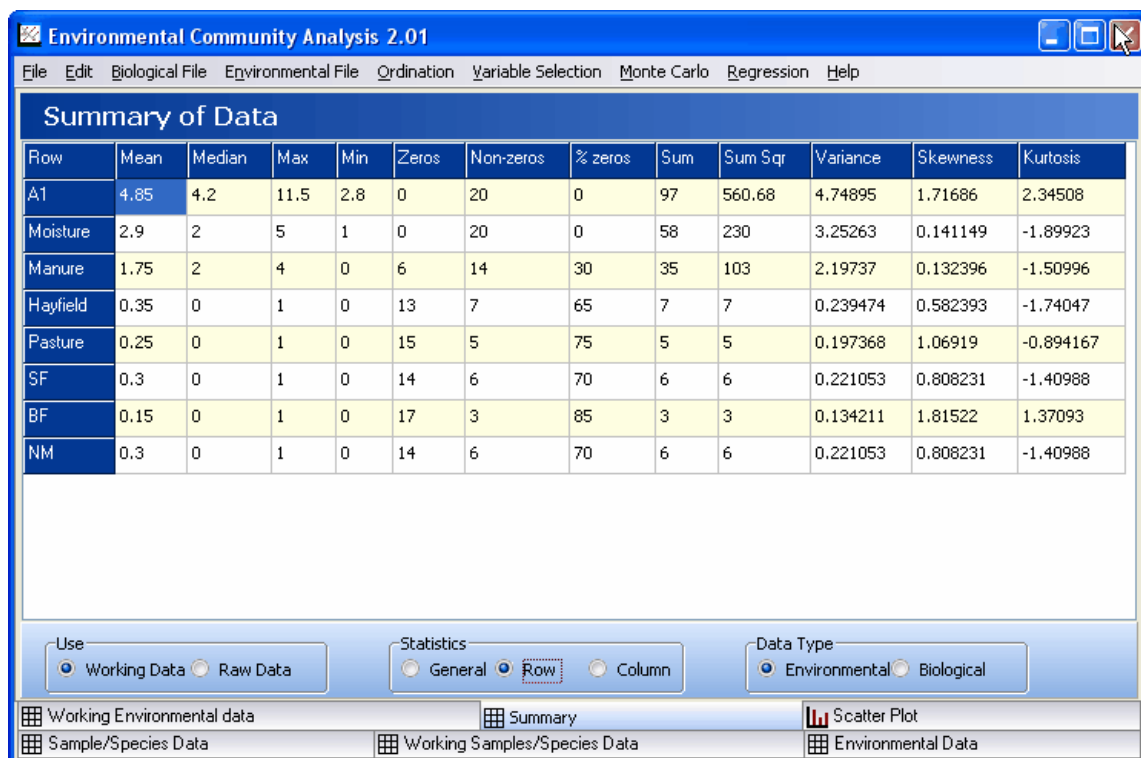
There is a paradox in gradient analysis: species respond to the environment, but they also modify the environment. For example, vegetation itself can be considered an environmental factor to which vegetation responds. There are a suite of variables derived from species data which might be useful in CCA and other constrained ordinations: maximum height of vegetation, total biomass, light penetrating through the canopy, woody plant cover, etc.

These variables might be quite informative in an exploratory analysis, though the ecologist must realize that it would be difficult to distinguish cause and effect. Species-derived variables should NOT be used in hypothesis testing, because the same data would be represented in both the dependent and the independent variables. This would lead to circular reasoning. A compromise would be to have these vegetation-derived variables considered "passive" or "supplementary" - i.e. they would be included in diagrams, but would not otherwise influence the ordination.

An extreme case of species-derived variables is to use dummy variables derived from a classification of samples. The classification could either be a result of a subjective procedure (e.g. the Braun-Blanquet approach or something less formal) or a multivariate analysis. When the centroids of these dummy variables are plotted along with species scores in a CCA biplot, you have an ideal display of the relationships between your classes, and the species that occur in them. Of course, any statistical tests would be inappropriate.

12.5 Data set statistics

Summary statistics for both the raw and working data sets are displayed by clicking on the **Summary** tab.



You can choose summary statistics for either the raw or working data sheets. Statistics can be generated for Rows, columns and general summary statistics for the data. These statistics are particularly useful prior to undertaking many analyses. For example analyses should not be undertaken on a data set that comprises large numbers of zeros. Use Summary statistics to determine the number of zero elements in your data.

To obtain general statistics on the raw data select the Raw Data radio button in the 'Use' panel situated below the grid to select Raw Data.
Statistics for the individual rows and columns of either the raw or working data matrix are selected using the Statistics radio buttons situated below the grid.

Values for each row or column are shown numbered in the order they appear in the original data matrix. The statistics calculated are as follows:

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

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

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

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

Zeros - This is the number of zero entries in the data matrix.

Non-zeros - This is the number of non-zero entries in the data matrix.

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

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

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

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

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

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

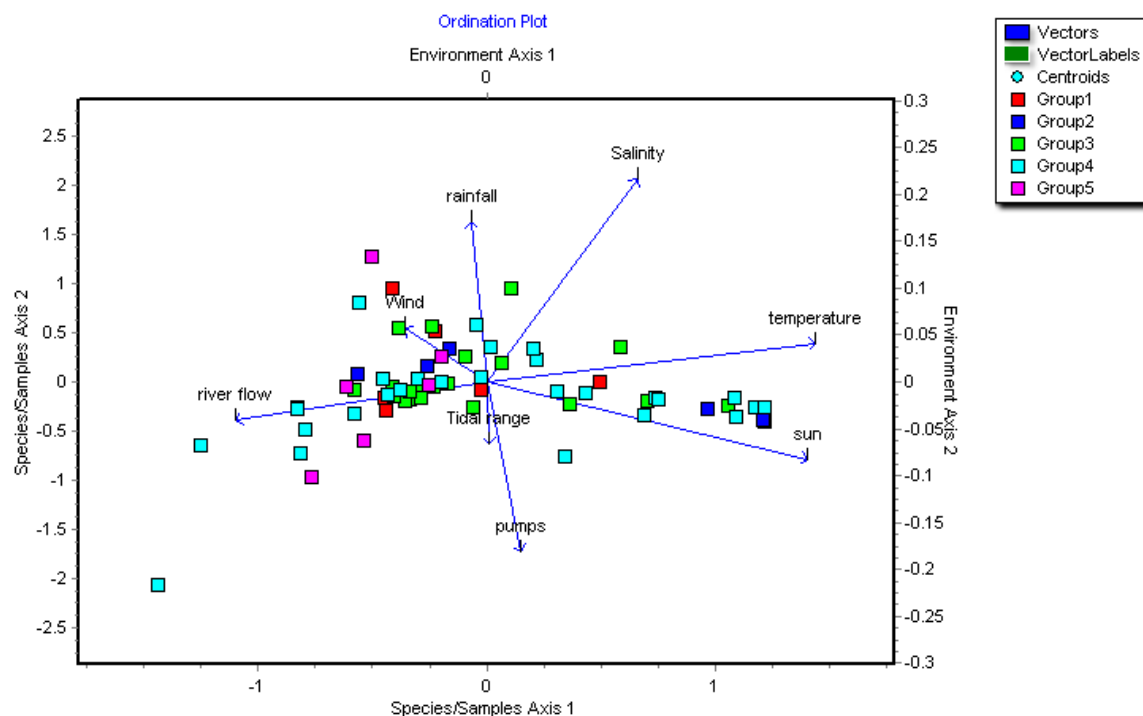
Part



13 Groups

It is often useful to place and display samples in groups.

For example, in the graph below, the species of fish caught in different years are split into several groups and displayed as different colours.



Choose **Edit|Groups** from the drop-down menu to create groups and assign individual samples to groups.

Assignment to groups.

1. First type a name for a group in the box to the right of the label 'Enter name here'. We have added the word Estuarine.

Groups

Sort by Name Sort by Group

Sort samples alphabetically

/	Sample Name	Group Name
1	Hinkley	Unassigned
2	W thur	Unassigned
3	Sizewell	Unassigned
4	Wylfa	Unassigned
5	Fawley	Unassigned
6	Oldbury	Unassigned
7	Heysham	Unassigned
8	Dung B	Unassigned
9	Hartlepool	Unassigned
10	Kingsnorth	Unassigned
11	Torness	Unassigned
12	Coolkeeragh	Unassigned
13	Ballylumford	Unassigned
14	Kilroot	Unassigned
15	Belfast West	Unassigned

Type a name for a group in the box below. To assign a , select it in the left-hand panel and click the Assign to Group button.

Enter name here Estuarine

List of group names Unassigned

Type group name here.

Assign To Group

Unassign Selected

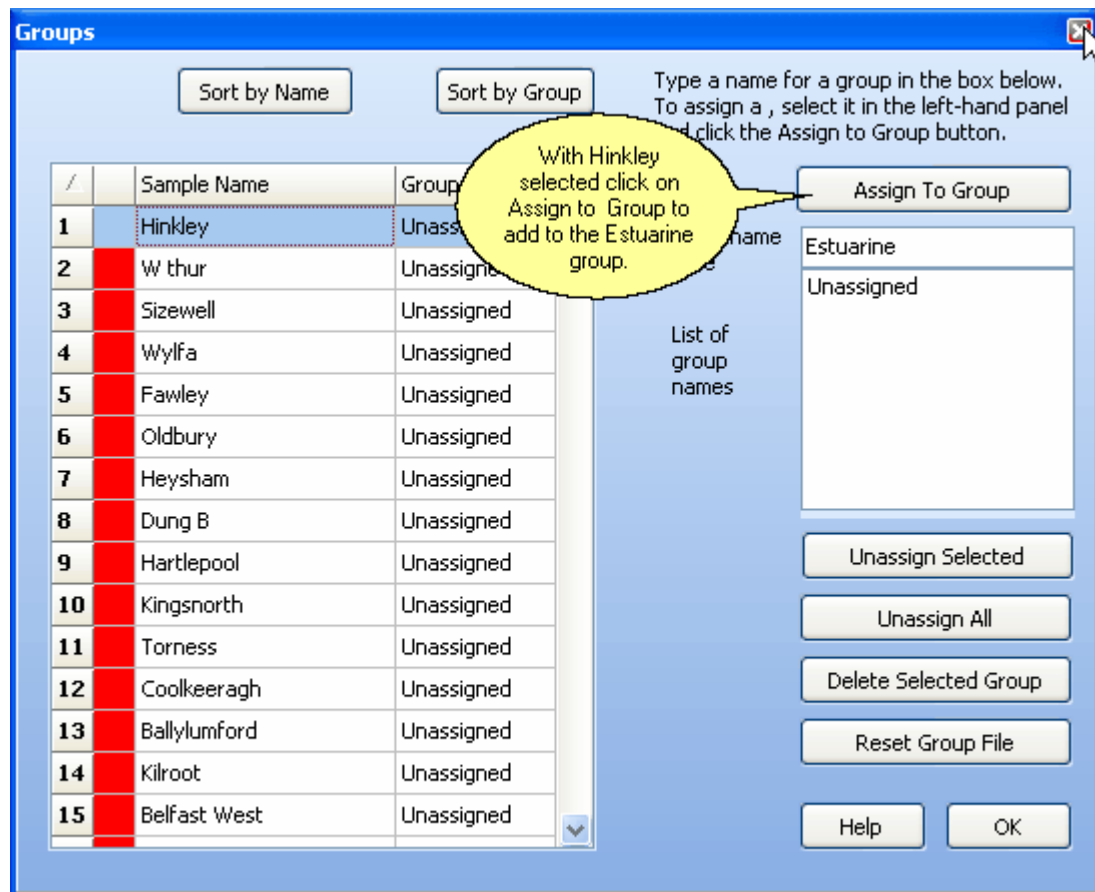
Unassign All

Delete Selected Group

Reset Group File

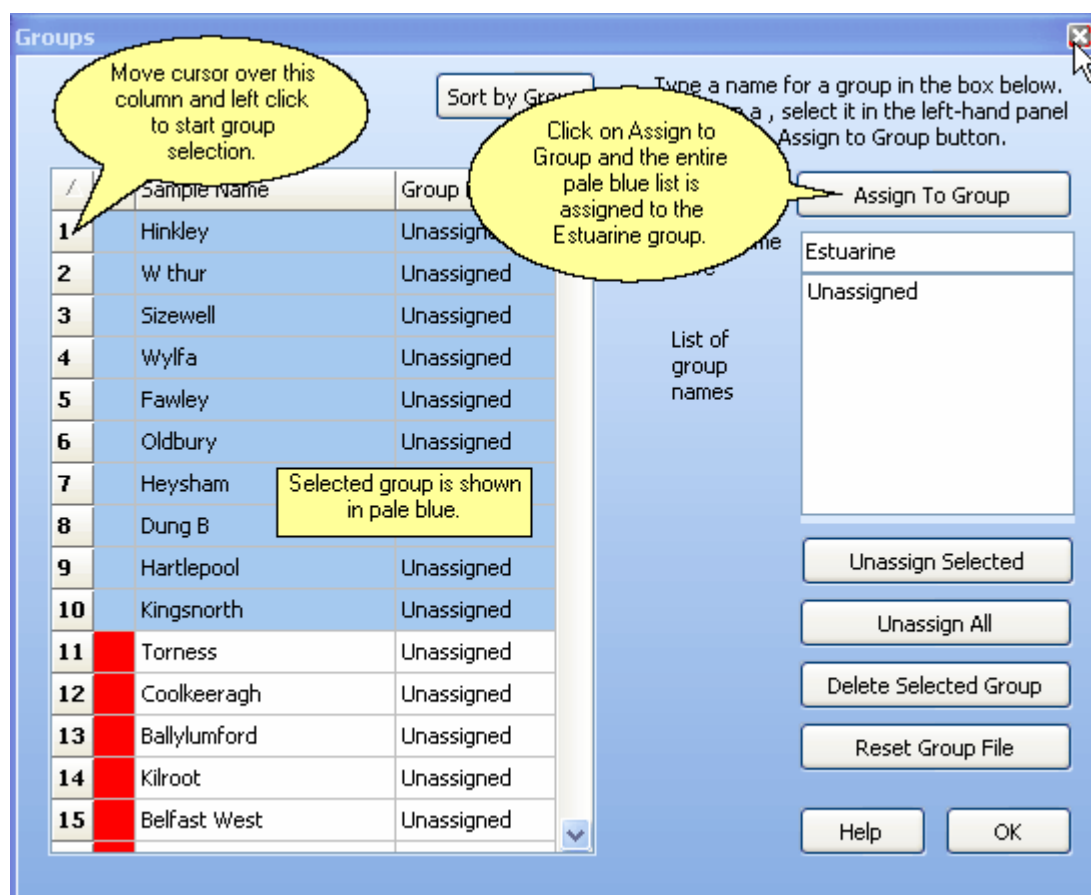
Help OK

2. Select a sample to allocate to this group from the left-hand list of samples in the data set: here we have selected Hinkley..



3. Click the Assign To Group button and the selected site is added to the group.

4. Blocks of samples can be assigned by clicking on the first sample required in the left hand column of numbers, holding down the left hand mouse button and dragging the cursor down to select a group, and then Clicking on **Assign to Group**.



To save the group allocations remember to save your dataset from the File menu. As well as creating the main data set, this will create a file of the same name as your main data file, with the .pcg file extension; your group assignments will be saved with your data.

You can remove individual samples from a group, or all group assignments, using the **Unassign Selected** and **Unassign All** buttons. If you wish to alter group names, you can remove any group name by selecting it in the 'List of group names' box and clicking the **Delete Selected Group** button. The group members for that group will revert to 'Unassigned'. You can sort your samples by Group or Name using the two buttons above the sample list.

NOTE: If you have allocated your samples to groups, you will find that your groups will be lost if you transpose your data. The solution is to save your working data using File | Export, then reload the data into ECOM; you can then re-assign samples to groups.

The data that controls the membership of groups is held in a file with the extension .pcg. There is one group file created for each data set saved from ECOM. The .pcg file is not meant to be edited manually, but via the Edit Group dialog within the program.

It is a text file with the following format:

Sample Name, Group Name, ""

Note that there are no headers on the columns. As an example:

```
3784,Rich Acres,""
4128,Rich Acres,""
4132,Rich Acres,""
4428,Quartz Gabbro,""
4436,Quartz Gabbro,""
44.44,Quartz Gabbro,""
4441,Hybrid,""
```

4445,Hybrid,""

If no groups have been set then each sample belongs to a group called Unassigned; e.g.:

G1,Unassigned,""

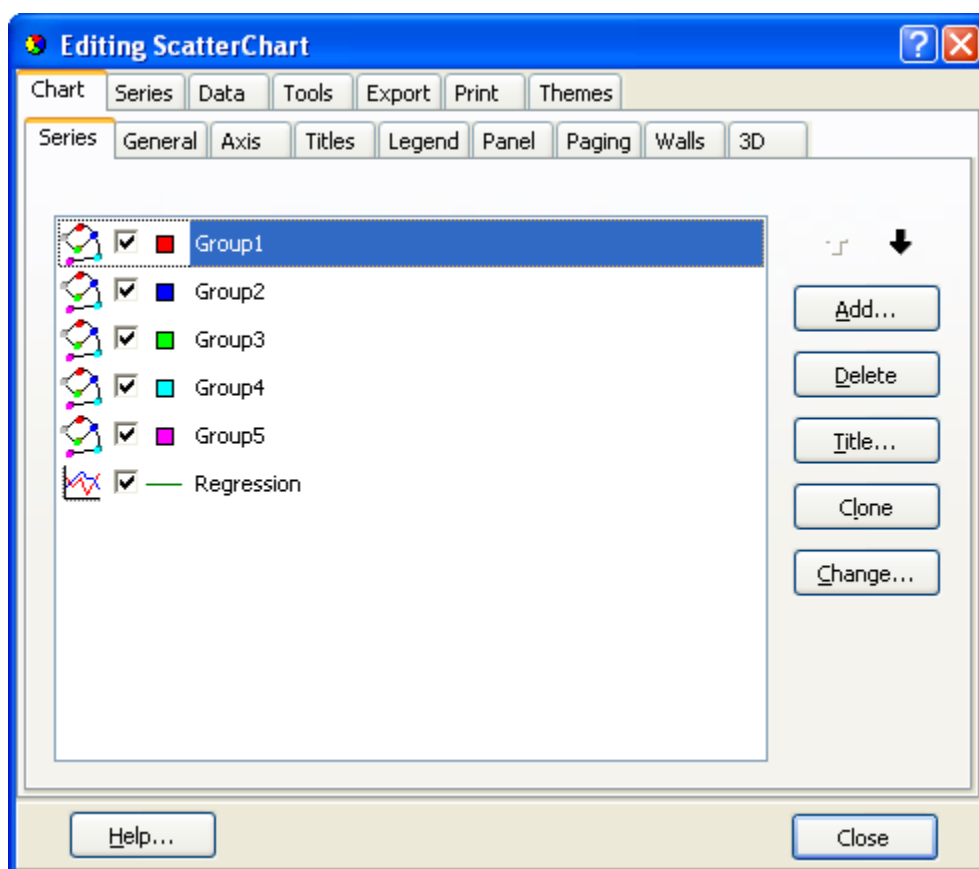
G2,Unassigned,""

X1,Unassigned,""

X2,Unassigned,""

When you have performed an analysis, if you enable the plot legend, the groups will be displayed on the legend.

If you wish, you can change the order in which the Groups are displayed; click the Edit Chart button, and on the Series tab, select a Group, and use the Up or Down arrow to move it into the order you want:



Part



14 Canonical Correspondence Analysis

CCA seeks an ordination of the sites (samples) in terms of their biological components which is optimised in terms of the influence of environmental factors.

[What is CCA?](#)

Environmental variables

[Selecting Environmental variables](#)

[Linear combinations of environmental variables](#)

[Dummy Environmental variables](#)

[Transforming environmental variables](#)

[Circular environmental variables](#)

Biological variables

The normal situation is for the biological data to comprise a sample (column) by species (row) array. The species data would comprise quantitative counts of the number of individuals in each sample. (See [Creating and editing a data set](#)).

Presentation of Results - Your output

CCA produces an ordination of the species and sites. It also produces biplots that show the inferred ranking of the species along an environmental variable. The output from ECOM is divided into a number of different sections of a tabbed notebook. These are described below.

[Eigenvalues](#)

[Species Scores](#)

[Sample Scores](#)

[Ordination Plot](#)

[Rank Plot](#)

[Correlation of Env. Vars](#)

[Binary Environments](#)

[Biplot Scores](#)

[Canonical Coef](#)

14.1 What is CCA?

Canonical correspondence analysis (CCA; ter Braak 1986, 1994) is an ordination method in which the ordination of the biological (main) matrix by correspondence analysis or reciprocal averaging is constrained by a multiple regression on the variables included in the environmental matrix.

In ecological terms, the ordination of sites and species is constrained by their relationships to environmental variables. If the environmental variables included are major determinants of community structure and abundance of species changes along these environmental gradients then this technique aids interpretation of community structure and identification of the features that mould it.

The alternative approach is to use ordination techniques such as Correspondence Analysis or Principal Components Analysis which ordinate the community (sites-species) data alone and then undertake an auxiliary analysis to identify the environmental variables that correlate most strongly with the ordination axes.

CCA excels at representing community data sets where: (1) species responses are unimodal (hump-shaped), and (2) the important underlying environmental variables have been measured. Note that condition 1 causes problems for methods assuming linear response curves (PCA) but causes no problems for CCA, according to ter Braak (1986, 1994). Condition 2 results from the environmental matrix being used to constrain the ordination results, unlike any other ordination

technique apart from Canonical Correlation. For this reason, CCA has been called a method for "direct gradient analysis" (ter Braak 1986).

CCA is currently one of the most popular ordination techniques in community ecology. It is, however, one of the most dangerous in the hands of people who do not take the time to understand this relatively complex method. The dangers lie in several areas: (1) Because it includes multiple regression of community gradients on environmental variables, it is subject to all of the hazards of multiple regression. Multicollinearity is a particular problem and it is easy to believe that a relatively high coefficient of multiple correlation implies a highly significant result which it may not. Further, it must be remembered that the method uses linear regression, it is quite likely that the response of the community to changes in an environmental variable may not be linear. (2) As the number of environmental variables increases relative to the number of observations, the results become increasingly dubious as the appearance of very strong relationships becomes inevitable. (3) Statistics indicating the "percentage of variance explained" can be calculated in several ways, each for a different question, but users frequently confuse these statistics when reporting their results.

CCA does not explicitly calculate a distance matrix. But CCA, like CA and PCA, is implicitly based on the chi-squared distance measure where samples are weighted according to their totals (Chardy et al. 1976; Minchin 1987a). This gives high weight to species whose total abundance in the data matrix is low, thus exaggerating the distinctiveness of samples containing several rare species (Faith et al., 1987; Minchin 1987a).

14.2 Selecting Environmental variables

The choice of environmental variables determines the outcome of CCA. For an exploratory analysis, include all the variables that you think are important determinants of the community. If there are other variables that you do not believe to be important but are easy to measure, then these should also be included during an exploratory analysis. You can always subsequently remove superfluous variables that add little to your insight or are difficult to interpret. Remember, if you are testing a hypothesis about the influence of selected variables on a community then the post-hoc removal of variables until you get an interesting result is not the way to proceed.

The number of environmental variables can range from 1 to more than the number of samples. If only 1 environmental variable is used then there is only one canonical axis and it is not possible to produce a 2-dimensional graph. You can, if desired, produce a 2-dimensional image in which the second axis is the first residual axis. If you have at least as many environmental variables as you have samples then your ordination is no longer constrained, and a simple correspondence analysis would result. When using ECOM, the number of environmental variables must be less than the number of samples.

Also see

[Linear combinations of environmental variables](#)
[Dummy Environmental variables](#)
[Transforming environmental variables](#)

Since the statistical significance of a CCA analysis is determined by a randomization test, there is no need to transform data to fulfill statistical assumptions. However, transformations can be used to dampen the influence of outliers. The choice of transformation impacts the location of sample scores, species scores, and environmental scores. A dampening transformation (e.g. square root) tend to make samples and species more evenly spread out. Only rarely will transformation of environmental variables change the overall interpretation of an ordination.

References cited

See also selected references for self-education.

Roberts, D. W. 1986. Ordination on the basis of fuzzy set theory. *Vegetatio* 66:123-31.

14.3 Linear combinations of environmental variables

An environmental variable cannot be a linear combination of other variables. For example, sediments can be defined by % clay, % sand, and % silt, which must add to 100%. Thus if you give the % of clay and sand then the % silt in the sample can be calculated as $100 - \% \text{clay} - \% \text{sand}$. The % silt is a linear combination of the other variables. Data that include linear combinations will produce a singular matrix which cannot be solved. ECOM will remove variables to avoid this problem. However, you should avoid this problem by eliminating unnecessary variables (eg only include two of the 3 particle types), as ECOM will not detect situations where rounding errors result in a situation where the linear combination is not exact; eg when you have entered 33.3% for all 3 sediment variables. If this occurs then the results can be unreliable as the numerical methods might not find an accurate solution.

This problem also occurs with the use of dummy variables.

See also:

[Selecting Environmental variables](#)
[Dummy Environmental variables](#)
[Transforming environmental variables](#)

14.4 Dummy Environmental variables

Some aspects of the environment cannot be described using continuous variables. For example land use is better described by categorical variables. Categorical variables need to be coded as dummy variables if they are to be used in either CCA or RDA. Dummy variables are binary; they take the value 1 or 0.

For every categorical variable with K categories, only K-1 dummy variables can be included in the analysis (see [Linear combinations of environmental variables](#)).

For example, suppose in a study of a chalk stream fauna at 5 sites we categorise the stream bed as unmodified, recently dredged or dredged more than 10 years ago, then these data can be entered as two variables as follows:

Site	unmodified	recently dredged
1	0	0
2	1	0
3	1	0
4	0	1
5	0	1

Site 1 was dredged more than 10 years ago, Sites 2 & 3 are unmodified and sites 4 & 5 have been recently dredged. Note that we do not need to explicitly include the category 'dredged more than 10 years ago'. Indeed, you **must not** include it if you are to avoid the environmental variables becoming linear combinations of each other.

See also:

[Selecting Environmental variables](#)
[Linear combinations of environmental variables](#)
[Transforming environmental variables](#)
[Circular environmental variables](#)

14.5 Transforming environmental variables

[Palmer \(1993\)](#) recommends logarithmic transformations of soil nutrient concentrations when undertaking vegetational analysis "because a 1-unit difference in nutrient concentration is probably much more important at low concentrations than it is at high concentrations".

Linear transformations of variables will have no effect on the results of a CCA. Therefore, changing the units of a variable (e.g. centimeters to millimeters) or the base point (e.g. elevation above a fixed point to elevation above sea level) will have no effect.

See also

[Data transformations](#)

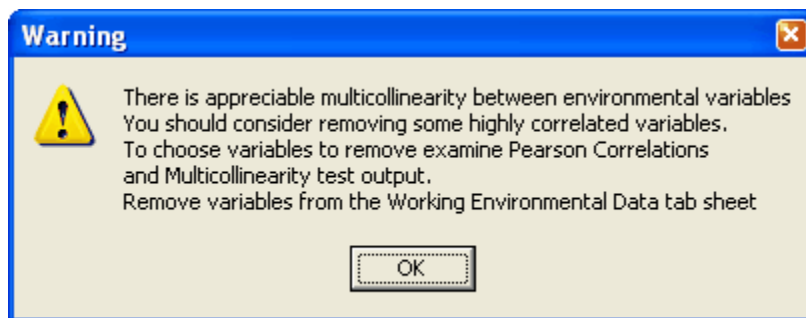
[Selecting Environmental variables](#)

[Linear combinations of environmental variables](#)

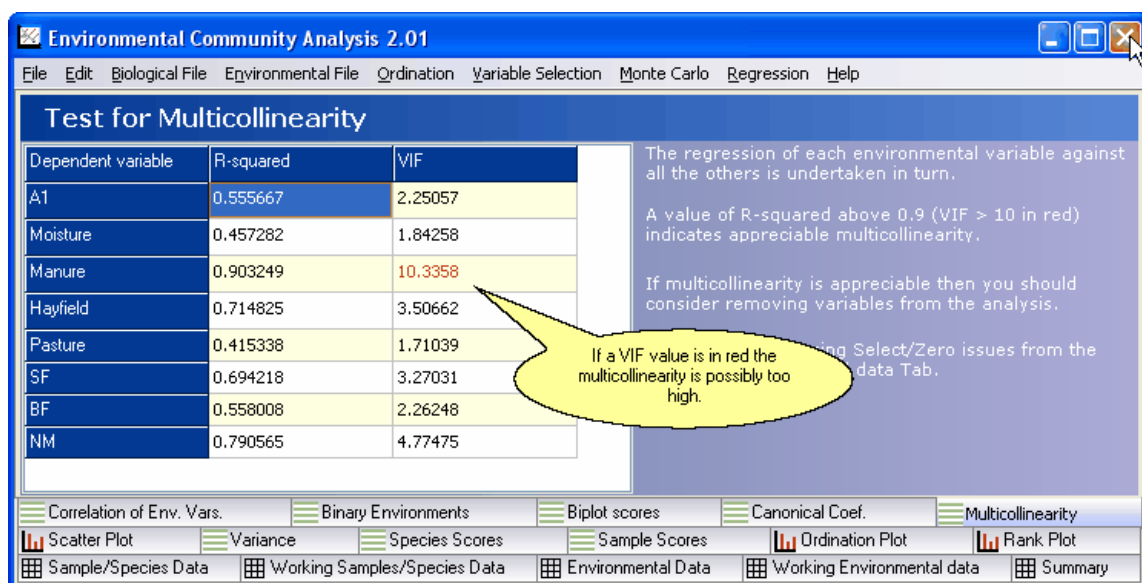
[Dummy Environmental variables](#)

14.6 Multicollinearity

If the environmental variables are highly correlated then results obtained by any multiple regression method can be highly unreliable. ECOM tests for appreciable collinearity between the environmental variables and will give a warning message if appreciable levels of collinearity are present.



The results of the multicollinearity test are presented in a tabbed sheet.



A multiple regression is undertaken for each of the environmental variables in turn with all the other environmental variables acting as the independent variables. Values of R-squared close to 1 or Variance inflation factors (VIF) well above 1 are indicative of multicollinearity and you should consider removing one of a group of highly correlated variables from the analysis. High VIF values are highlighted in red. Use [stepwise regression](#) and [Variable Selection](#) to identify which environmental variables to retain for your analysis

If you should obtain very large values for VIF it is likely that some of your variables add up to a constant. This often happens if you are using dummy variables see [Linear Combinations of Environmental Variables](#).

14.7 Circular environmental variables

Two common examples of circular data are direction (aspect) and day of the year.

Aspect (or compass direction of a slope) can be transformed by trigonometric functions (Roberts 1986). The simplest way to do this is to create two variables, "northerliness" and "easterliness" as follows:

$\text{easterliness} = \cos(\text{aspect})$

$\text{easterliness} = \sin(\text{aspect})$

Northerliness will take values close to 1 if the aspect is generally northward, close to -1 if the aspect is southward, and close to 0 if the aspect is either east or west. Easterliness behaves similarly, except that values close to 1 represent east-facing slopes.

Another approach is to use [dummy variables](#) e.g. N, NE, E, SE, S, SW, W, NW. This approach has the additional advantage of allowing flat surfaces (i.e. no aspect) to have a category.

The day of the year is circular because early January will have low values, and late December will have high values, yet the climates at the two times will be similar. As with compass direction one approach is to construct dummy variables Winter, Spring, Summer and Autumn. If sampling occurred over a limited number of months the Julian date can be used as a variable.

For more advanced work with circular or repeating variables Pisces produce a dedicated package 'Axis', more details of which are available [here](#).

14.8 Species Scores

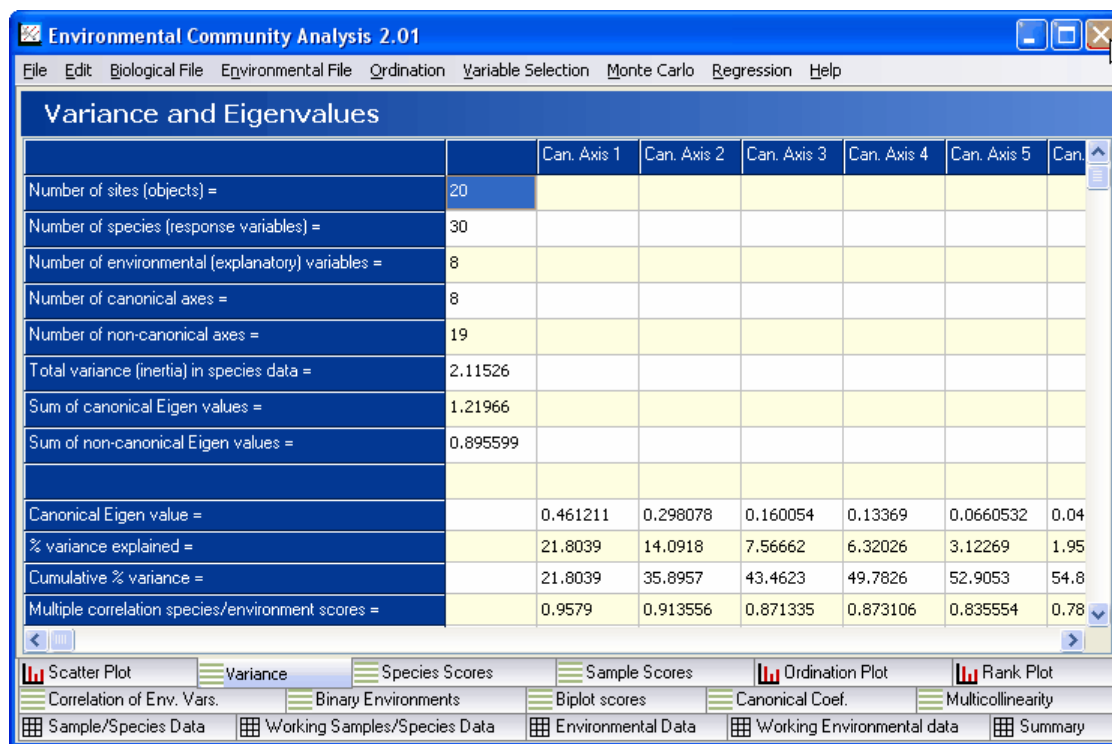
This grid gives the score for each species for each axis and are the values that are plotted to form the species ordination (see [Ordination Plot](#)).

	Can. Axis 1	Can. Axis 2	Can. Axis 3	Can. Axis 4	Can. Axis 5	Can. Axis 6	Can. Axis 7	Can. Axis 8	Axis 9
Achmil	0.840181	-0.381648	-0.0275585	-0.334142	0.107333	-0.308985	-0.0187439	0.00102603	-0.31794
Agrsto	-0.770423	0.500034	0.114276	-0.0800852	0.0555424	0.189701	-0.103764	0.0585975	0.341028
Airpra	-0.739535	-1.78737	1.07692	0.531809	0.248846	0.234991	-0.0990732	-0.108114	-2.75678
Alogen	-0.354077	0.96996	0.347032	0.138889	-0.205414	0.0220627	-0.0645755	0.0187861	-0.04457
Antodo	0.385784	-0.778054	0.0413796	0.2262	0.336673	-0.110704	-0.0687928	0.341086	-1.13791
Belper	0.712212	-0.169055	0.257959	-0.717673	0.378433	0.000144958	-0.407914	0.14445	0.226436
Brohor	0.832422	-0.0178449	0.160855	-0.857988	0.231838	-0.419679	0.0389218	0.0799226	-0.00261
Chealb	-0.93085	1.64736	0.693199	0.0270757	0.219168	-0.371949	-0.571582	-0.221614	-0.71823
Cirarv	0.397591	0.845227	0.987728	-0.666637	0.444475	1.22314	-0.345974	-0.0698164	0.035059
Elepal	-1.497	0.0904499	-0.574501	-0.382865	0.0684001	0.0891398	0.224726	0.171137	0.682068
Elyrep	0.637705	0.381922	0.371916	-0.00365445	0.41975	0.0946009	-0.161038	0.14486	0.272509
Emprig	-1.20563	-1.54905	1.45152	0.558478	0.476743	-0.252786	0.302853	-0.343742	-2.78162
Hyprad	-0.560475	-1.48629	0.702576	0.21677	-0.396043	-0.0476902	0.0508105	-0.113965	-2.2345

These results can be printed or exported from the File menu and copied from the Edit menu. See [Export Active Grid](#)

14.9 Variance & Eigenvalues

For both Redundancy Analysis and Canonical Correspondence Analysis this tab presents tabulated data for the eigenvalues of each axis and the amount of variance present in the original data that is accounted for by the axis. It also gives summary information on your data set. The total variance or what is more strictly termed inertia is the sum of the canonical and non-canonical Eigen values.



Environmental Community Analysis 2.01

File Edit Biological File Environmental File Ordination Variable Selection Monte Carlo Regression Help

Variance and Eigenvalues

		Can. Axis 1	Can. Axis 2	Can. Axis 3	Can. Axis 4	Can. Axis 5	Can. Axis 6
Number of sites (objects) =	20						
Number of species (response variables) =	30						
Number of environmental (explanatory) variables =	8						
Number of canonical axes =	8						
Number of non-canonical axes =	19						
Total variance (inertia) in species data =	2.11526						
Sum of canonical Eigen values =	1.21966						
Sum of non-canonical Eigen values =	0.895599						
Canonical Eigen value =		0.461211	0.298078	0.160054	0.13369	0.0660532	0.04
% variance explained =		21.8039	14.0918	7.56662	6.32026	3.12269	1.95
Cumulative % variance =		21.8039	35.8957	43.4623	49.7826	52.9053	54.8
Multiple correlation species/environment scores =		0.9579	0.913556	0.871335	0.873106	0.835554	0.78

Scatter Plot Variance Species Scores Sample Scores Ordination Plot Rank Plot

Correlation of Env. Vars. Binary Environments Biplot scores Canonical Coef. Multicollinearity

Sample/Species Data Working Samples/Species Data Environmental Data Working Environmental data Summary

The output is explained below.

Number of sites - This is the number of columns in the Sample/Species array loaded as the biological file. It may be the number of transects, quadrats, pitfall traps etc.

Number of species - This is the number of rows in the Sample/Species array loaded as the biological file. It is the total number of taxa in the data set.

Number of environmental variables - This is the number of rows in the Environmental array loaded as the biological file. It is the total number of taxa in the data set.

Number of canonical axes - This is the number of axes that can be related to environmental variables.

Number of non-canonical axes - This is the number of axes that hold the residual variance unexplained by a linear combination of the environmental variables.

Total variance - This is the sum of the variance explained by the canonical and non-canonical eigenvalues.

Canonical Eigenvalues - These are the eigenvalues for the canonical axes.

% variance explained - This is the amount of the total variance in the data that is explained by each of the canonical axes.

Cumulative % variance - This is the sum total of the variance explained by the addition of each new canonical axis.

Multiple correlation species/environment scores - The analysis produces scores for each site based on either their environmental variables or species composition. This row gives, for each canonical axis, the correlation between these two scores for all the sites. The value presented is the multiple correlation coefficient produced by a multiple regression of the environmental variables and the site (sample) scores derived from the weighted average of the species as the dependent variable. This value should be viewed with caution see Problems with Species-Environment Correlation.

Kendal rank correlation species/environment scores - The analysis produces scores for each site based on either their environmental variables or species composition. This row gives, for each canonical axis, the Kendal rank correlation between these two scores for all the sites.

Non-canonical Eigenvalues - These are the eigenvalues for the non-canonical axes and represent the variance that could not be explained by a linear combination of the environmental variables. These are the results of the residual analysis.

% variance explained - This is the amount of the total variance in the data that is explained by each of the non-canonical axes.

Cumulative % variance - This is the sum total of the variance explained by the addition of each

new non-canonical axes.

These results can be printed or exported from the File menu and copied from the Edit menu. See [Export Active Grid](#)

14.10 Sample Scores

This grid gives the score for each sample for each axis and are the values that are plotted to form the sample ordination (see [Ordination Plot](#)).

Environmental Community Analysis 2.0

FileEditEnvironmental FileOrdinationVariable SelectionMonte CarloRegressionHelp

Sample Scores

	Sample	Can. Axis 1	Can. Axis 2	Can. Axis 3	Can. Axis 4	Can. Axis 5	Can. Axis 6	Can. Axis 7	Can. Axis 8	Axis 9	Axis 10
Scores derived from the species scores.	site 1	0.828	0.2713	0.3741	-0.4579	0.05526	0.4691	0.4734	0.1945	0.3575	-0.1571
	site 2	0.587	0.1367	0.2143	-0.6226	0.2155	-0.3602	-0.06872	0.1257	0.1487	0.02165
	site 3	0.2139	0.5512	0.3601	-0.2332	-0.02291	0.2499	-0.08441	0.06371	0.07138	-0.1563
	site 4	0.1612	0.5081	0.5086	-0.3828	0.1255	0.3975	-0.2442	0.03666	0.01898	-0.2007
	site 5	0.7786	-0.1312	-0.3658	0.2281	0.4775	0.006008	-0.09848	-0.3295	-0.2205	0.08397
	site 6	0.7002	-0.2011	-0.7198	0.5468	0.06579	0.2269	0.1466	0.2084	0.02993	0.5697
	site 7	0.7023	-0.08256	-0.3851	0.2372	-0.0406	0.002383	0.2099	-0.01863	-0.133	0.07363
	site 8	-0.4737	0.3914	0.04313	-0.1009	-0.02968	-0.05874	0.4766	0.09018	-0.07391	-0.3002
	site 9	0.03619	0.5486	0.1958	0.4539	0.09343	-0.3062	0.02195	-0.08344	0.3886	-0.5212
	site 10	0.654	-0.3277	-0.09435	-0.5329	-0.0704	-0.4168	-0.0146	-0.1533	0.008558	0.3478
	site 11	0.3446	-0.5089	-0.05848	-0.1884	-0.9637	0.07238	0.081	0.1672	-0.2066	-0.4958
	site 12	-0.2402	0.7874	0.2095	0.6674	-0.4836	-0.2573	-0.4836	0.04583	-0.2346	-0.125
	site 13	-0.3205	0.8837	0.363	0.2396	-0.1174	-0.2915	-0.2442	0.1192	-0.3888	-0.03755
	site 14	-1.372	-0.1452	-1.018	-0.6597	-0.04204	0.0992	-0.3059	0.05871	0.1663	0.04996
	site 15	-1.33	-0.02697	-0.8679	-0.3145	0.4769	-0.03879	-0.04764	0.2293	0.1399	-0.1967
	site 16	-1.312	0.3771	-0.3207	-0.193	0.03168	0.2691	0.3928	-0.5466	0.3302	0.719
	site 17	0.2636	-1.512	0.4262	0.3307	0.2181	-0.09243	-0.08407	-0.5293	-1.472	0.7284
	site 18	0.211	-0.8161	0.0587	-0.02711	-0.3461	0.1987	-0.3496	0.03245	0.6954	-0.648
	site 19	-0.4514	-1.569	1.066	0.6305	-0.004099	-0.1294	-0.1945	0.3915	-1.506	-0.1314
	site 20	-1.359	-0.5476	0.1054	0.1198	0.3317	0.215	0.4121	-0.1996	1.38	0.4507
Sample										Row sum (weights)	
Scores that are linear combinations of environmental variables and weights	site 1	0.6018	0.2366	0.5107	-0.08501	-0.05455	0.3431	-0.05665	0.04561	18	
	site 2	0.7083	-0.0573	0.09135	-0.6029	0.00749	-0.233	0.005138	0.2806	42	
	site 3	0.2655	0.4692	0.3791	-0.2549	0.1222	0.248	-0.06837	0.006707	40	
	site 4	0.27	0.4615	0.3952	-0.2437	0.1142	0.2488	-0.06378	0.01109	45	
	site 5	0.8093	0.04026	-0.5161	0.02526	0.4506	-0.01844	-0.09998	-0.3275	43	
	site 6	0.8105	0.1224	-0.3938	0.5711	7.613E-5	0.02596	-0.01683	0.267	48	
	site 7	0.563	-0.2023	-0.3259	0.08139	-0.1232	0.2792	0.3397	-0.009064	40	
	site 8	-0.5738	0.3711	-0.1149	-0.08633	0.216	-0.06541	0.504	0.03682	40	
	site 9	0.1067	0.1367	0.0713	0.7684	0.04079	-0.3089	0.06045	-0.1107	42	
	site 10	0.523	-0.2443	0.2746	-0.4441	-0.03342	-0.4041	-0.05026	-0.1682	43	
	site 11	0.3615	-0.3663	-0.393	-0.3337	-0.7405	-0.1117	0.03198	-0.01607	32	

Sample Scores

Ordination Plot

Rank Plot

Correlation of Env. Vars.

Binary Environments

Biplot scores

Canonical Coef.

Multicollinearity

Sample/Species Data

Working Samples/Species Data

Environmental Data

Working Environmental data

Summary

Scatter Plot

Variance

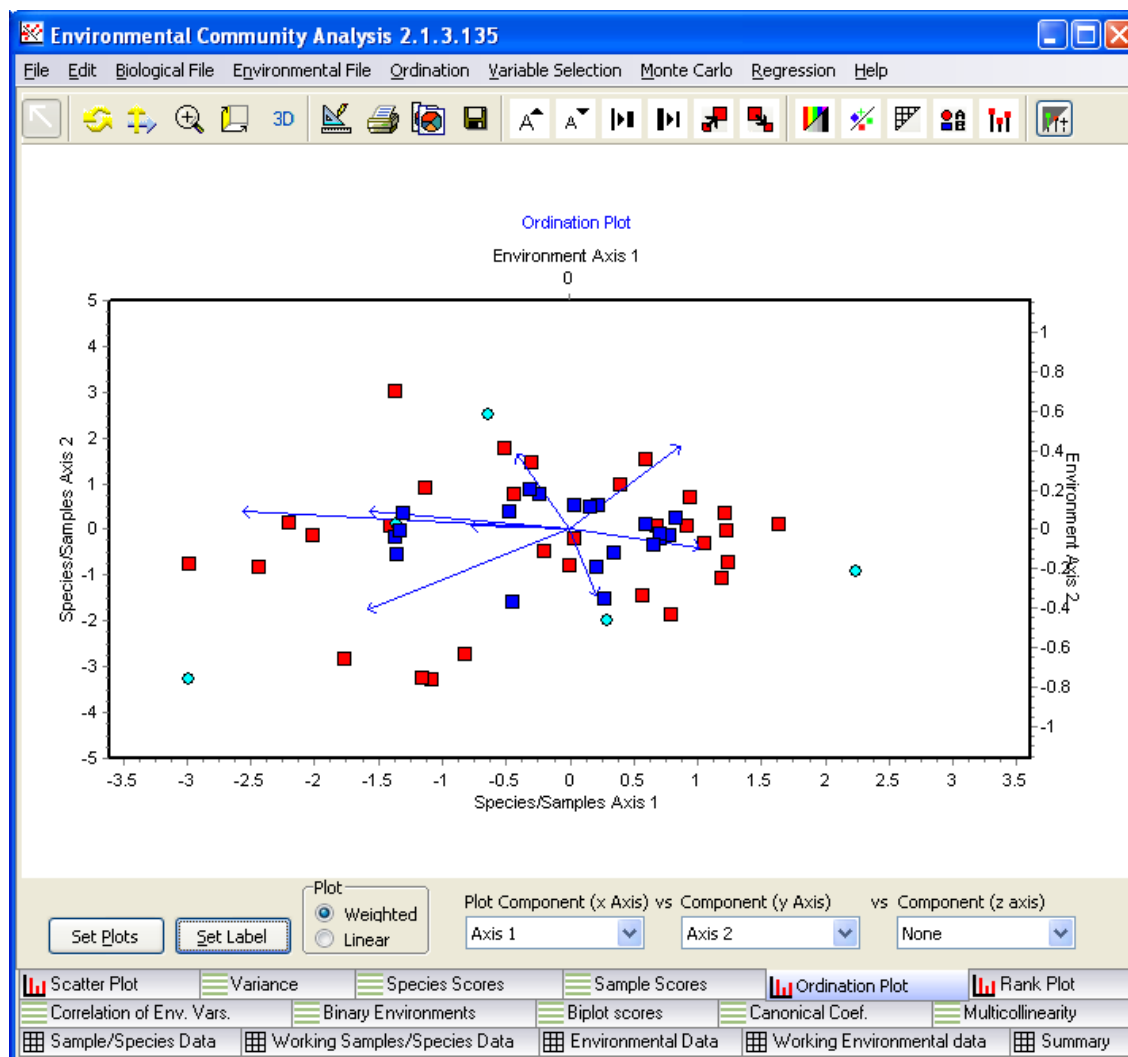
Species Scores

For each sample there are two sets of scores and both are presented in the grid. The upper set of scores, labelled 'Scores derived from the species scores' are calculated using the species data. The lower set labelled 'Scores that are linear combinations of environmental variables and weights' are derived from the linear combinations of the environmental variables. The column labelled 'Row sum weights' is the total sum of the observations for each environmental variable and gives the weighting applied to each environmental variable during the calculations.

These results can be printed or exported from the File menu and copied from the Edit menu. See [Export Active Grid](#)

14.11 Ordination Plot

This tab shows the ordination plot of the species and site scores. The graph can also show the environmental vectors. It will produce a figure that can be used to interpret the similarity between sites in terms of their species composition and species in terms of their habitat requirements. When the tab is first selected the graph will show species, sites and environmental vectors plus their labels.



If you have a fair number of species, sites or environments it can be confusing and impossible to read; many display options are available to produce a figure that allows you to see clearly the relationships within your data. Below the graph there are two buttons **Set Plots** and **Set Label** which will display options that can be used to select the data and labels that will be displayed, and drop-down menus.

The Plot radio button can be used to plot either the sample scores derived from the weighted species averages or the sample scores derived from the linear regression of the environmental variables. You should normally plot the default which is the weighted average sample (sites) scores.

The Plot Component drop-down boxes allow the selection of the axis for plotting. Generally, you will only be interested in plotting the first 2 or 3 axes. If you select an axis in the right hand drop-down menu then you will be presented with a 3D graph.

You can change almost every aspect of your graph. To change graph options, print and export see [Graph display options](#).

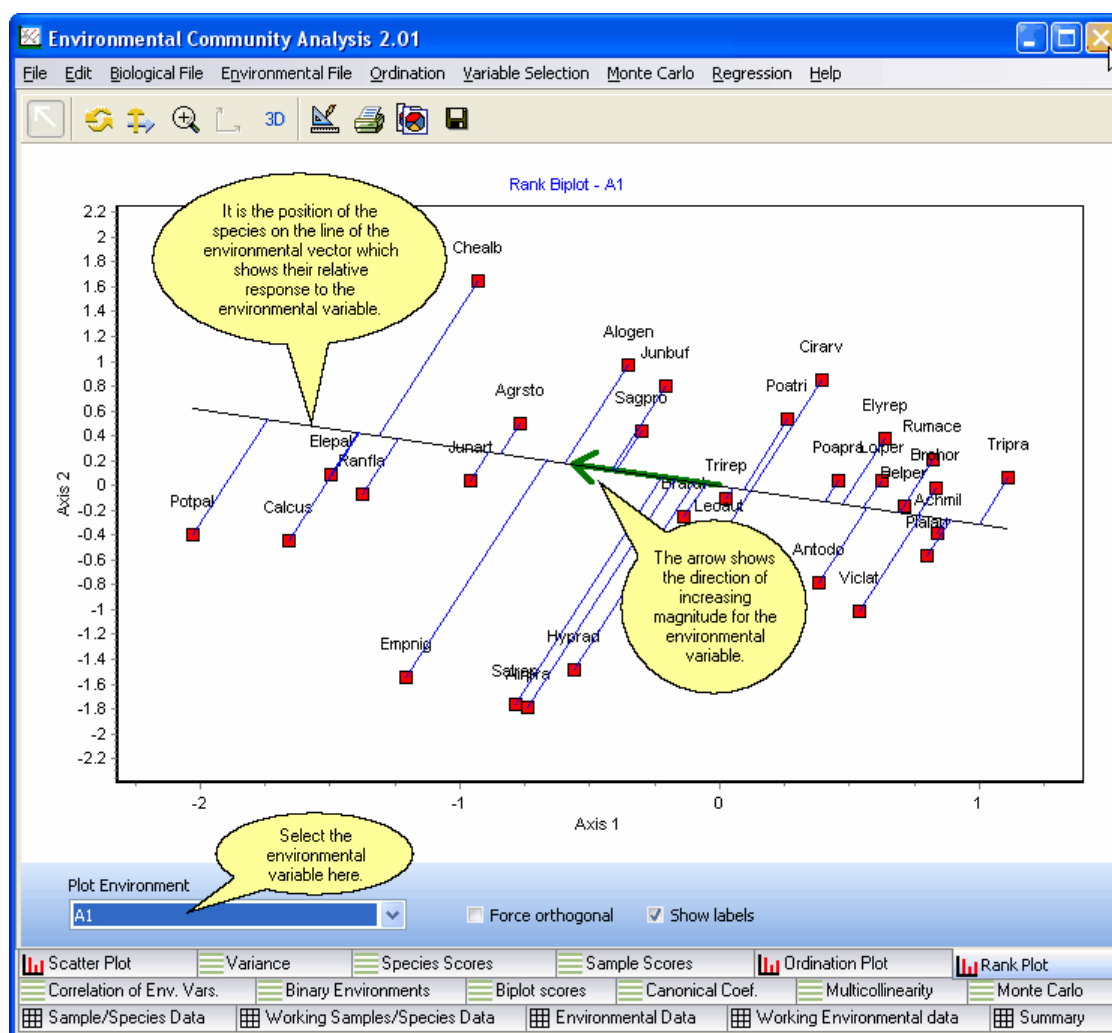
To pan and zoom within the graph area see [Zooming and panning on graphs](#)

To export your graph see [Export Tee Chart](#).

To print your graph see [Printing and exporting your results](#)

14.12 Rank Plot

This plot shows the inferred ranking of the species (or other taxa) along an environmental variable presented in the biplot. It will enable you to identify species that are particularly positively or negatively correlated with the chosen environmental variable. It will also allow you to present the species in order of preference for the chosen environmental variable. The position of each species is projected onto the line formed by the environmental vector. Please note that while the lines are projected at right angles, they will not appear so until you adjust the shape of the graph by changing the window or the axes. To adjust the graph axes click on the Edit graph button above the plot and select Axis.



The environment is selected using the **Plot Environment** drop-down menu below the graph and labels can be added or removed by clicking on the **Show labels** tick box.

To export your graph see [Export Tee Chart](#).

To print your graph see [Printing and exporting your results](#)

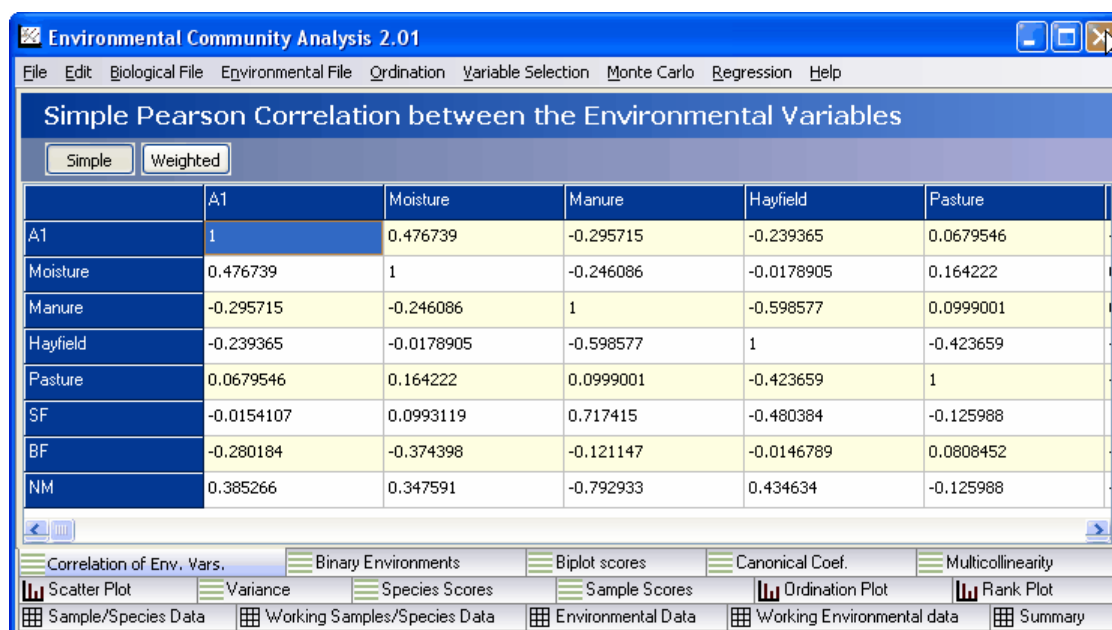
To change graph options, print and export see [Graph display options](#).

To pan and zoom within the graph area see [Zooming and panning on graphs](#)

14.13 Correlation of Env. Vars

This tab gives the weighted and unweighted Pearson correlations between the environmental variables. When two environmental variables are highly correlated, only one of them is likely to be of value as a predictor of species composition. If they are perfectly correlated ($R = 1$), then they both cannot be used in the analysis as it will result in a singular matrix (i.e. it has no inverse matrix). The best analysis will tend to be produced following the selection of environmental variables that are not highly correlated.

When running a Redundancy analysis, the Correlation between Environmental Variables is shown in a single grid as shown. These are the Simple Pearson Correlations.



When running a Correspondence Analysis two buttons will be shown, the simple, (see above) is selected by pressing the 'Simple' button, and the Weighted Pearson Correlations weighted by the row sums of the species matrix is selected with the 'Weighted' button.

These results can be printed or exported from the File menu and copied from the Edit menu. See [Export Active Grid](#).

14.14 Binary Environments

While quantitative environmental variables can best be represented on a biplot as a vector (see [Correlation of Env. Vars](#)) some environmental variables such light or dark are best coded as a 0 or 1. These are termed binary variables (see [Dummy Environmental variables](#)). Such variables can be represented on a biplot as a vector but are best displayed as a single point - a centroid. This grid gives the co-ordinates of the binary variables in the data set. N/A (for not applicable) is marked beside quantitative environmental variables.

Environmental Community Analysis 2.01

File Edit Biological File Environmental File Ordination Variable Selection Monte Carlo Regression Help

Centroids for Binary Environmental Variables

Binary environmental variables are represented in the data as either 0 or 1

	Can. Axis 1	Can. Axis 2	Can. Axis 3	Can. Axis 4	Can. Axis 5	Can. Axis 6	Can. Axis 7
A1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Moisture	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Manure	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Hayfield	0.105657	-0.844836	0.352585	0.394256	0.427867	-0.506349	-0.184643
Pasture	-0.488361	0.0545132	-0.747686	-0.368582	-0.773894	0.329416	1.14506
SF	-0.227826	1.06807	0.646397	-0.132717	-0.267593	0.640674	-0.475085
BF	0.803015	-0.385612	0.0656163	-1.28782	-0.825304	-1.29159	-0.0427428
NM	-1.07264	-1.38797	-0.00150449	0.0800611	0.36188	0.264395	-0.44522

Correlation of Env. Vars. Binary Environments Biplot scores Canonical Coef. Multicollinearity Monte Carlo

Scatter Plot Variance Species Scores Sample Scores Ordination Plot Rank Plot

Sample/Species Data Working Samples/Species Data Environmental Data Working Environmental data Summary

These results can be printed or exported from the File menu and copied from the Edit menu. See [Export Active Grid](#)

14.15 Biplot Scores

This tab gives the co-ordinates of the environmental variables which are used in the biplots.

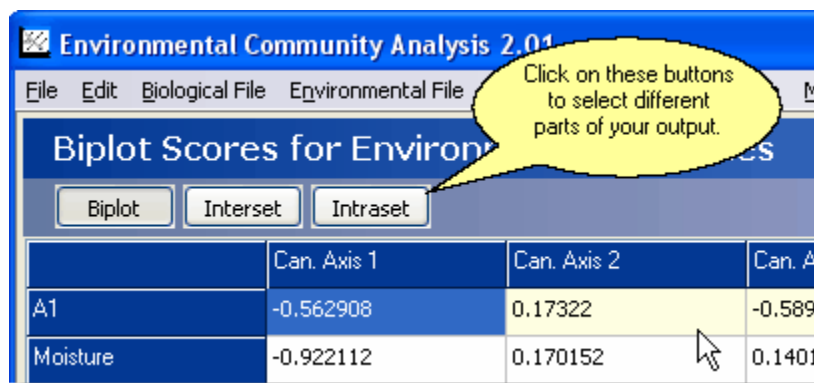
When a canonical Correspondence Analysis is undertaken you will be offered three buttons above the grid that can be used to show the correlations between the environmental variables and the site (sample) scores for each axis. These correlations will help to identify which environmental variables are most influential for each axis. For each axis, the environmental variable that has the greatest positive or negative correlation will be the most important determinant. These correlations should not be used in isolation as they can be misleading, you should also consider the size of the Eigenvalue (see [Variance & Eigenvalues](#)) for the axis to assess if the environmental variables explain an appreciable proportion of the total variability.

Use the buttons at the top of the form to show three different parts of the output .

Biplot - This button presents a grid of the coordinates of the tip of the environmental variable vectors used in the ordination biplots to show species relationships to the environmental variable (see [Rank Plot](#)).

Inter - This button presents the correlation between the environmental variables and the site (sample) scores calculated using the weighted average of the species. It is termed the Inter-set correlation.

Intra - This button presents the correlation between the environmental variables and the site (sample) scores calculated using the regression of the environmental variables with the site scores derived from the species. It is termed the Intra-set correlation. It is always larger than the intra-set correlation.

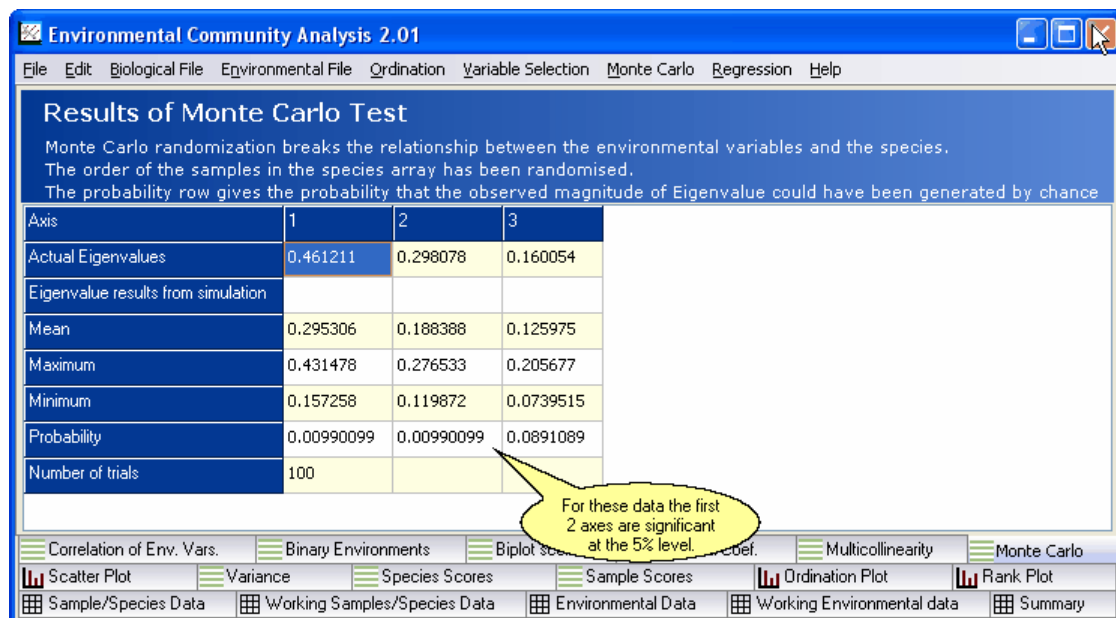


These results can be printed or exported from the File menu and copied from the Edit menu. See [Export Active Grid](#).

14.16 Monte Carlo Simulations

Monte Carlo randomization tests are only undertaken if selected from the Randomization drop-down menu. This menu allows you to select 100, 1000, and 10000 trials. It is a good idea to initially run only a hundred simulations as this procedure can take a long time if you have a large data set.

For each simulation the order of the samples in the sample-species array is randomly shuffled and the Eigenvectors calculated as usual. By shuffling the order of the samples we are breaking the correlation between the environmental variables and the species. Thus there will not be any relationship. The probability measure given in the Monte Carlo output grid is the probability that an Eigenvalue as large as that observed could have occurred by chance. Also given in the table is the largest and smallest Eigenvalues observed and the mean Eigenvalue. The calculations are only undertaken for the 3 axes with the largest Eigenvalues as it is unlikely that you will have significant variation explained by subsequent axes.

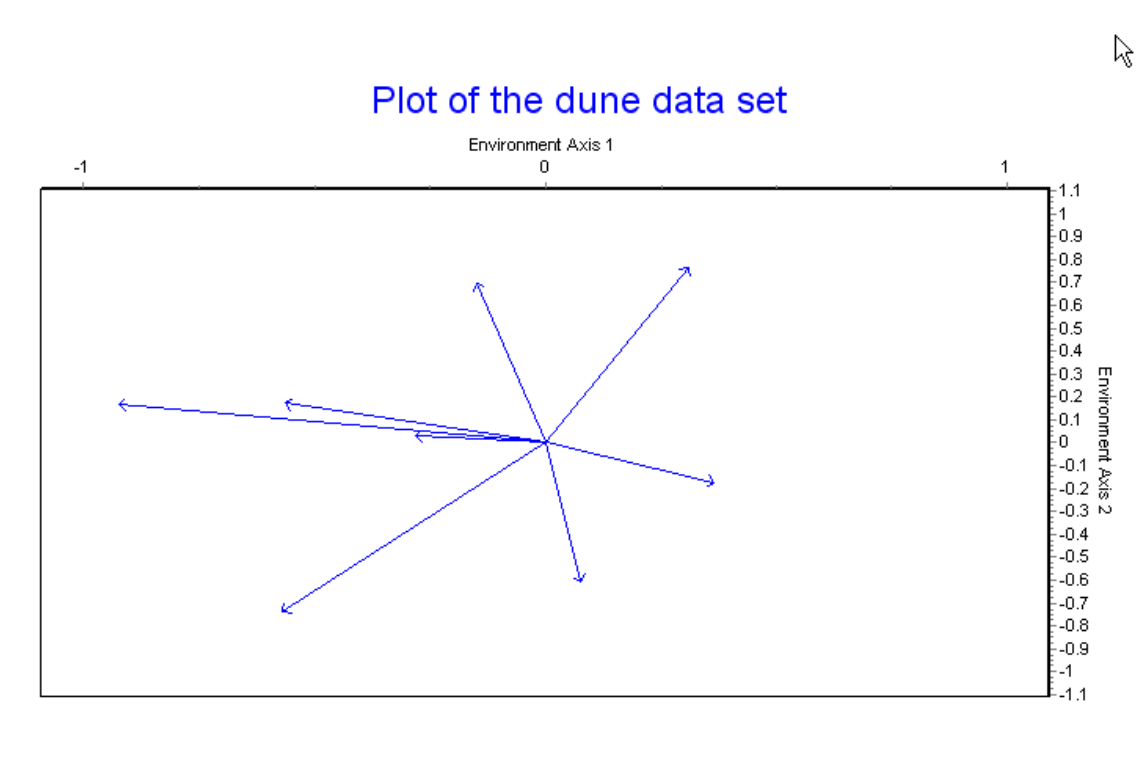


If the observed probability for an axis is greater than 0.05 then there is a greater than 5% chance that the observed magnitude of the Eigenvalue could occur by random chance and the axis is unlikely to explain any significant part of the variability in the species data.

14.17 Environment plot

This graph shows the environmental variables plotted on the ordination axes. It will allow you to see which variables are closely correlated with particular axes. For example, a high positive correlation with an axis will be shown by a variable that has a long arrow pointing in almost the same direction as the axis. A variable with an arrow at almost right angles to an axis indicates that it is almost uncorrelated with the axis.

You are able to plot the environmental variables with the ordination space defined by the first four axes only. Other axes are unlikely to be of any importance.



To export your graph see [Export Tee Chart](#).

To print your graph see [Printing and exporting your results](#)

To change graph options, print and export see [Graph display options](#).

To pan and zoom within the graph area see [Zooming and panning on graphs](#)

14.18 Canonical Coef.

This tab gives the canonical coefficients for each of the canonical axes. These are the multiple regression coefficients which are used to calculate the sample (site) score from the environmental data.

Environmental Community Analysis 2.01

File Edit Biological File Environmental File Ordination Variable Selection Monte Carlo Regression Help

Multiple regression of environmental variables on sites scores

These are the multiple regression coefficients and are termed the Canonical Coefficients

	Can. Axis 1	Can. Axis 2	Can. Axis 3	Can. Axis 4	Can. Axis 5	Can. Axis 6	Can. Axis 7
A1	-0.12431	0.264591	-0.745758	-0.566287	0.579271	-0.0707113	-0.463586
Moisture	-0.683977	0.36843	0.47107	-0.0141947	0.382659	-0.710278	0.53673
Manure	0.0313016	0.125651	0.531079	-1.72539	1.97434	0.403249	0.964128
Hayfield	0.0623942	-0.205739	0.23454	-0.418507	0.535087	-0.0673364	0.0222702
Pasture	-0.210232	-0.204218	-0.353888	-0.228341	-0.630095	0.384651	0.367728
SF	-0.207335	0.108869	0.402716	0.12857	-1.2086	0.416257	-0.916616
BF	-0.0767479	-0.0810615	0.33577	-1.29989	0.104555	-0.490629	-0.0301483
NM	-0.391624	-0.795956	0.576567	-1.16101	0.640715	0.817448	0.154328

Correlation of Env. Vars. Binary Environments Biplot scores Canonical Coef. Multicollinearity Monte Carlo

Scatter Plot Variance Species Scores Sample Scores Ordination Plot Rank Plot

Sample/Species Data Working Samples/Species Data Environmental Data Working Environmental data Summary

These results can be printed or exported from the File menu and copied from the Edit menu. See [Export Active Grid](#)

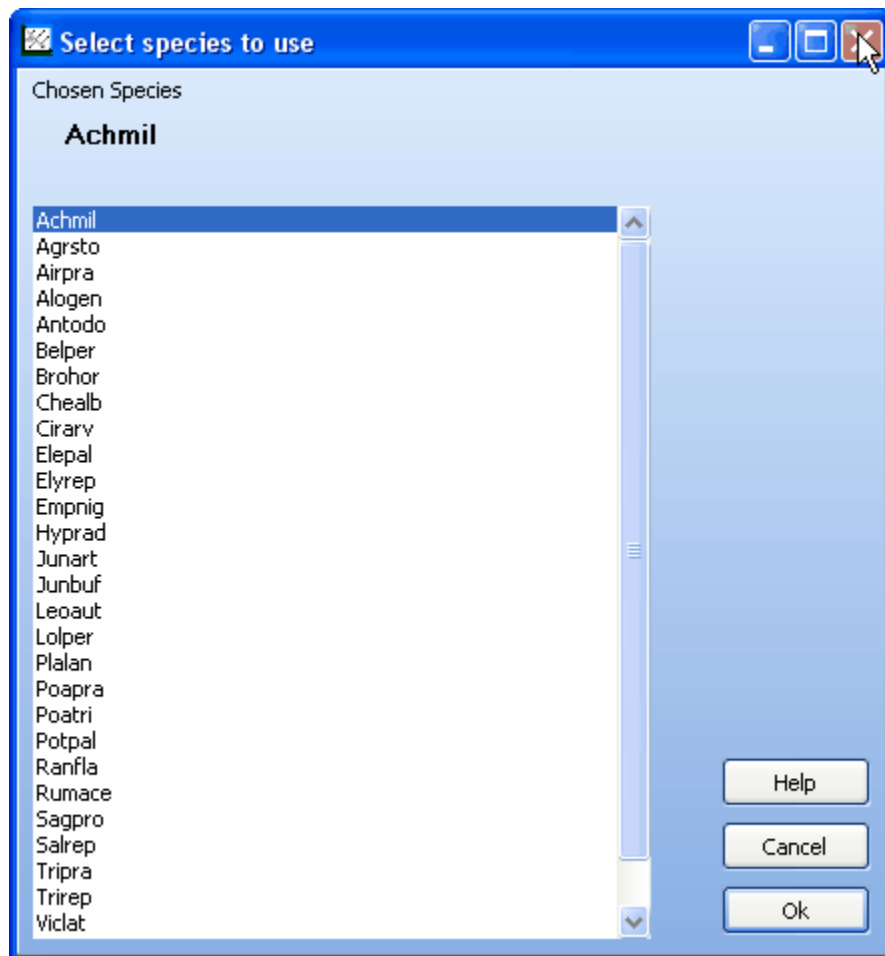
Part



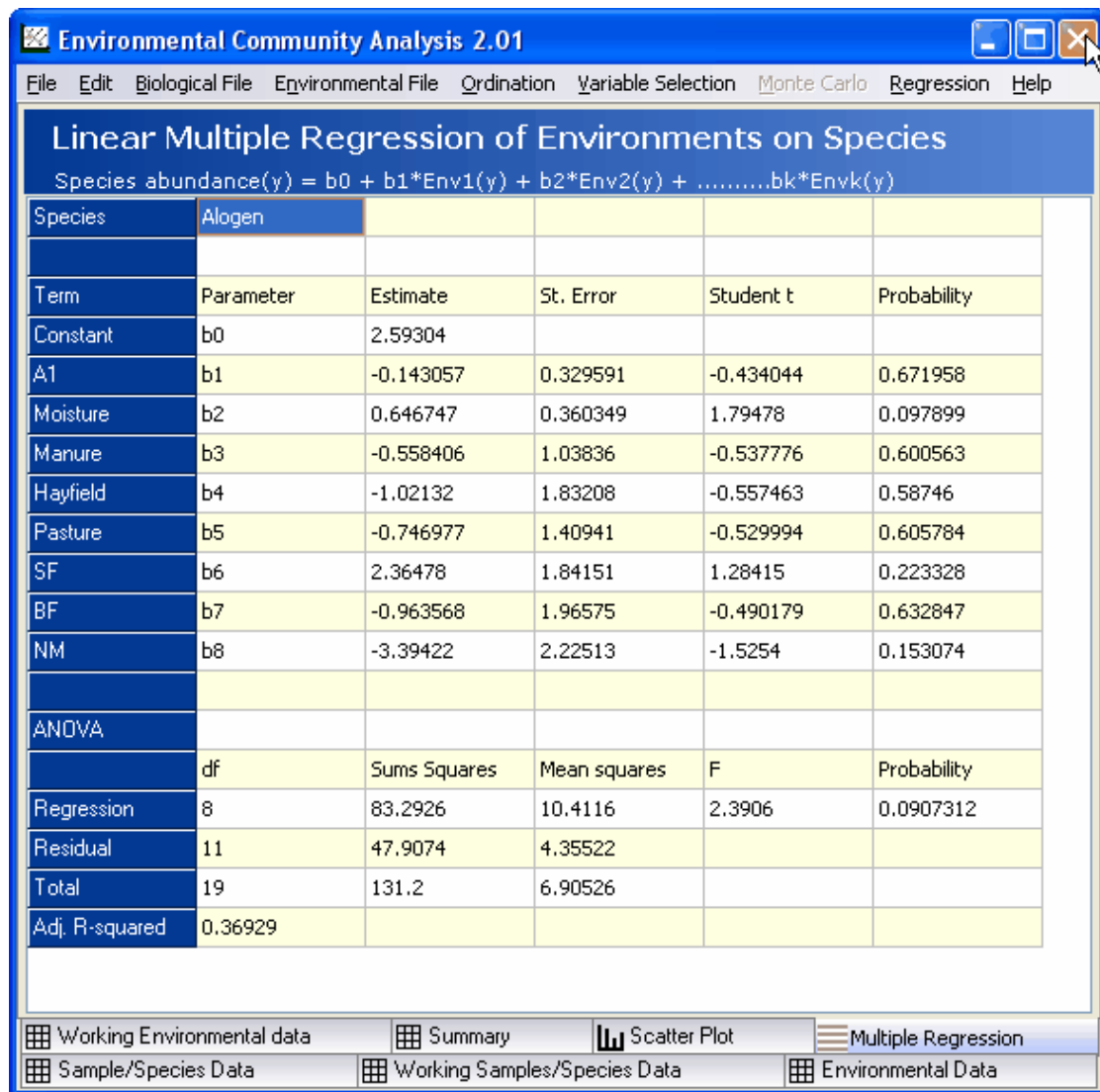
XV

15 Multiple Regression

Select the species against which to run the regression in the lookup box.



The multiple regression results tab is displayed once the Multiple Regression method has been run by selecting Regression|Multiple Regression from the drop-down menus at the top of the main form. This window presents the results of the multiple linear regression analysis of the environmental parameters on the abundance of the selected species. The environmental parameters are the independent variables and the species abundance is the dependent variable.



The fitted equation is of the form:

Species abundance (y) = b0 + b1*Environment 1 + b2*Environment 2 +..... bk*Environment k,

where b0, b1, b2.....bk are the fitted parameters.

The upper part of the output grid gives the estimated parameters, their standard error, the Student t value and its estimated probability. To be significant at the 5% level a parameter should have a probability < 0.05.

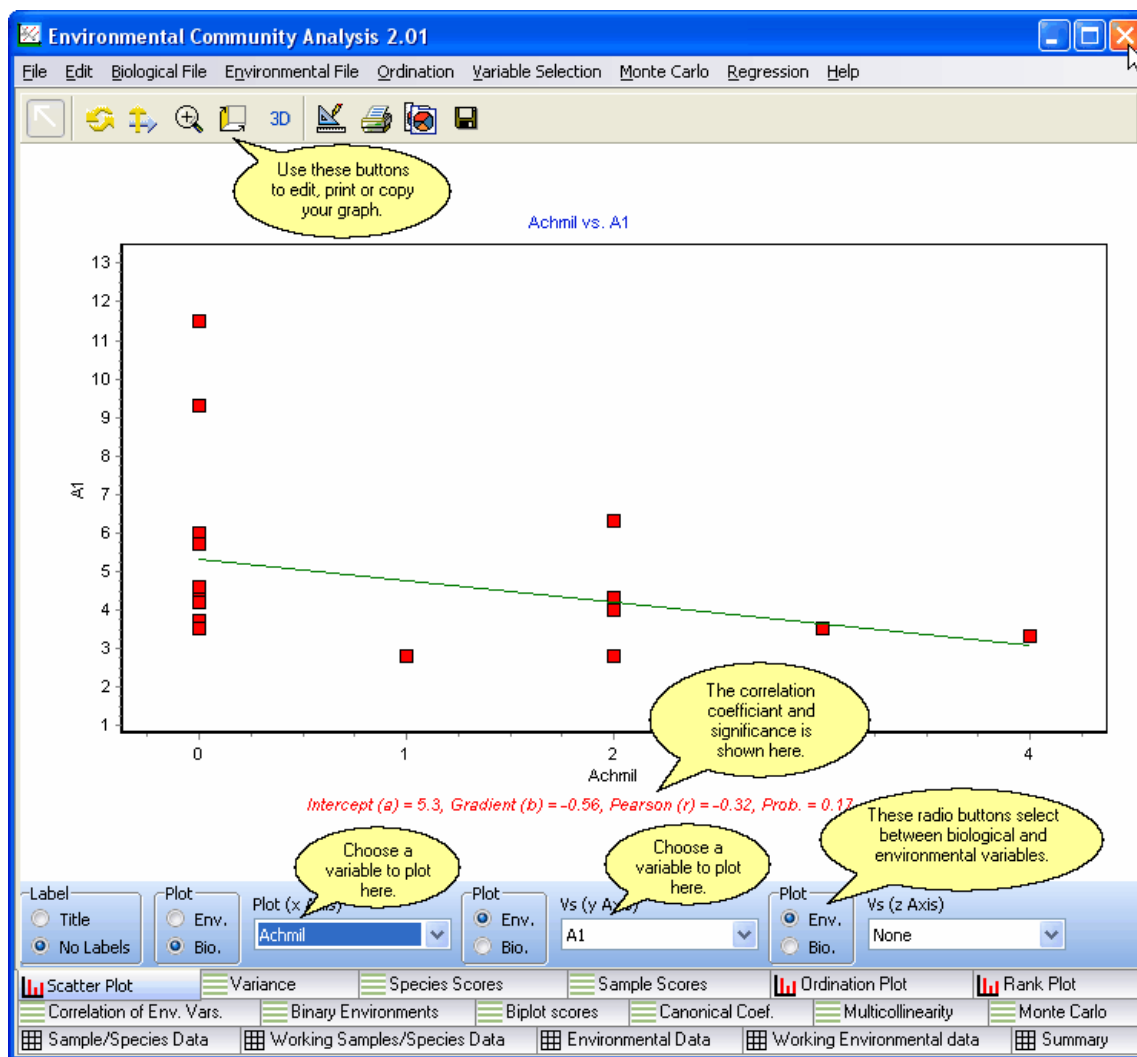
The lower part of the grid gives the Analysis of variance table, the F value and its probability. The regression explains a significant part of the variation in species number at the 5% level if the probability value is < 0.05.

Pearson correlations between any two variables (biological or environmental) can be obtained from the scatter plot tab - see [Scatter Plot](#).

15.1 Scatter Plot

The scatter plot allows you to plot two and three dimensional graphs of any variables from either the biological or environmental working data grids.

The variables for plotting are selected using the panel below the graph. For each axis, radio buttons are offered to allow the selection of variables from either the biological or environmental working array. The variable is selected from the drop-down lists.



A two dimensional graph will be produced if the z axis is set to none.

Note that the labels can be shown or hidden. Here they have been hidden to make the graph more readable.

To export your graph see [Export Tee Chart](#).

To print your graph see [Printing and exporting your results](#)

To change graph options, print and export see [Graph display options](#).

To pan and zoom within the graph area see [Zooming and panning on graphs](#)

Part



XVI

16 Stepwise Multiple Linear Regression

From Version 2 onwards, ECOM features forward and backward stepwise multiple regression.

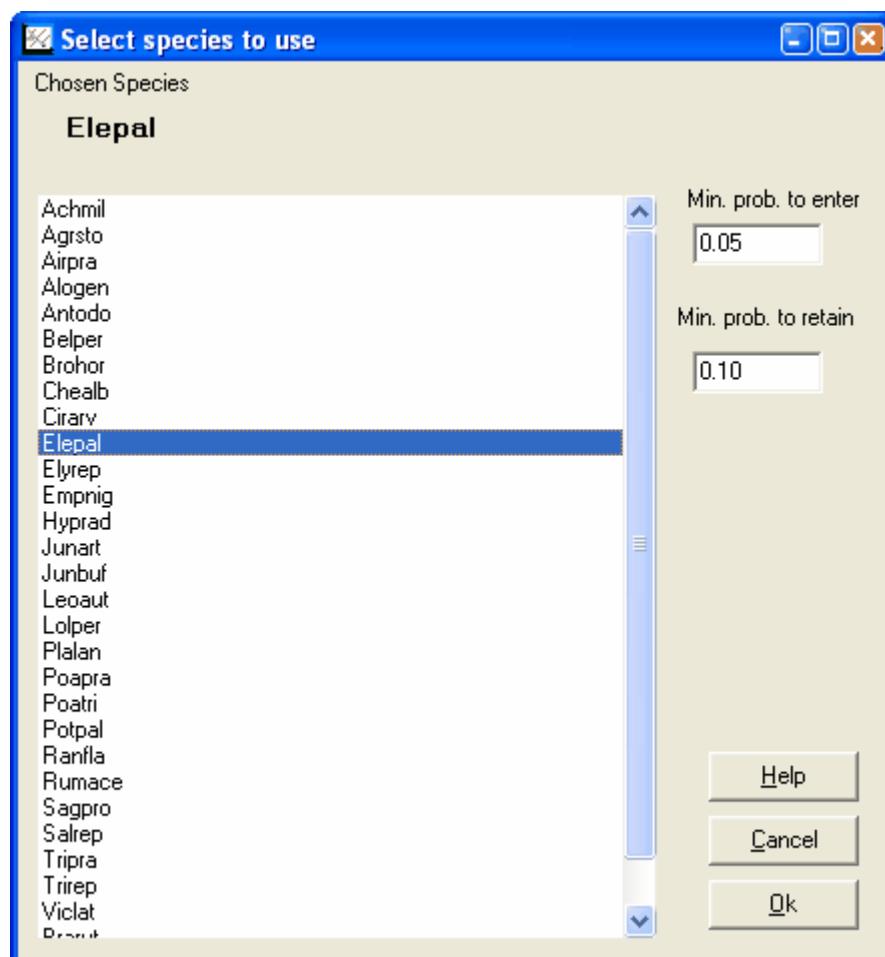
[Forward Stepwise Regression](#)

[Backward Stepwise Regression](#)

Load both your species and your environmental datasets as described and make your selection from the drop-down on the Regression menu.

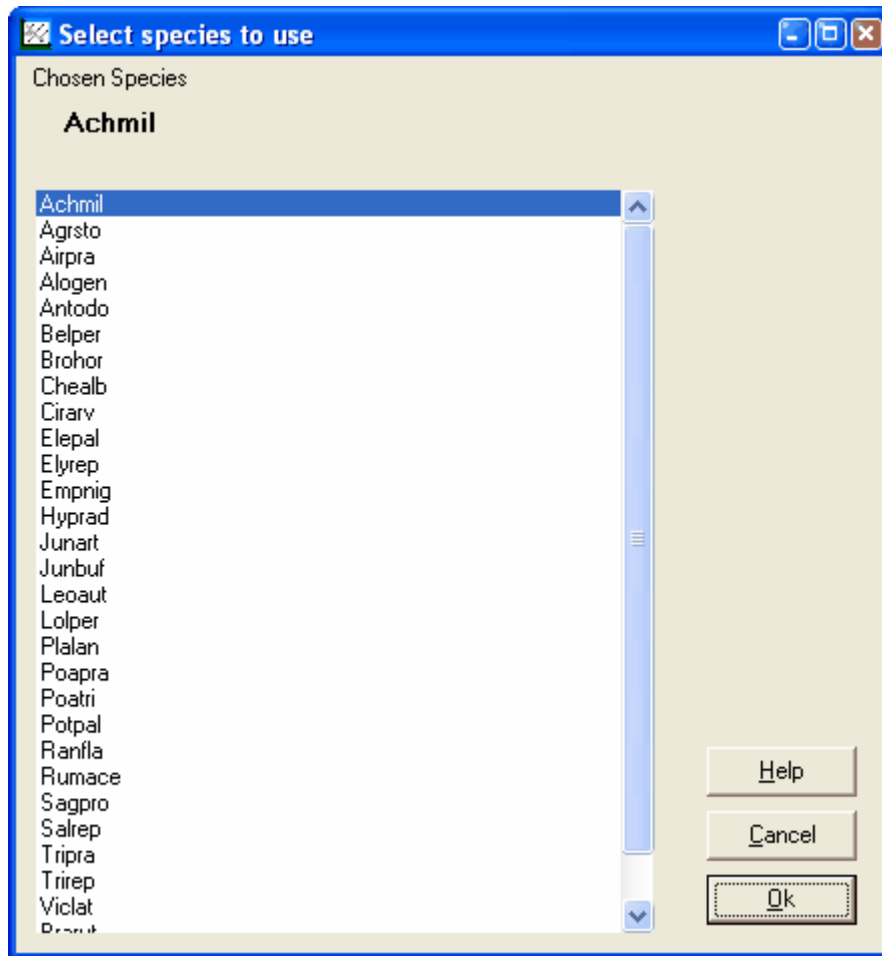
16.1 Variable Selection

When using either [forward](#) or [backward](#) stepwise multiple regression a dialog box is displayed to enable you to select the species against which you wish to test your environmental variables. With forward stepwise regression, the dialog has boxes for you to enter and adjust the minimum probabilities required to enter variables into the model and to remove them again if they have little value (thereby preventing all variables from being entered, and complicating the model).



Click OK to run the analysis and display the results on the Regression tab.

With backward stepwise regression, because variables are removed, rather than being added there is no minimum probability for removal and thus the probability boxes are not displayed.



Click OK to run the analysis and display the results on the Regression tab.

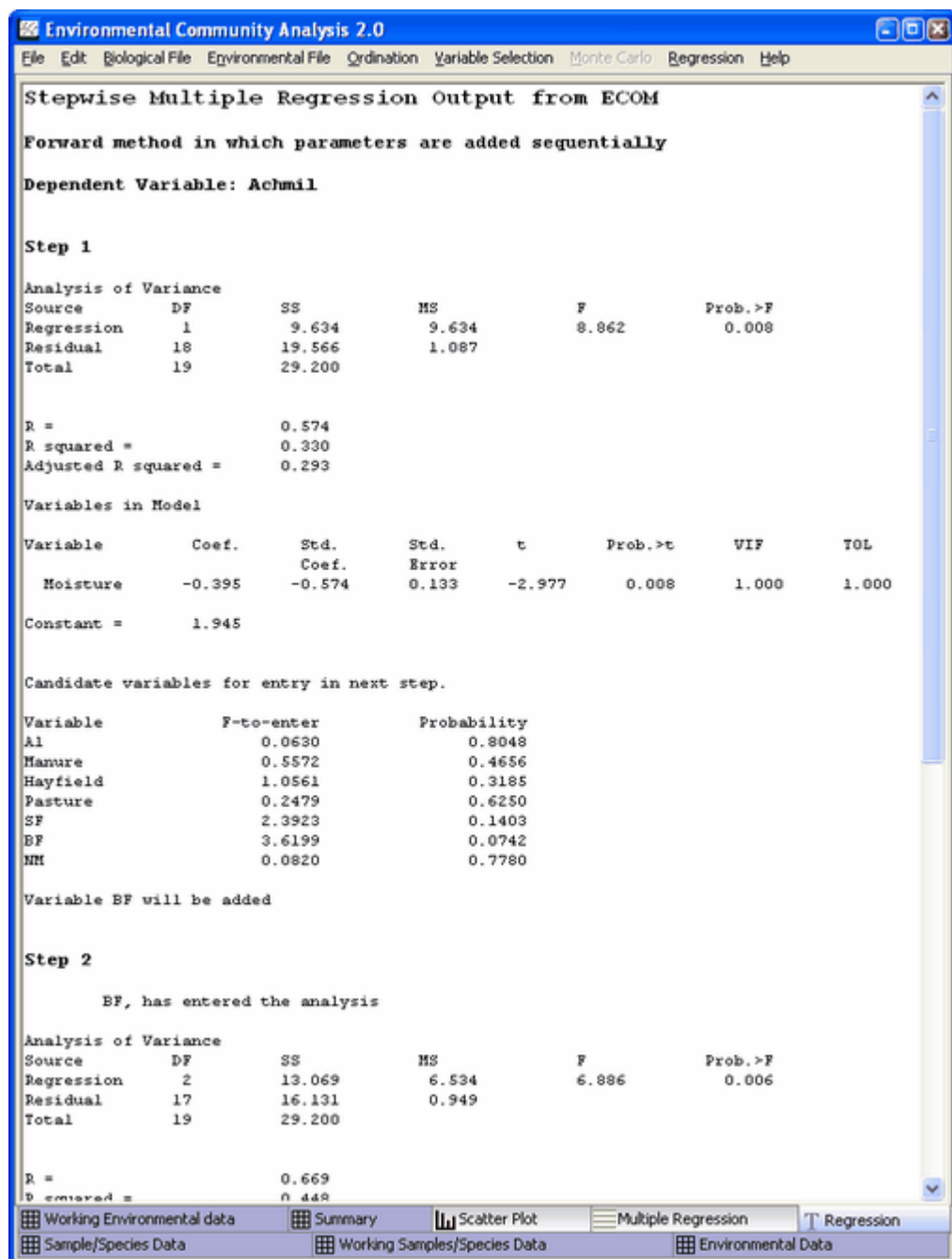
16.2 Foward Stepwise Linear Regression

Forward stepwise regression starts with no variables in the model. It then adds the most significant explanatory variable, that is, the one with the lowest p-value, at each step, until all variables have been added. By scrutinising the overall fit of the model variables will be automatically added (and, if they do not improve the fit, removed again) until the optimum model is found.

Load both your species and your environmental datasets as described in [Opening a Data Set](#)

From the Regression menu on the Toolbar select Forward Stepwise. A list of dependent variables (usually species) is displayed.

The results report shows the sequences of the procedure as Steps:



Step 1 - Shows the effect of introducing the most explanatory variable into the model, and a list of candidate variables that will be entered next, providing they fit the selection criteria.

Step 2 - Shows the effect of introducing the next most explanatory variable into the model, and a list of candidate variables that will be entered next, providing they fit the selection criteria.

...and so on.

In most species/environmental datasets it is unusual for more than three variables to be included in any explanatory model.

A common problem with environmental data sets is that of [multicollinearity](#) between explanatory variables. ECOM automatically checks for multicollinearity and, if detected, will take you to a screen that details the problem and allows you to remove one of the correlated variables.

16.3 Backward Stepwise Linear Regression

Backward stepwise regression starts with all explanatory variables included in the model. It then removes the least significant explanatory variable, that is, the one with the highest p-value, at each step, until all variables have been added. By scrutinising the overall fit of the model variables will be automatically removed until the optimum model is found.

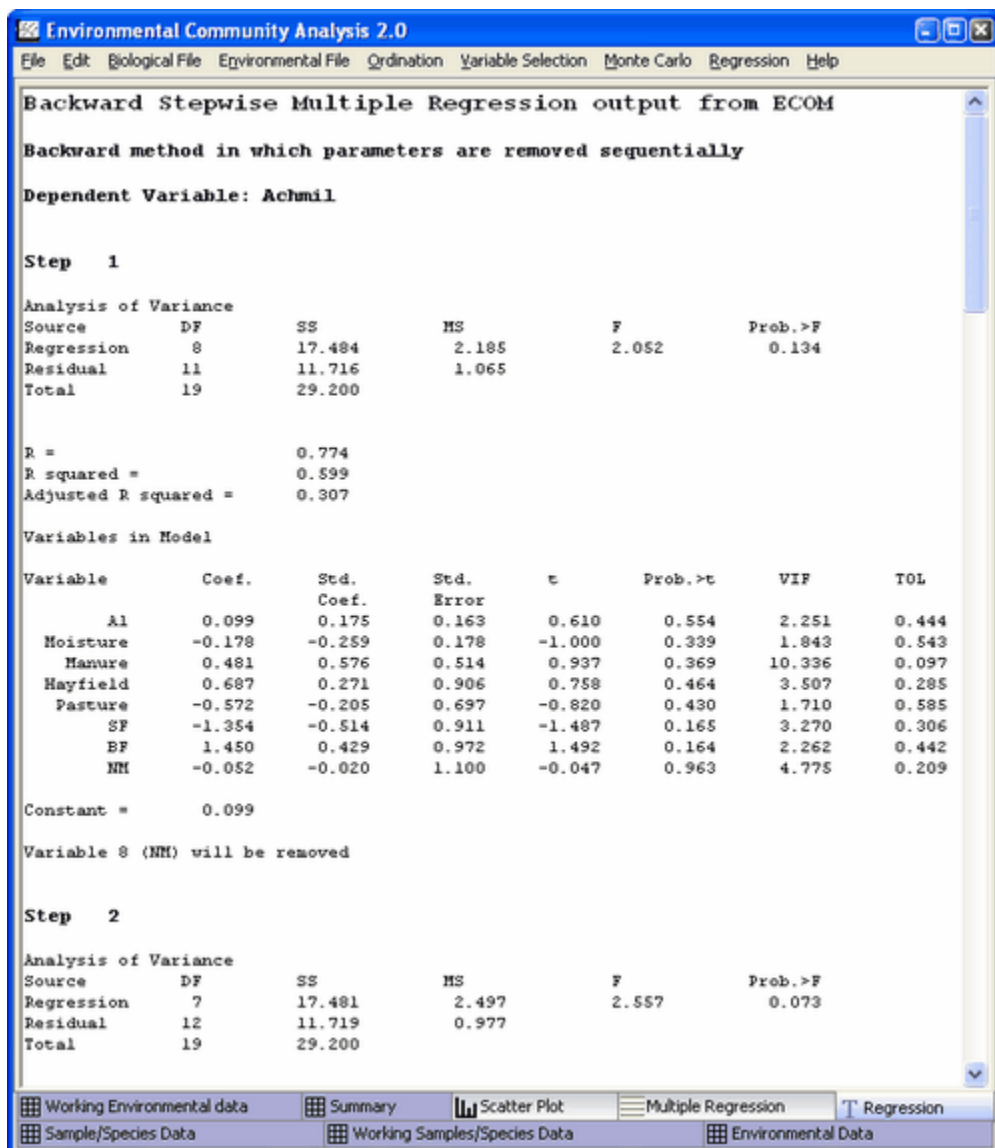
Load both your species and your environmental datasets as described in [Opening a Data Set](#)

From the Regression menu on the Toolbar select Backward Stepwise. A list of dependent variables (usually species) is displayed.

Select the species against which you wish to test your environmental variables.

Click OK to run the analysis and display the results on the Regression tab.

The results report shows the sequences of the procedure as Steps:



Step 1 - Shows the effect of including all the explanatory variables into the model, with the individual p-values.

Step 2 - Shows the effect of removing the least explanatory variable from the model.

...and so on.

In most species/environmental datasets it is unusual for more than three variables to be included in any explanatory model.

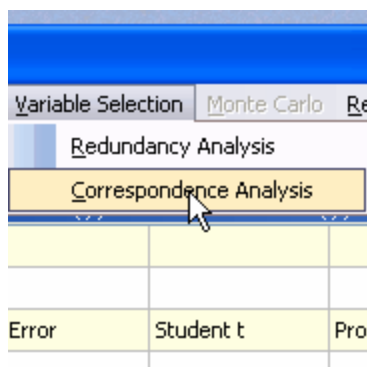
A common problem with environmental data sets is that of [multicollinearity](#) between explanatory variables. ECOM automatically checks for multicollinearity and, if detected, will take to you a screen that details the problem and allows you to remove one of the correlated variables.

Part



17 Variable Selection

The Variable Selection option enables you to use either Redundancy Analysis or Correspondence Analysis to identify the three variables which contribute most to the variability of the response variables and which should therefore be retained in any of the multivariate analyses.

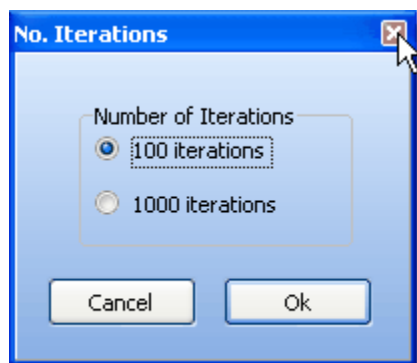


It also performs a randomization and calculates significance levels.

See also:

[Sample/Species Data](#)
[Working Sample/Species Data](#)
[Environmental Data](#)
[Working Environmental Data](#)
[Redundancy Analysis](#)
[Correspondence Analysis](#)

When selecting Redundancy or Correspondence analysis from the Variable Selection menu, you are presented with a dialog from which you can



select the number of iterations to be run. Warning: On a large dataset, selecting 1000 iterations could be somewhat time-consuming.

17.1 Sample/Species Data

This grid holds the main data used in the analyses. It comprises either full quantitative or presence/absence data on the taxa present in the samples.

Environmental Community Analysis 2.01

File Edit Biological File Environmental File Ordination Variable Selection Monte Carlo Regression Help

Sample - species data as loaded from file

C:\Documents and Settings\Peter Henderson\Desktop\ECOM 2007\Data\Dune_species.csv

	site 1	site 2	site 3	site 4	site 5	site 6	site 7	site 8
Achmil	1	3	0	0	2	2	2	0
Agrsto	0	0	4	8	0	0	0	4
Airpra	0	0	0	0	0	0	0	0
Alogen	0	2	7	2	0	0	0	5
Antodo	0	0	0	0	4	3	2	0
Belper	0	3	2	2	2	0	0	0
Brohor	0	4	0	3	2	0	2	0
Chealb	0	0	0	0	0	0	0	0
Cirarv	0	0	0	2	0	0	0	0
Elepal	0	0	0	0	0	0	0	4
Elyrep	4	4	4	4	4	0	0	0
Empnig	0	0	0	0	0	0	0	0
Hyprad	0	0	0	0	0	0	0	0
Junart	0	0	0	0	0	0	0	4
Junbuf	0	0	0	0	0	0	2	0
Leoaut	0	5	2	2	3	3	3	3
Lolper	7	5	6	5	2	6	6	4
Plalan	0	0	0	0	5	5	5	0
Poapra	4	4	5	4	2	3	4	4
Poatri	2	7	6	5	6	4	5	4

Sample/Species Data Working Samples/Species Data Environmental Data Working Environmental data Summary Scatter Plot

See [Opening a data set](#).

17.2 Working Samples/Species Data

This form allows the user to make a variety of changes to the raw sample/species data prior to undertaking an analysis. Initially, the user will be presented with a grid filled with the raw data; this can be adjusted using the options in the panel below the data grid.

Type of Adjustment

☒ Transform
☐ Relative
☐ Handling zeros
☐ Transpose

Reload Raw Data
 Log 10
 Log 10 (x+1)
 Log e
 Log e (x+1)
 Square Root

This will cause the working data to revert to the untransformed raw data.

Selected Variable
 -All-

Reload Raw Data Submit

See the subjects below for further details of transformations etc

[Data transformations](#)

[Relative adjustments](#)

[Dealing with zeros](#)

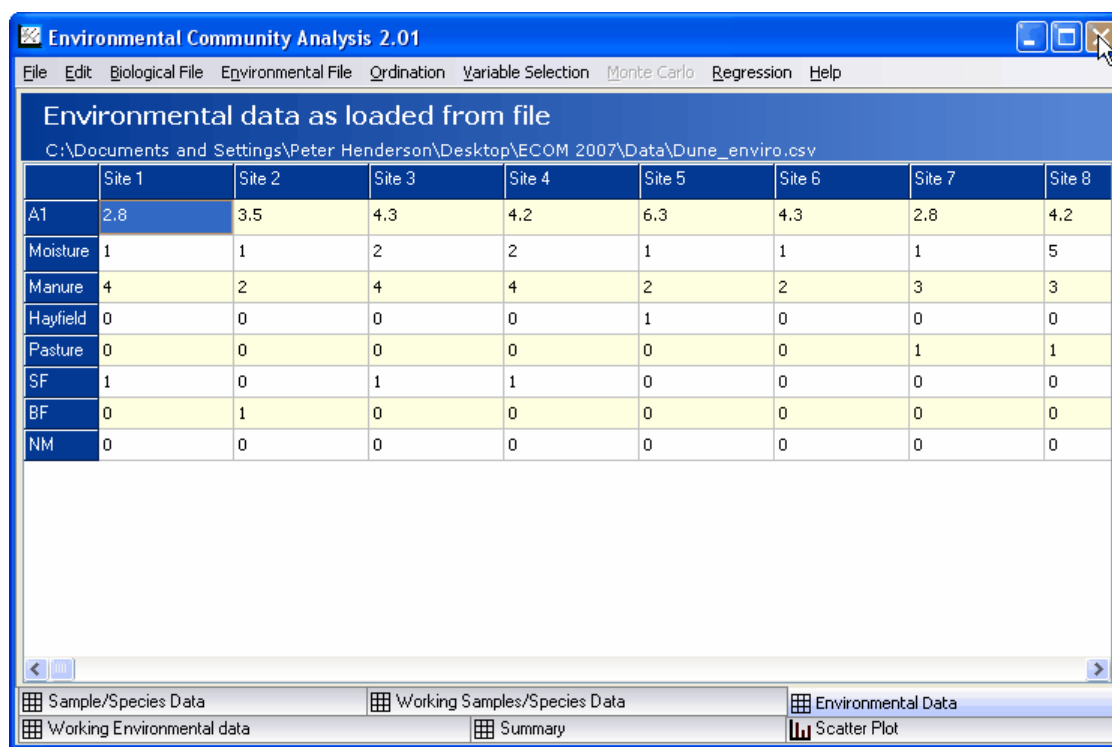
[Transposing data](#)

This form can also be used to resolve any mismatch between the samples in the Biological and Environmental arrays.

Also see [Opening a data set](#)

17.3 Environmental Data

This grid holds the environmental data used in the analyses. It comprises either full quantitative or binary data on the environmental (physical) variables for each site. You can mix quantitative and semi-quantitative and binary variables.



Environmental data as loaded from file
C:\Documents and Settings\Peter Henderson\Desktop\ECOM 2007\Data\Dune_enviro.csv

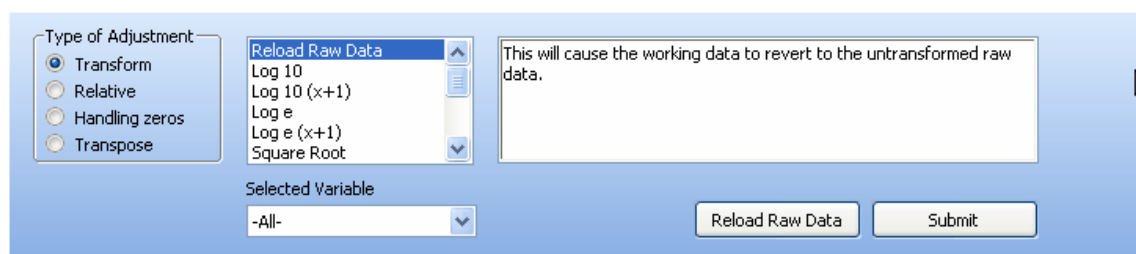
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8
A1	2.8	3.5	4.3	4.2	6.3	4.3	2.8	4.2
Moisture	1	1	2	2	1	1	1	5
Manure	4	2	4	4	2	2	3	3
Hayfield	0	0	0	0	1	0	0	0
Pasture	0	0	0	0	0	0	1	1
SF	1	0	1	1	0	0	0	0
BF	0	1	0	0	0	0	0	0
NM	0	0	0	0	0	0	0	0

Sample/Species Data Working Samples/Species Data Environmental Data
Working Environmental data Summary Scatter Plot

See [Opening a data set](#)

17.4 Working Environmental Data

This form allows the user to make a variety of changes to the raw Sample/Environmental variables data prior to undertaking an analysis. Initially, the user will be presented with a grid filled with the raw data; this can be adjusted using the options in the panel below the data grid.



Type of Adjustment

- ☒ Transform
- ☐ Relative
- ☐ Handling zeros
- ☐ Transpose

Selected Variable: -All-

Reload Raw Data

Submit

This will cause the working data to revert to the untransformed raw data.

See [Opening a data set](#)

See the subjects below for further details of transformations etc

[Data transformations](#)

[Relative adjustments](#)

[Dealing with zeros](#)

[Transposing data](#)

17.5 Redundancy Analysis

Redundancy Analysis is an extension of multiple regression when there is more than one response (species) variable. If there is only one species then RA reduces to a simple multiple regression. It was first proposed by Rao (1964) but has been little used by biologists. A good description of the procedure is to be found in [Legendre & Legendre \(1998\)](#).

The input data comprises a biological data set of samples/species and another set of environmental/explanatory variables. The results are normally presented as a biplot.

[Working samples/species data](#)
[Sample/species data](#)
[Environmental data](#)
[Working environmental data](#)
[Eigenvalues](#)

17.6 Correspondence Analysis

Canonical Correspondence Analysis (CCA) is an ordination method widely used among ecologists. This method allows the analysis of a species abundance matrix with respect to a second table of explanatory environmental variables. Both the species and environmental variable tables must refer to the same samples and must therefore have the same number of samples. CCA is a correspondence analysis of the species data using linear constraints derived from a multiple linear regression of the environmental variables against the species scores. The result is an ordination that shows the relationship between the species, sites and environmental variables. It is termed a constrained ordination method as it is an extension of the correspondence analysis with the environmental data acting as the constraint.

CCA is suited to community data sets where: (1) species responses to environmental variables are unimodal (hump-shaped), and (2) the important underlying environmental variables have been measured. If you have not measured the most important environmental variables, or your environmental variable measurements are subject to large errors, CCA will not yield satisfactory, or believable, results.

CCA is currently one of the most popular ordination techniques in community ecology. It is, however, one of the most dangerous in the hands of people who do not take the time to understand this relatively complex method. The dangers lie principally in several areas: (1) Because it includes multiple regression of community gradients on environmental variables, it is subject to all of the hazards of multiple regression. These are well documented in the statistical literature, but often not fully appreciated by newcomers to multiple regression. (2) As the number of environmental variables increases relative to the number of observations, the results become increasingly dubious, even though an appearance of very strong relationships is inevitable. (3) Statistics indicating the "percentage of variance explained" can be calculated in several ways, each for a different question, but users frequently confuse these statistics when reporting their results.

CCA does not explicitly calculate a distance matrix. But CCA, like CA and PCA, is implicitly based on the chi-squared distance measure where samples are weighted according to their totals. This gives high weight to species whose total abundance in the data matrix is low, thus exaggerating the distinctiveness of samples containing several rare species.

Part



18 Printing and saving results

Output can be saved as a file, copied to the clipboard or printed. See topics below for further details.

[Printing text and grid output](#)

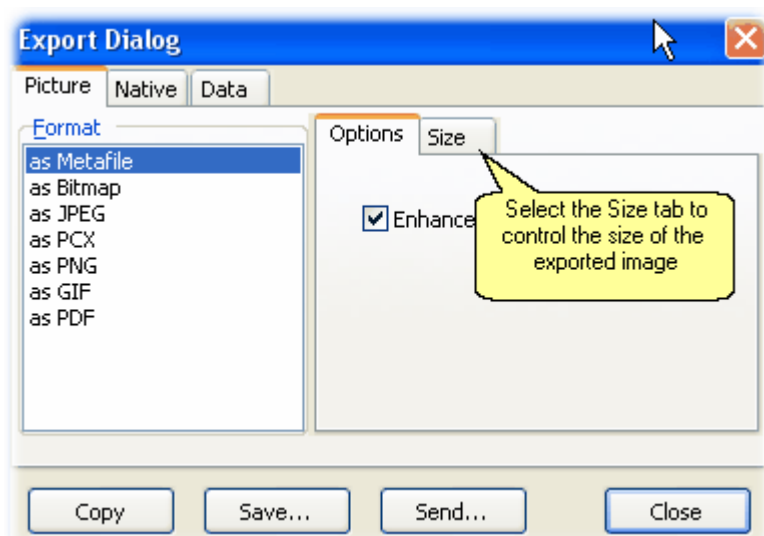
[Preparing charts for output](#)

[Printing charts](#)

[Exporting charts](#)

18.1 Exporting charts

To export the image of a chart select **File: Export**. This will open the Export Dialog.

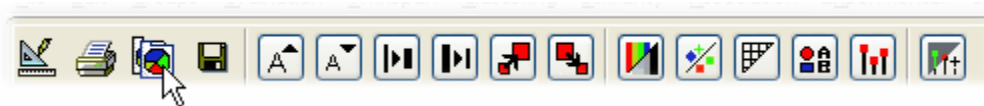


The chart can be saved in a number of different file formats: Enhanced Metafile (*.emf), Bitmap (*.bmp), JPEG (*.jpg), PCX, PNG, GIF or Native (*.tee). Each file format has advantages and disadvantages.

- 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.
- GIFs are also compressed files useful for web sites; they have a considerably smaller colour palette than JPEGs or BMPs; 256 colours, as opposed to many millions. This means that while a JPEG or BMP image can show a smooth gradation of colour (for instance in a graded background), saved as a GIF image, it will appear broken up into jagged zones of colour. GIF images are therefore better suited to images showing large discrete blocks of single 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 available from the [Pisces website](#). Tee files tend to be very small indeed; often less than 1KB.

This export dialog can also be used to export the data shown on the chart.

You can also copy a chart to the Windows clipboard using the Copy button on the graphics toolbar:

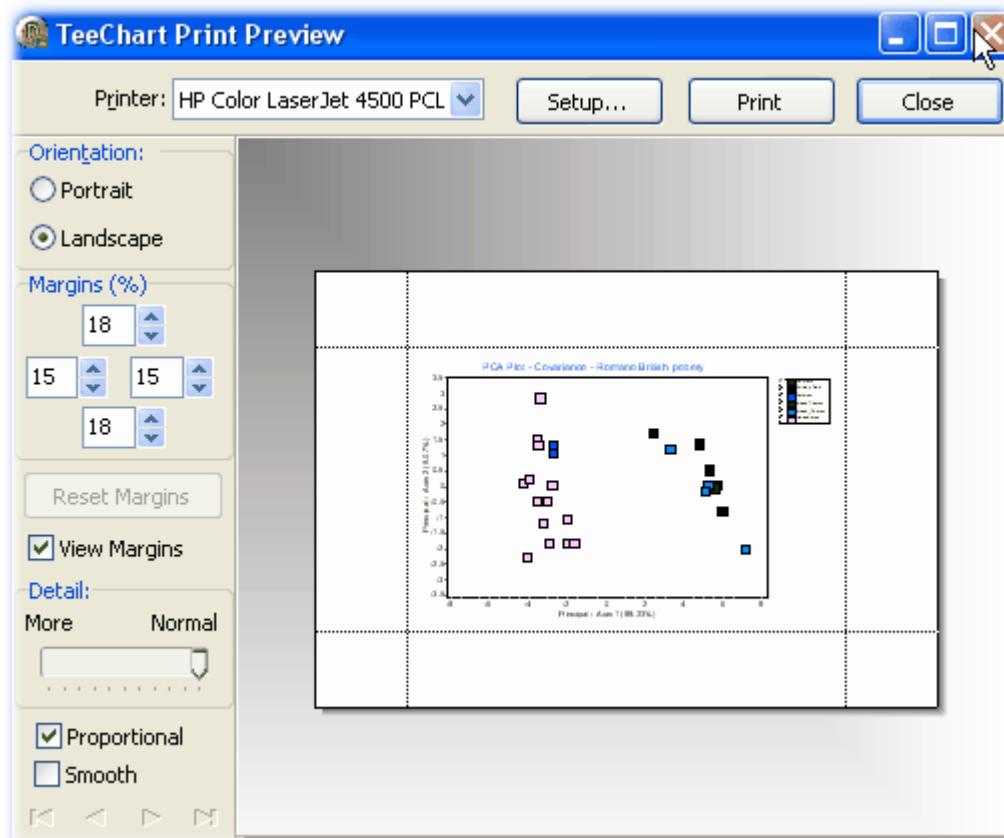


18.2 Printing charts

The graphs and other results can be printed using **File: Print**, or the **Print button** on the toolbar.



The Print Preview option will show the page layout and allow image size, margins, and paper orientation to be changed.



If you have the full version of Adobe Acrobat (**not** the free Acrobat Reader) 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.

18.3 Preparing charts for output

The graph option buttons on the Chart Toolbar 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.

Rotate - (Two arrows in a circle) Use this button to rotate the graph. Click the button then move onto the graph and hold down the left-hand mouse button and move the mouse.

Move - (Two arrows at right angles) Use this button to move the graph. Click the button then move onto the graph and hold down the left-hand mouse button and move the mouse.

Zoom - (Magnifying glass) Use this button to zoom in onto the graph. Click the button then move onto the graph and hold down the left-hand mouse button and move the mouse. This feature is very sensitive to mouse movements and some practice may be required. See [Zooming and panning on graphs](#).

Depth - (3D box) Use this button to change the depth shown on the graph. Click the button then move onto the graph and hold down the left-hand mouse button and move the mouse.

3D - Use this to toggle the 3D view on and off.

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, and even to email it using the 'Send' button. For more information on chart editing use the Tee Chart help system available from the Help button on the chart edit box.

Print - Use this button to print the graph

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

Save - Save the file in a variety of different formats.

Increase font - This will increase the font size of the chart titles.

Decrease font - This will decrease the font size of the chart titles.

Increase line thickness - This increases the thickness of plotted lines such as Eigenvectors.

Decrease line thickness - This decreases the thickness of plotted lines.

Increase symbol - Increase the size of a plotted point.

Decrease symbol - Decrease the size of a plotted point.

Change between colour/grey scale - Change between a colour and greyscale plot.

Change symbol set - Change the plotting symbols.

Add grid - Add graph grid lines to the plot.

Add legend - Add legend to the plot. N.B. This legend can be used to select groups for plotting.

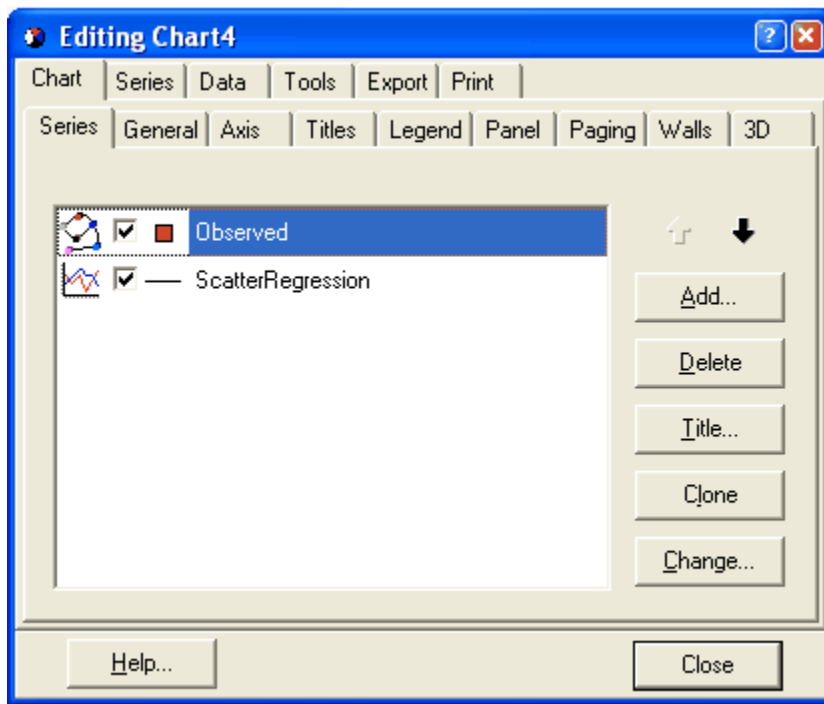
Add stalks - Add stalks to points - useful for 3D plots.

Select theme - Use to select a particular style of graph, for example black background.

Moving Chart Labels

Click the Edit button to change a wide range of aspects of the graph, add titles, and use various tools to customise the chart. One of the most useful tools allows you to drag the chart labels - very useful if you have a cluster of points with overlapping labels. Click the Edit button, then the Tools tab of the Edit dialog, the Add... button, and choose 'Drag Marks'. When the chart is displayed, you will find that the cursor displays as a hand symbol on hovering over a data label, allowing you to drag the label to the desired position.

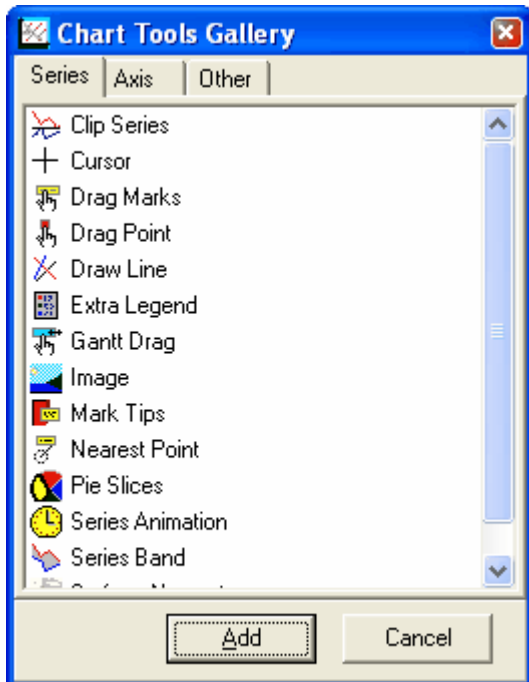
NOTE: any alteration of other aspects of the graph will cause the labels to return to their original position. Therefore, moving the labels should be the last alteration you make before printing or exporting the graph.



Drawing Lines

You may find it useful to draw lines on a chart, to separate groups or clusters of points.

- Click the Chart Edit button, choose the Tools tab - Add... button, and from the list of tools, choose Draw Line. If you click on the Pen button, you can specify the type and colour of the line.

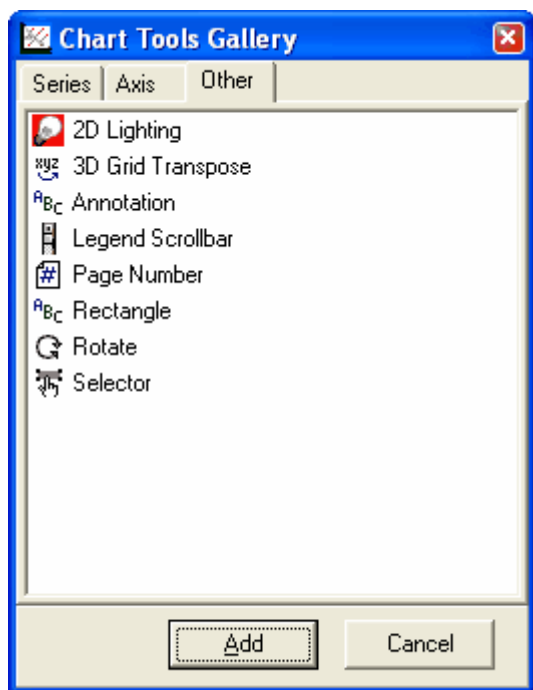


- When you have made your choice, click 'Close' to return to the program, where you can simply draw lines with the mouse. Lines, once drawn, can be dragged to the position required.
- To alter the line type or colour, open the Tools tab again, click on Draw Line, and make the changes you require.
- To remove lines from a plot, go to the Tools tab, click on Draw Line, and press the Delete button.

Adding Annotations

Annotations can be very useful if, for instance, there are a large number of points on your chart and you wish only to label a few of them. To add annotations:

- With the chart showing, click on the set-square icon to show the editing screen, select the Tools tab; click Add, and from the 'Other' tab on the Chart Tools Gallery screen, select Annotation from the list of tools.

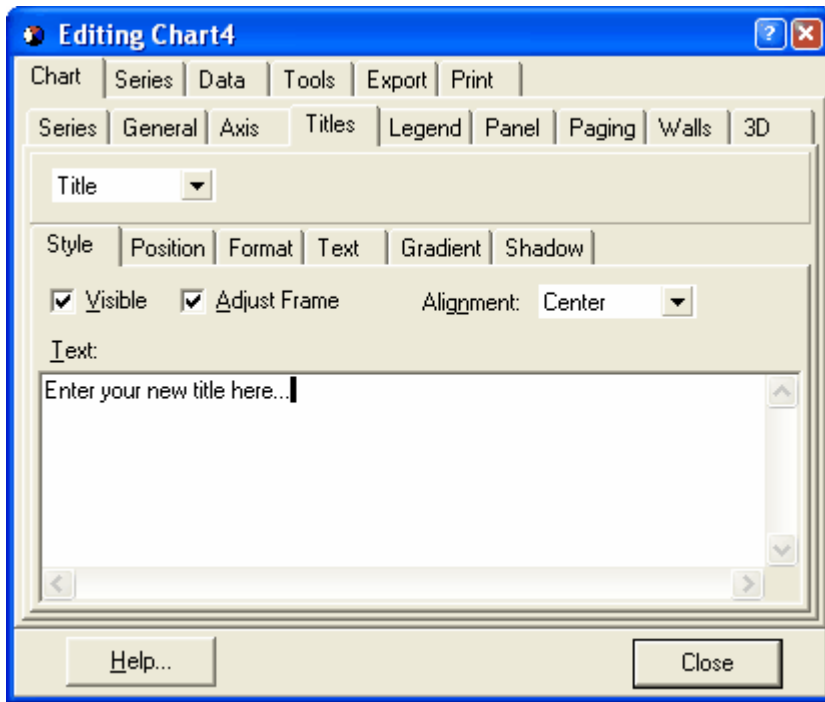


Enter the text you want on your annotation into the text box on the Options tab, and then use Position: Custom to manoeuvre the label into the desired position. Use the options on the other tabs to change font, shading, outline and other details.

To add another annotation, click Add, and select Annotation again.

Altering chart titles

- With the chart showing, click on the set-square icon to show the editing screen. Click on the Titles tab (on the Chart tab) to show the following screen:

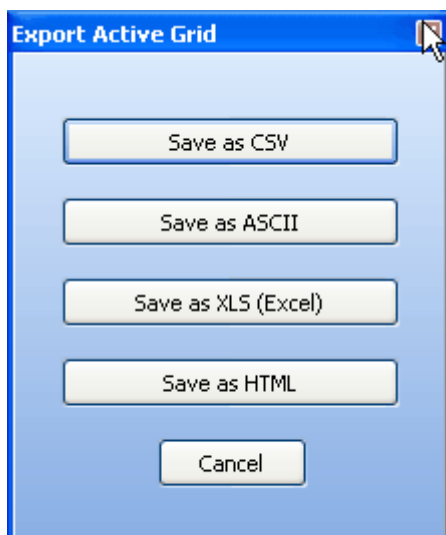


Enter the new title text into the box and click 'Close' to save.

Note that there are very many options on the Edit screen and it is worth a few minutes' experimentation to see what sort of effects are available to enhance your charts for publication or display.

18.4 Exporting text and grid output

To save text and grid results to file choose File|Export and follow instructions in the dialog box.



Any active window can be printed using **File|Print** or exported to other programs using **File|Export**, and data or graphs can be copied to the clipboard using **edit|copy** in the normal Windows

fashion.

Copies can be made to the Windows clipboard or to a file by selecting **Edit|Copy** or, in the case of graphs, by clicking on the Copy button above the graph. Images of graphs can be stored as bitmaps, metafiles, enhanced metafiles or TEE files (see [Export Tee Chart](#)). For Windows users who wish to include a graph in another document the enhanced metafile format is to be preferred as it will allow the image to be changed in size without loss of resolution of the text.

18.5 Graph display options

Many aspects of your graphs can be altered and the results printed or exported using the series of buttons arranged above the graphs.



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, and even to email it using the 'Send' button. For more information on chart editing use the TeeChart help system available from the Help button on the chart edit box.

Print - Use this button to print the graph

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

Save - Save the file in a variety of different formats.

Increase font - This will increase the font size of the chart titles.

Decrease font - This will decrease the font size of the chart titles.

Increase line thickness - This increases the thickness of plotted lines such as Eigenvectors.

Decrease line thickness - This decreases the thickness of plotted lines.

Increase symbol - Increase the size of a plotted point.

Decrease symbol - Decrease the size of a plotted point.

Change between colour/grey scale - Change between a colour and greyscale plot.

Change symbol set - Change the plotting symbols.

Add grid - Add graph grid lines to the plot.

Add legend - Add legend to the plot. N.B. This legend can be used to select groups for plotting.

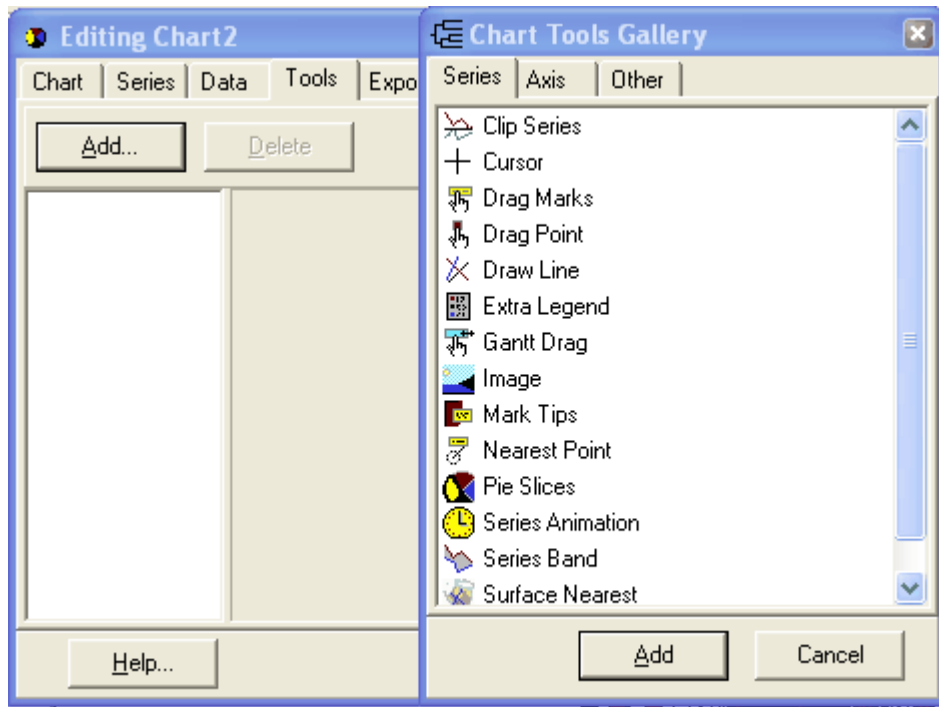
Add stalks - Add stalks to points - useful for 3D plots.

Select theme - Use to select a particular style of graph, for example black background.

Moving Chart Labels

Click the Edit button to change a wide range of aspects of the graph, add titles, and use various tools to customise the chart. One of the most useful tools allows you to drag the chart labels - very useful if you have a cluster of points with overlapping labels. Click the Edit button, then the Tools tab of the Edit dialog, the Add... button, and choose 'Drag Marks'. When the chart is displayed, you will find that the cursor displays as a hand symbol on hovering over a data label, allowing you to drag the label to the desired position.

NOTE: any alteration of other aspects of the graph will cause the labels to return to their original position. Therefore, moving the labels should be the **last** alteration you make before printing or exporting the graph.



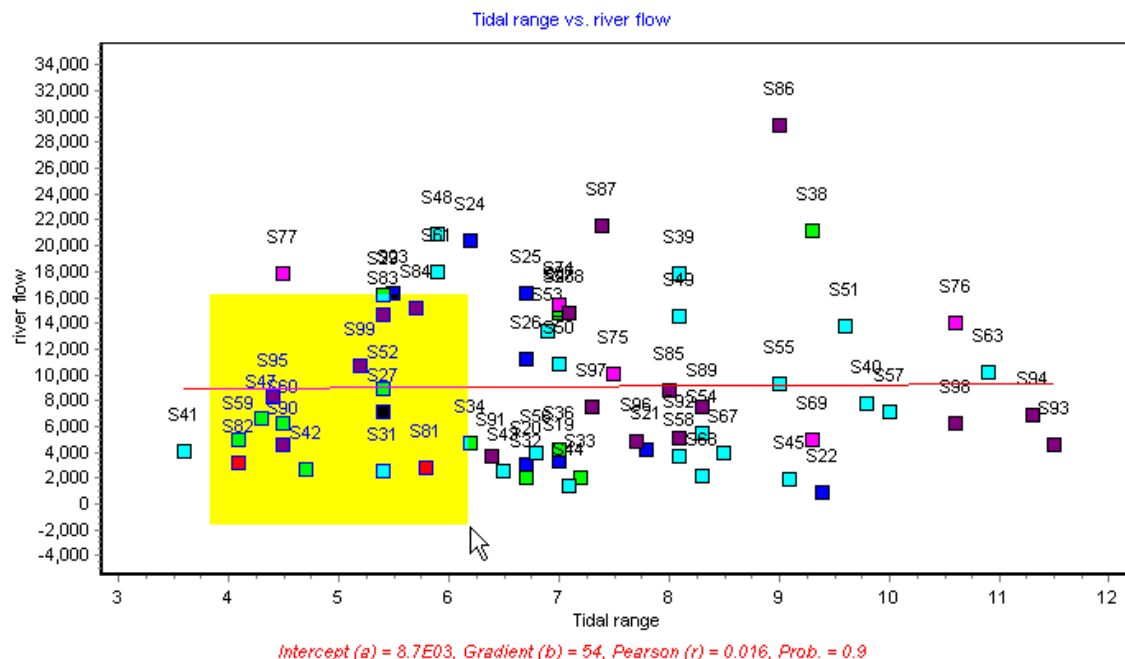
18.6 Tips for improving graphical output

It frequently happens that the default ordination graph will have overlapping labels or labels that overlap the axes. To stop the labels overlapping the axes enter the [graph options menu](#) and manually select the maximum and minimum to be a greater or lesser value than the default for each axis. It is sometimes possible to reduce the overlap of labels by using the 3D options. Alternatively, it may be useful to display the ordination from different directions so that different labels lie on top. Generally it is best to avoid large species or sample names. For large figures consider labeling the species 1, 2, 3 etc. and putting a key in the caption.

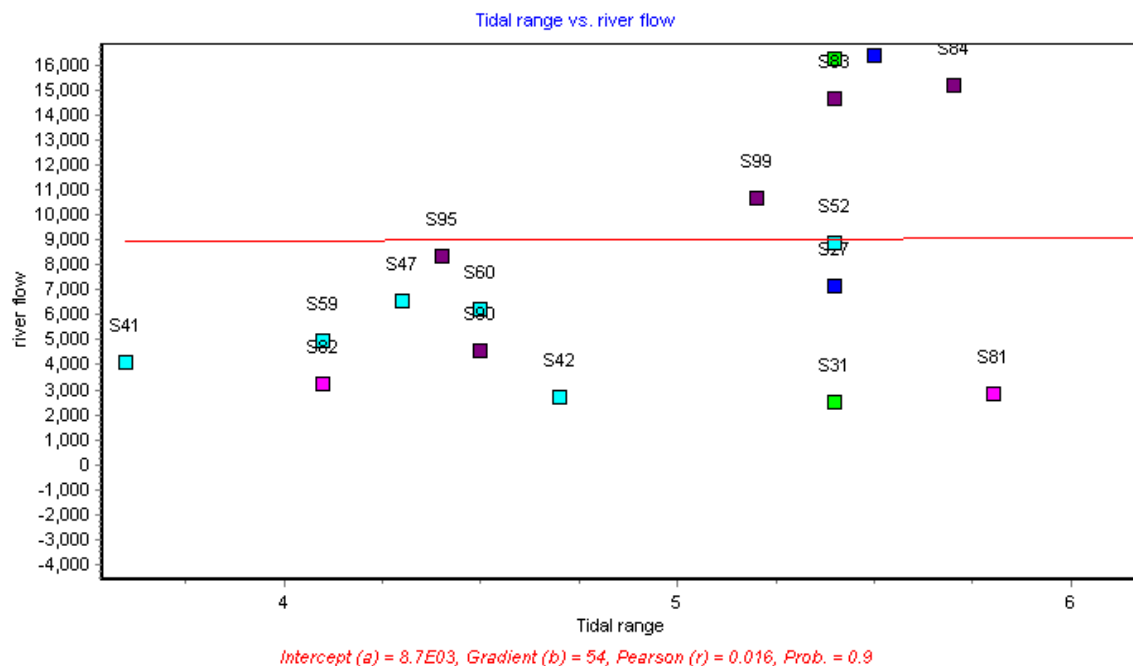
18.7 Zooming and panning on graphs

Tight clusters of points which cannot be differentiated can occur. To zoom in on an area, select the normal mode button (left hand button with arrow icon) then move to the top left corner of the area to be enlarged, then hold down the left hand mouse button and drag to the lower right hand corner of the area you require, and release the button. An enlarged view of the selected area will be displayed.

As you drag down from upper left to lower right corner of the area to be enlarged the area will appear as a pale yellow panel on the plot.



When you release the button this pale yellow area will fill the plot. Using this zoom facility can allow you to distinguish between labels of tightly-clustered points.



To return to the original view, hold down the left hand mouse button and move upwards from the lower right to the upper left corner and release.

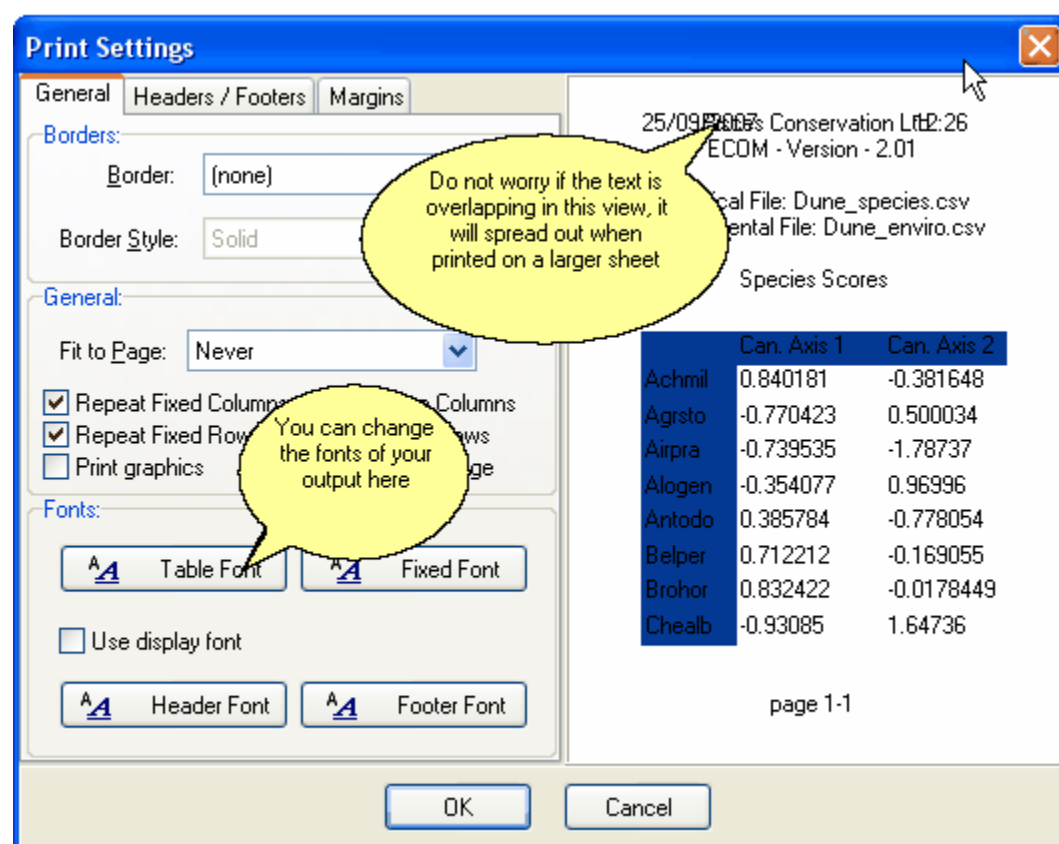
To pan the graph hold down the right hand mouse button and move the mouse.

Part



19 Printing text and grid output

With an output grid showing simply choose **File|Print** and a dialog box for printing options will be activated.



Note that in the dialog, if the header text is too long for the preview panel, it will be compressed and words may overlap. This is simply so that you can ascertain that all the text you require is actually there. When printed, it will be formatted correctly.

If you wish to make adjustments to the text and text positions in the header or footer, click on the headers/footers tab and make changes in the boxes and drop-down lists available.

Print Settings

General Headers / Footers Margins

Date / Time:

Date Position: Top Left

Date Format: dd/mm/yyyy

Time Position: Top Right

Title:

Title Position: Top Center

Title Text: Pisces Conservation Ltd
ECOM - Version - 2.01

Page Numbers:

Pages Position: Bottom Center

Prefix: page Separator: / Suffix:

25/09/2007 Pisces Conservation Ltd:30
ECOM - Version - 2.01

Biological File: Dune_species.csv
Environmental File: Dune_enviro.csv

Species Scores

	Can. Axis 1	Can. Axis 2
Achmil	0.840181	-0.381648
Agrsto	-0.770423	0.500034
Airpra	-0.739535	-1.78737
Alogen	-0.354077	0.96996
Antodo	0.385784	-0.778054
Belper	0.712212	-0.169055
Brohor	0.832422	-0.0178449
Chealb	-0.93085	1.64736

page 1-1

OK Cancel

Numerous options are again available and experimentation will pay dividends.

Part



XX

20 Programs for ecologists by PISCES Conservation Ltd

Software which allows you to quickly discover patterns and perform your analysis.

All our programs feature

- Designed to make difficult statistics easy.
- Easily import and export data.
- Proven high quality graphs.

More details of our programs, free downloads, FAQ and Technical Support information and much more can be found on our website at: www.pisces-conservation.com

Programs

[CAP - Community Analysis Package](#)

A fully-featured analytical package ideal for a very wide range of analytical techniques. CAP guides you through the complexities of multivariate statistics, with its 'Instant Assist' feature and a very extensive Help system that contains all the background to the methods that you could need. Includes a helpful worked example too! Buy bundled with ECOM for a substantial discount!

[ECOM](#)

The sister program to CAP, it complements CAP in providing analytical techniques to detect, visualise and order relationships from both species data and environmental variables. Used where you have environmental data as well as species data. Principal methods: Canonical Correspondence Analysis (CCA), Redundancy Analysis (RDA), Multiple Regression (MR) and Forward and Backward Stepwise Regression. (Discounts are available for buying CAP & ECOM as a bundle).

[Species Diversity and Richness IV](#)

A program to calculate and compare species diversity, estimate total species richness and study patterns of abundance.

[Axis](#)

For handling circular statistics. Essential in behavioural studies, biogeography and geology.

[Fuzzy Grouping](#)

Fuzzy Grouping offers two main methods of data analysis, firstly, fuzzy c-means and secondly, fuzzy ordination, for dealing with imprecise data and likelihood of group membership.

[Population Estimation by Removal Sampling](#)

A program to calculate the statistics from removal trapping experiments.

[Growth II](#)

To analyse growth curves from length and/or weight data - a major update to the program previously known as Simply Growth.

[Simply Probit](#)

A program for estimating lethal or environmental concentrations using toxicity data.

[Simply Tagging](#)

For estimating population size of both closed and open populations with

mark-recapture methods. Includes data simulation options to model experiments for open and closed populations.

[Density from Distance](#)

Offering a range of analytical techniques commonly used by ecologists to estimate animal and plant density from measurements of the distance between objects or from a selected line or point to the objects.

[Dynamica](#)

Dynamica allows you to explore how animal populations and communities change through time.

The program uses a long-term data set from the Severn Estuary, England. Since 1980 fish and crustacean samples have been collected from Hinkley Point 'B' power station. Data are presented on every fish and most large crustaceans known from the estuary, producing an exceptional data set for those interested in population dynamics.

[Hedgerow Assistant](#)

The reliable way to survey hedgerows under the 1997 Hedgerows Regulations.

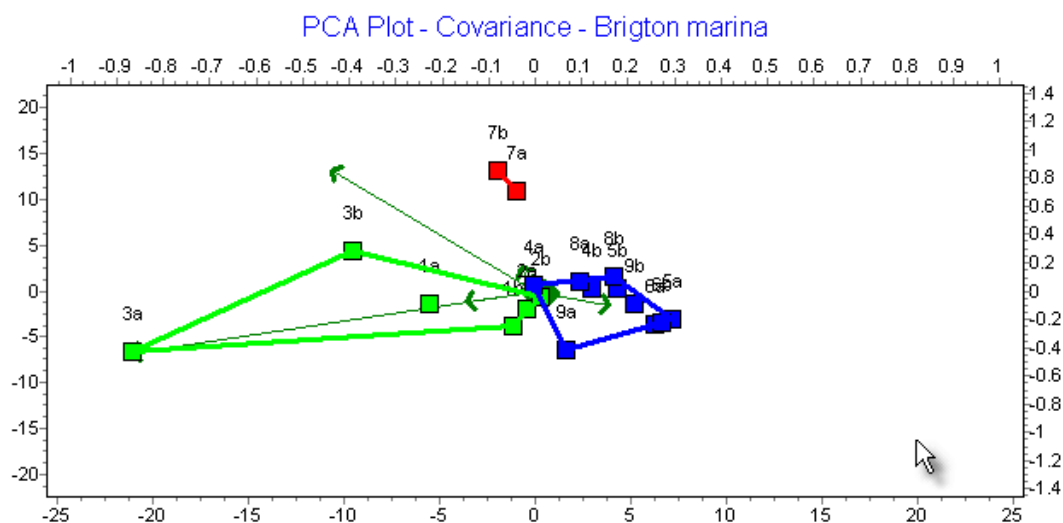
20.1 CAP

Community Analysis Package

For more information and a [free demo](#) visit our website [here](#).

CAP is a fast, easy to use, multivariate analysis package for Windows PCs. It has superb graphical output that gives you complete control over how you want your graphics to look. The program offers you a simple way to discover the power of multivariate statistics. Now in version 3, it is a proven performer that is both easy to use and affordable.

CAP allows you to identify your groups that are then displayed on the output, and even surround them with a perimeter, as can be seen below.



Example PCA output from CAP

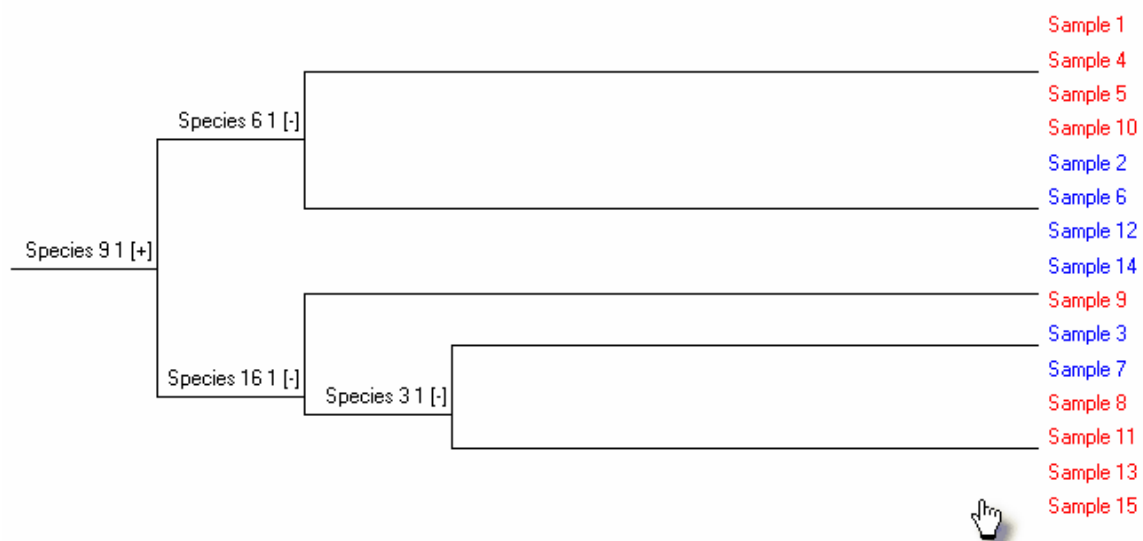
It offers all the analytical techniques commonly used by community ecologists, and researchers in other fields such as palaeontology, archaeology and the social sciences. It is particularly useful for ecological teaching because it allows students to quickly enter data, try different transformations, and explore a wide range of methods within a familiar Windows setting.

Its principal methods are:

- Two-way INDicator Species Analysis (TWINSpan)
- Principal Components Analysis (PCA)
- Non-metric Multidimensional Scaling (NMDS)
- Detrended Correspondence Analysis (DECORANA)
- Analysis of Similarity (ANOSIM) and Similarity Percentages (SIMPER)
- Agglomerative clustering
- Species Filtering
- Reciprocal Averaging
- 26 Similarity and distance measures
- Divisive cluster analysis
- Association analysis

CAP also feature 'Instant Assist' an (optional) always-on help panel to guide you through your analysis, to tell you what's appropriate and what the pitfalls of any technique are likely to be.

CAP is an indispensable part of every ecologist's analytical toolbox.



An example Twinspan dendrogram from CAP

20.2 ECOM

Ecological Community Analysis

For more information and a [free demo](#) visit our website [here](#).

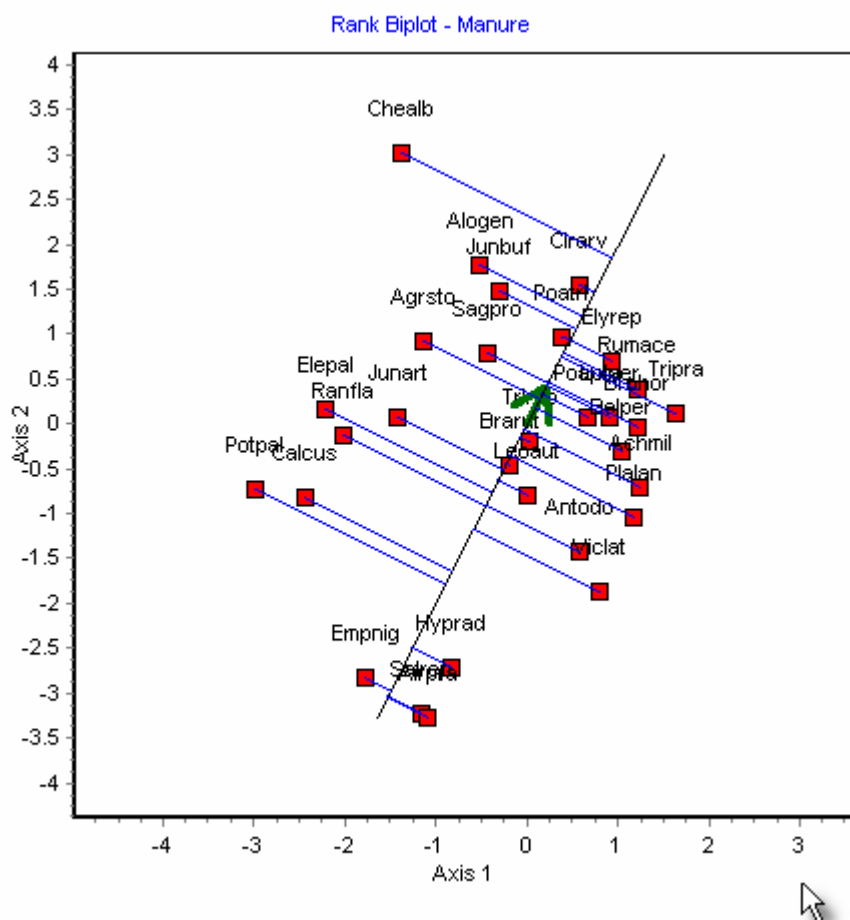
ECOM offers you a range of analytical techniques to discover, visualise and order relationships in multivariate data where the researcher has information on both the species present and the most influential environmental variables. For example, if you have sampled the animals from a series of habitats, and you have data on the

environment for each sample, then this program is for you. These methods are used relevant to several fields such as palaeontology, archaeology and the social sciences.

The principal methods are:

- Canonical Correspondence Analysis (CCA),
- Redundancy Analysis (RDA)
- Multiple Regression (MR).
- Stepwise Regression (SR)

A particular feature of **ECOM** is the graphical output that enables you to quickly visualise the complex relationships within your data set. For example, CCA output includes plots of sample and species ordinations, environmental correlation vectors and biplots of species arranged along individual environmental axes. All aspects of the graphs can be altered, and printed or exported in a wide variety of formats.



The program works on two separate input data sets holding the biological and physical data for the samples (sites). The environmental data may be either continuous variables such as pH, temperature or current speed; or binary variables such as soil or treatment type, which are scored as either a 0 or 1.

Backwards Stepwise Multiple Regression Output from ECOM					
Backward method in which parameters are removed sequentially					
Dependent Variable: Emprig					
Summary Table					
Step No	Vars Removed	R	R Sqr	Delta R Sqr	Vars In Model
1		0.557457	0.310758	0.000000	8.000000
2	Hayfield	0.557456	0.310757	0.000001	7.000000
3	BF	0.555576	0.308664	0.002093	6.000000
4	Manure	0.553667	0.306547	0.002117	5.000000
5	SF	0.550647	0.303212	0.003335	4.000000
6	Pasture	0.539043	0.290567	0.012645	3.000000
7	Moisture	0.449128	0.201716	0.088851	2.000000
8	A1	0.350438	0.122807	0.078909	1.000000

ECOM complements [CAP \(Community Analysis Package\)](#) and [Species Diversity and Richness IV \(SDR\)](#) which offer a range of techniques that only require biological data to undertake the analysis.

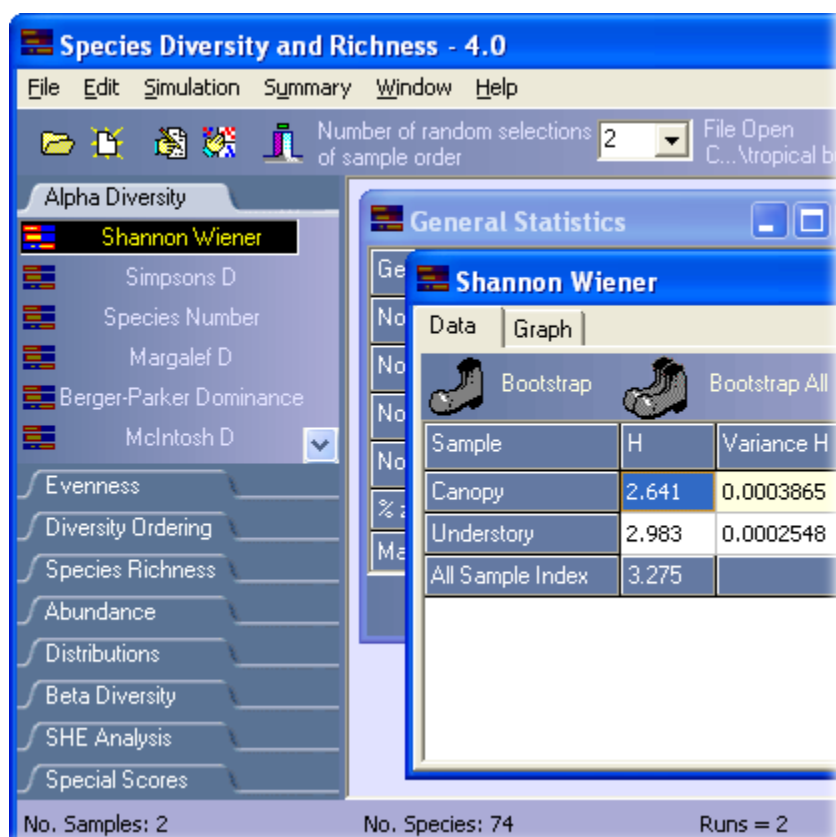
20.3 Species Diversity and Richness IV

Species Diversity and Richness

For more information and a [free demo](#) visit our website [here](#).

Calculate and compare species diversity, estimate species richness and study patterns of abundance in one easy to use program.

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.

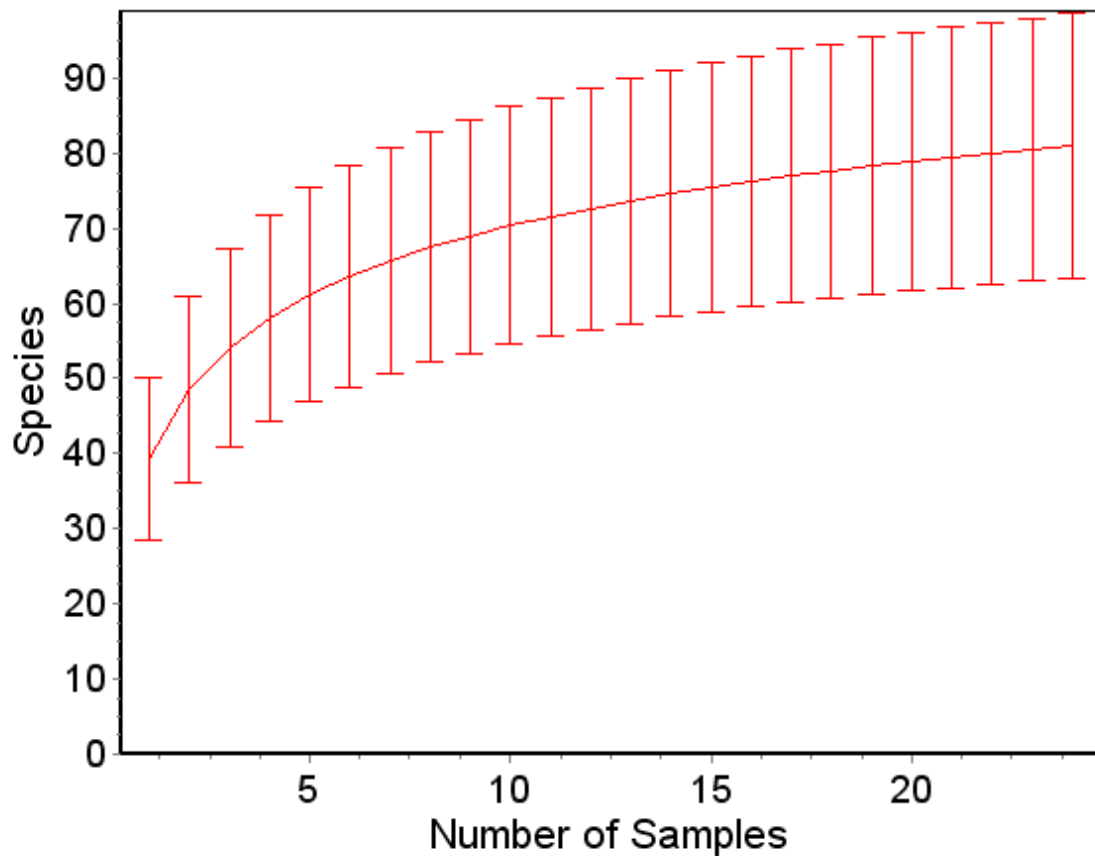


Species diversity and richness is a Windows program designed for both the professional ecologist and students.

Species diversity and richness version 4 offers considerable improvements over version 3. Existing users will immediately note the redesigned menu system and improved graphics, data input and printing. Probably the greatest recent improvements have been the extended range of methods and statistical tests on offer. The program offers all the analytical tools that a general ecologist needs for analysing and comparing the diversity of communities.

Methods available within *Species diversity and richness* include

- 11 Alpha diversity indices including Shannon's, Simpson's, Fisher's and Q statistic.
- 14 Evenness measures
- Bootstrapping of confidence intervals.
- Randomization tests to compare diversity between sites.
- Jackknife errors of total diversity/evenness
- 2 diversity ordering methods.
- 14 methods to estimate total species richness
- Rank order, K Dominance and species accumulation curves.
- 8 beta diversity measures
- SHE analysis
- Fitting of log series, log normal, geometric and broken stick abundance models
- Ability to generate simulated data
- Freshwater quality measures including BMWP, ASPT, LIFE & Irish quality rating, including the revised BMPS scoring system



Species diversity and Richness offers

An attractive user interface.

Extensive help system.

Easy data importation.

High quality graphical output.

and is ideal for

Community ecology research.

Undergraduate teaching.

Applied ecologists.

Ecologists with limited computer experience.

20.4 Axis

Axis

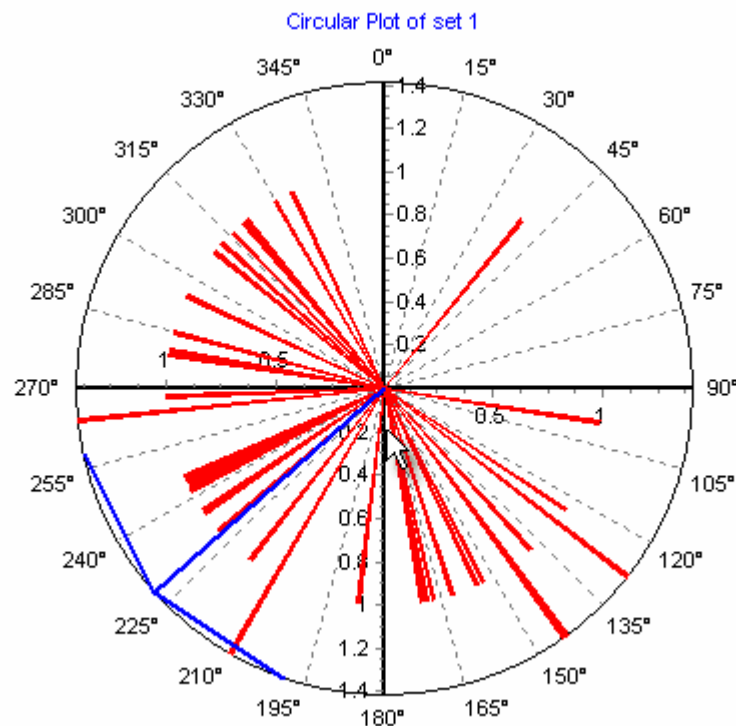
For more information and a [free demo](#) visit our website [here](#).

Finally a program that allows you to discover the relationships within your circular data. It offers a range of graphical and analytical techniques commonly used by biologists, geologists and archaeologists for the analysis of circular data.

Axis implements the principal graphical methods and statistical tests described by Fisher (1993) in *Statistical Analysis of Circular Data*. Periodic data of many kinds can be represented and analysed using the methods available within Axis.

While compass bearings are the most obvious examples of circular data, other examples could be the time of events over a 24-hour period, and the occurrence of activity over the lunar cycle.

Available plots within Axis include linear, circular and angular histograms, and optional smoothing by Fast Fourier Transform. Tests include: correlation between samples, uniformity or randomness, and specified mean direction. Simulated data sets can be created conforming to various distributions.



Axis has extremely versatile graphing capabilities, and can export images in .emf, .jpg, .png or .bmp formats. Tabulated data can be exported as plain text, .csv, or .xls files which will load into Excel or other spreadsheets - or even in HTML format.

Data sets can be imported from standard spreadsheets such as Excel, or created within the program.

20.5 Fuzzy Grouping

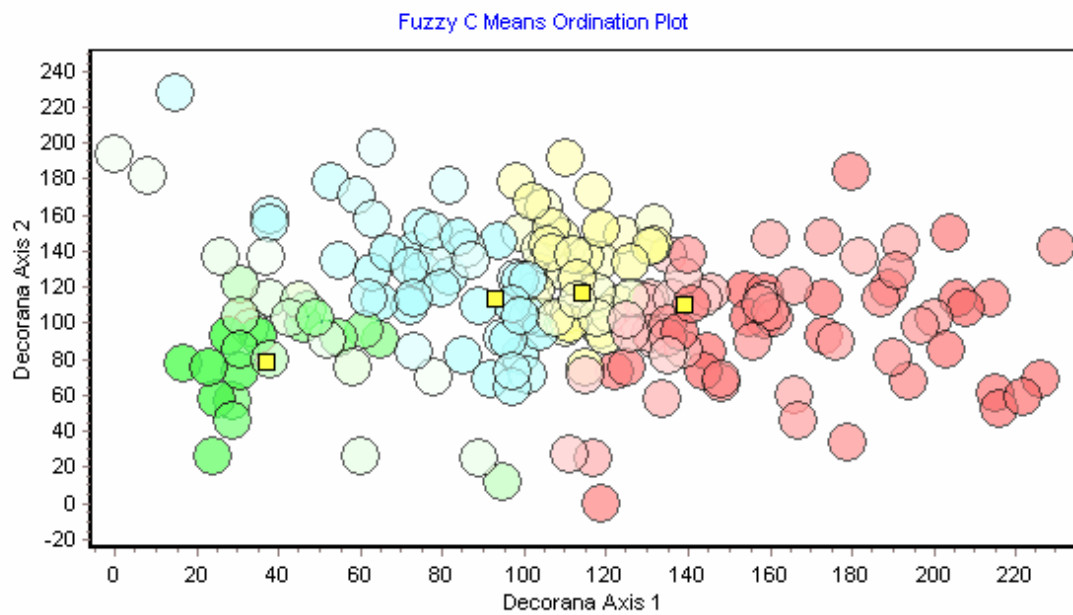
Fuzzy Grouping

For more information and a [free demo](#) visit our website [here](#).

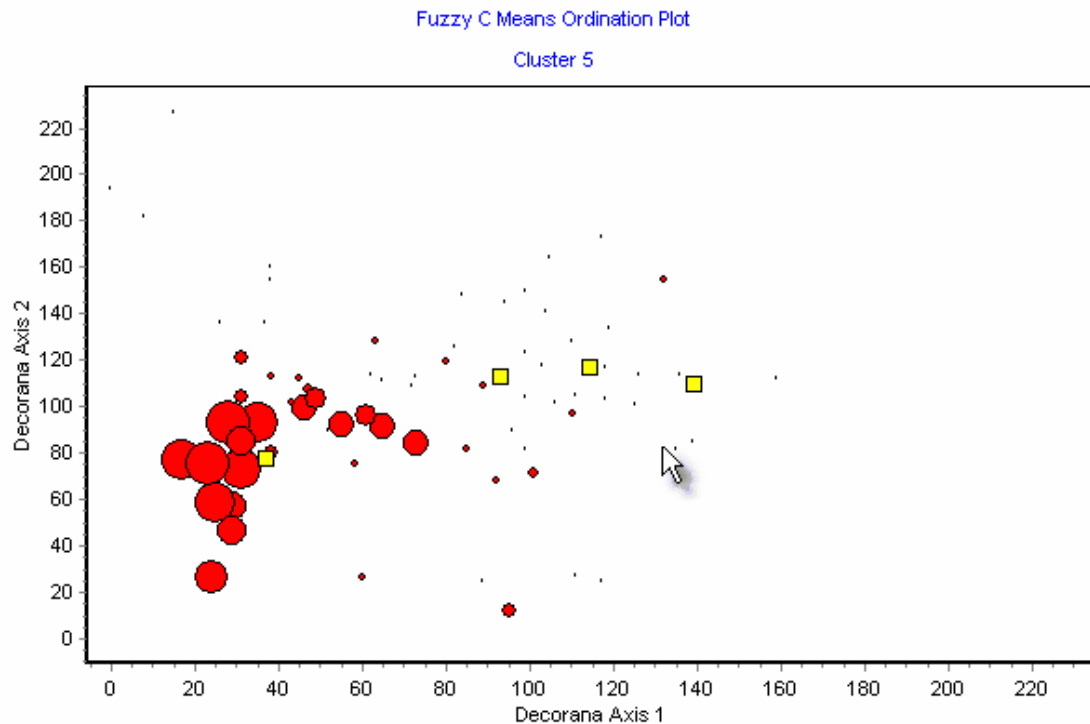
Fuzzy clustering methods allow samples to belong to more than one group. This

easy-to-use program for the PC allows the user to quickly carry out these techniques and is aimed to meet the needs of ecologists, palaeontology, archaeology and the social sciences. This has not been readily available before, and this has undoubtedly slowed the introduction of these techniques into these fields.

Lotfi Zadeh introduced the notion of a fuzzy set in 1965, as an approach for the handling of uncertain knowledge. Today these techniques are widely used over a range of scientific disciplines, and seem to be particularly appropriate to ecological analysis where the boundaries between groups such as communities or taxonomic units may be far from sharp.



In ordinary cluster analysis a sample or site either belongs, or does not belong, to a particular group or set. We can score membership as a value of 1 and non-membership as 0. In fuzzy clustering a degree of membership can be assigned as a value between 0 and 1 so that a value of say 0.8 would indicate a high probability that the sample belonged to the particular group in question. Thus, for a data set comprising many samples that can be hypothesized as being divisible into c groups, each sample has a degree of membership of belonging to each of the c groups.



Fuzzy Grouping offers three main methods of data analysis, first, fuzzy c-means, second, fuzzy ordination and finally fuzzy linear discriminant analysis. The first technique is appropriate for data that you suspect can be divided into c-groups but for which you may have no *a priori* information on group membership. The second, fuzzy ordination is appropriate when you have information on possible group membership. The second will allow you to plot a selected number of fuzzy groups in a space in which their differences are displayed to best effect and a measure of the ability of the groups to explain the total variability presented.

To help users to understand how to use Fuzzy Grouping, the instructions are written from an ecological viewpoint.

Fuzzy complements [CAP \(Community Analysis Package\)](#) and [ECOM \(Ecological Community Analysis\)](#)

20.6 Simply Tagging

Simply Tagging

For more information and a [free demo](#) visit our website [here](#).

Estimate the population size for both closed and open populations using mark-recapture methods. Simply Tagging is designed to be used by undergraduates and professionals and allows them to simulate data and explore the accuracy and reliability of mark-recapture methods.

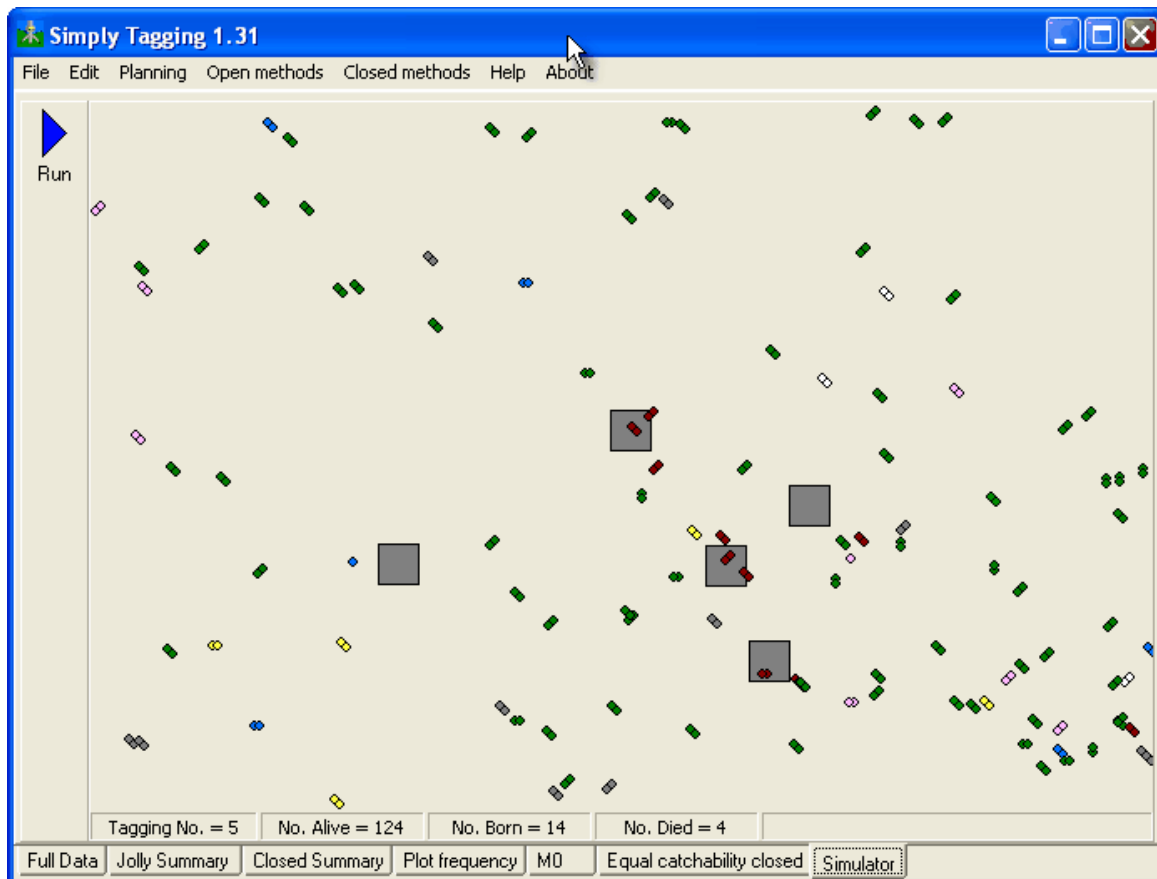
Main features:

- Data can be entered as individual animal capture histories or as summary tables.
- Ability to subset data by sex or age.

- Wide range of methods available for closed populations including Schnabel census and methods which assume behavioural differences between individuals.
- Undertakes full Jolly-Seber or constant survival and probability of capture model.
- Plots of population estimates and frequency of capture.

As with all PISCES software, Simply Tagging has an attractive user interface, extensive help system and good quality graphical output.

Tagging also includes a data simulation option to model mark-recapture experiments for open and closed populations. The user can vary sampling intensity, population size, animal behaviour and birth/death rates.



20.7 Population Estimation by Removal Sampling

Removal Sampling

For more information visit our website [here](#).

Removal Sampling allows you to calculate the statistics from removal trapping experiments. It contains three easy to use methods: constant & variable probability of capture (Zippin's maximum likelihood, ML) and a regression model.

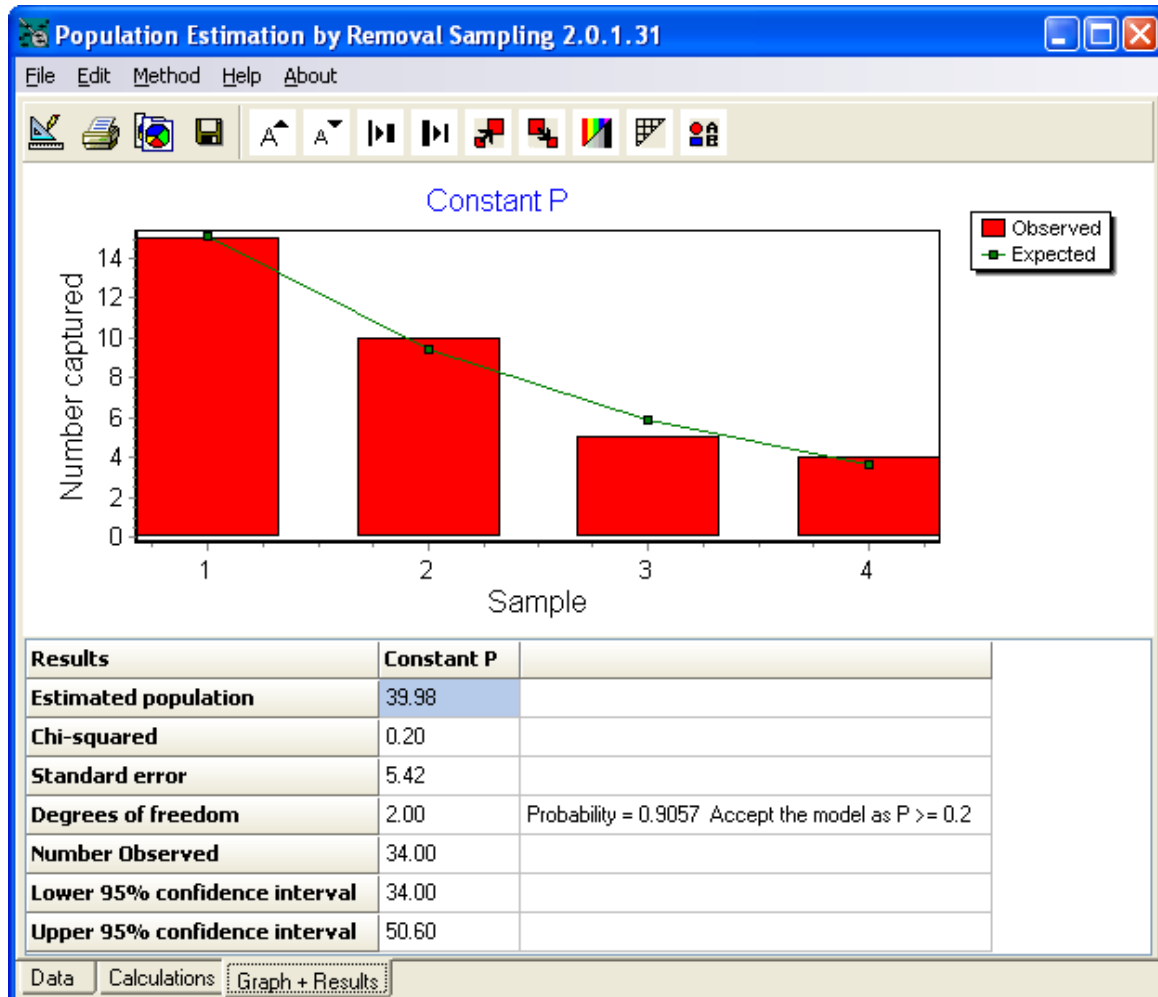
Using the ML methods the program can calculate a population estimate, the upper & lower 95% confidence intervals & the probability of fit of the model.

The program allows instantaneous graphical comparison of the various methods & rapid sub-division of the data. Data from spreadsheets such as Excel can be imported

as CSV files.

Removal Sampling is particularly useful for:

1. The estimation of fish densities using data collected by electric fishing;
2. Small mammal population estimation using traps.
3. Studies on insects that live in small, discrete populations that can be sampled using techniques such as sweep netting or pitfall traps.



20.8 Density from Distance

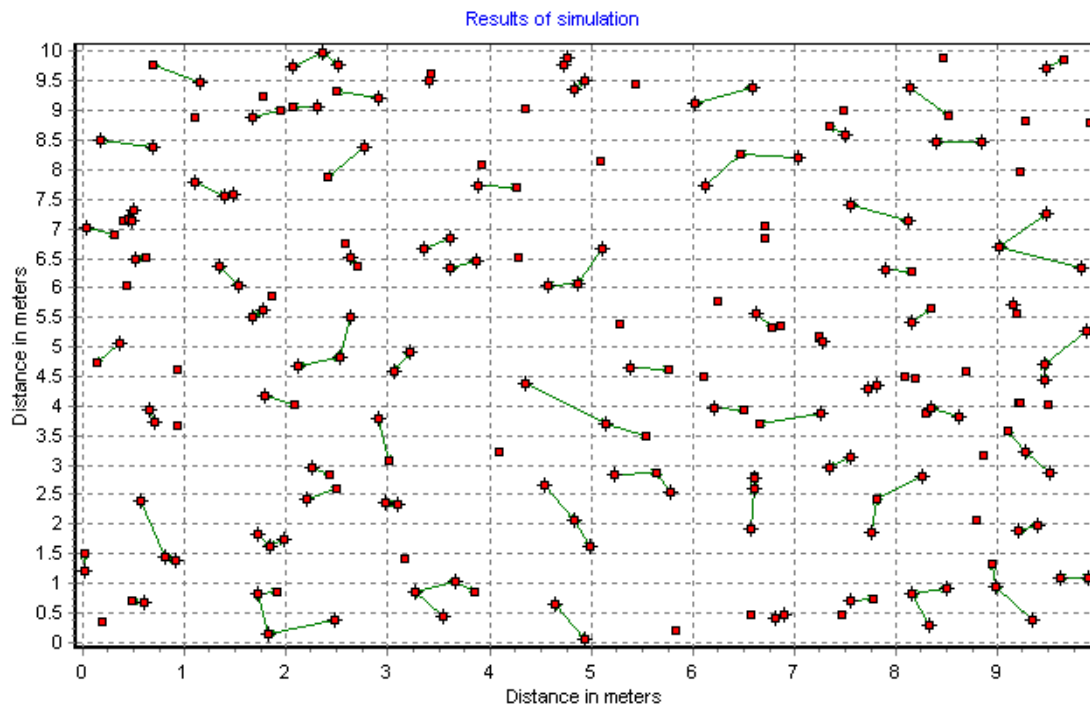
DfD - Density from Distance

For more information visit our website [here](#).

A range of analytical techniques are available in Density from Distance for you to estimate animal and plant density from measurements of the distance between objects, or from a selected line or point to the objects.

DfD has been designed for ease of use on Windows PCs. While designed for research use, the program is particularly useful for ecological teaching because it allows students to quickly enter or simulate data and explore a range of methods within a familiar Windows setting. The program is particularly useful for botanists, and zoologists studying birds or large mammals.

An example of a simulation using Nearest Neighbour:



20.9 Growth II (Simply Growth)

Growth II

A major update of our Simply Growth program, Growth II is the easy way for you to analyse growth curves from length and/or weight data.

A full list of the New Features in Growth II

New Models

- Exponential
- 4 Parameter Gompertz
- 4 Parameter logistic
- 3 Parameter Weibull
- 4 Parameter Weibull
- Morgan-Mercer-Flodin (MMF)
- Janoschek
- 4 Parameter Richards
- 5 Parameter Richards
- Non Linear Seasonally adjusted von Bertalanffy
- Linear Seasonally adjusted von Bertalanffy

New Appropriate Model Statistics

- Akaike Information Criterion (AIC)
- Schwarz Criterion (SC)

Improved Models

- Non Linear Von Bertalanffy

- 3 Parameter Gompertz
- 3 Parameter Logistic

Improved Graphic

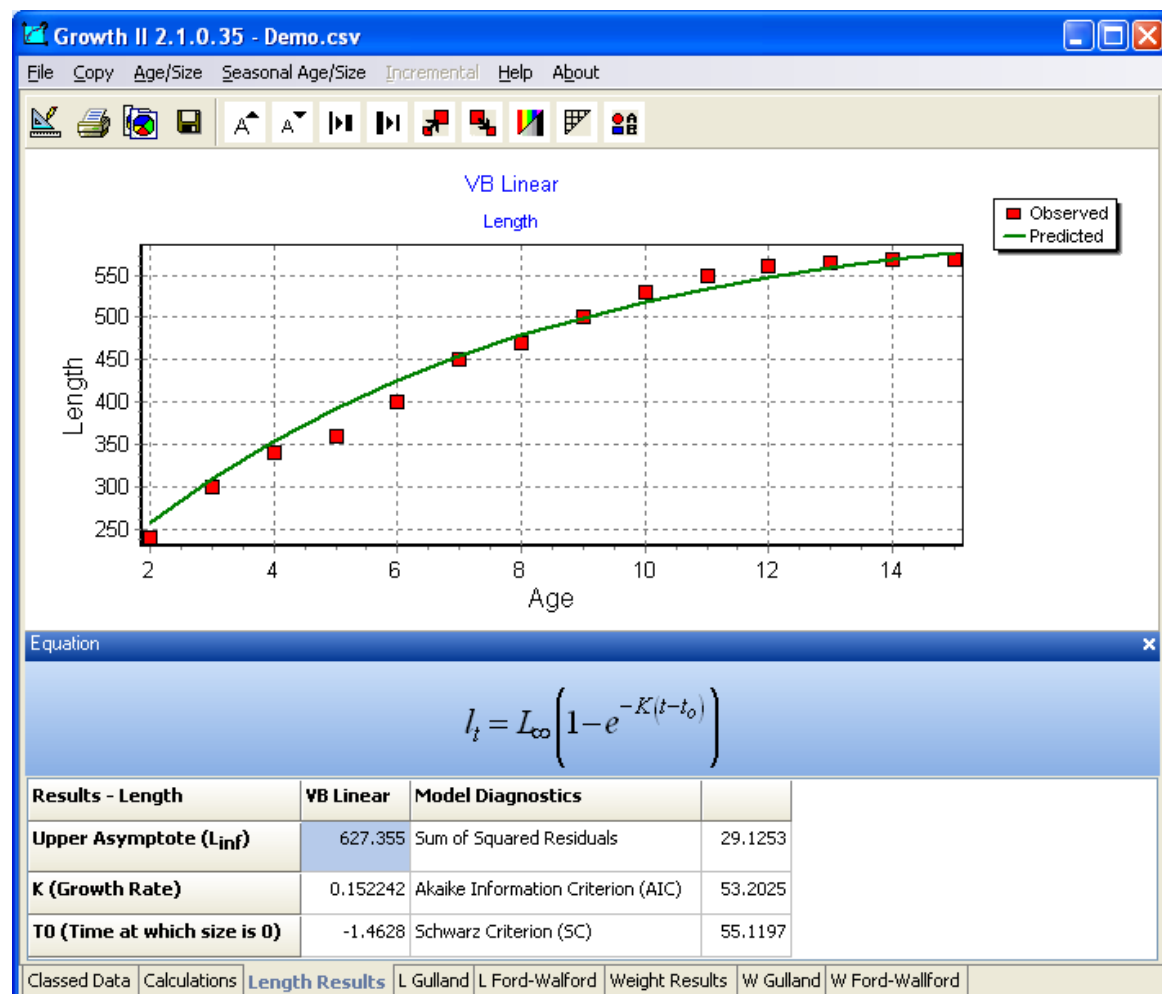
- Export Charts as BMP, WMF, EMF, JPEG, PCX, GIF and PDF
- Updated graphics
- New tools on each graph making publication quality charts easier
- Active clicking on graphs – takes you to the editor for each part of the graph
- Interactive results and Equation – click on a parameter and the corresponding part of the equation will be highlighted

Improved data entry and editing

- Redo/Undo on grids
- Paste in to grid
- Output to grids
- Opens Excel (*.xls) files

New Help

- Completely rewritten Help
- Explanations of the models
- Tutorials and worked examples



For more information visit our website [here](#).

20.10 Dynamica

Dynamica

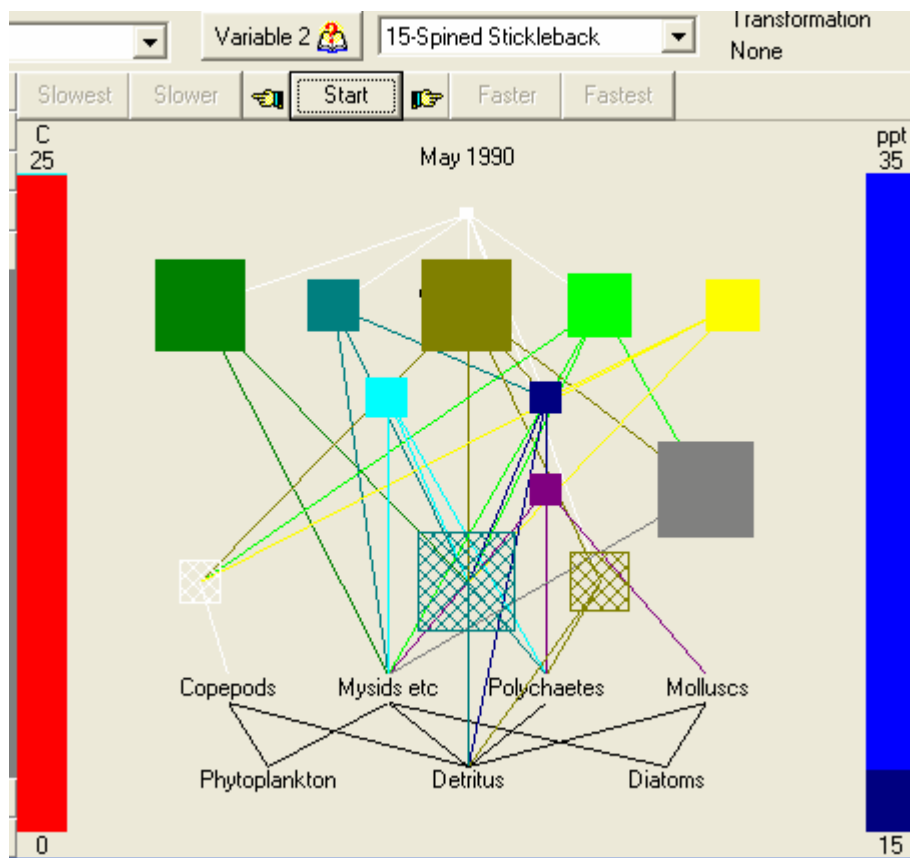
For more information visit our website [here](#).

You can explore real data, and reach your own conclusions about how animal numbers are inter-related and change through time. Data is from the Hinkley Point 'B' power station, Severn Estuary, UK

Dynamica features:

- Monthly abundance estimates of fish and crustaceans from January 1981 to January 1995.
- Time series for more than 100 species are available for study.
- Contains environmental data such as temperature, salinity and tides.
- Food webs that change through time.
- Statistics allow you to explore the complex inter-relationships between species and environmental factors.
- Length data for the common species of fish.
- Scientific papers from this data
- Details about data collection and habitat are also given.

The strong theme running through Dynamica is the study of community-level dynamics. Books often portray communities and food webs as static entities. This program will show you just how dynamic food webs are. Perhaps the most popular community statistic is a diversity index; Dynamica will let you compare different indices and appreciate how they change through time.



Dynamica has been designed to be useful to anyone with an interest in natural time series. It presents a range of elementary techniques for time series analysis, including data transformations, autocorrelations, moving averages and Fourier analysis. The program presents a wide range of natural dynamic behaviours which make good examples for study.

More information is available on the Hinkley Point sampling [on our website](#)

20.11 Hedgerow Assistant

Hedgerow Assistant

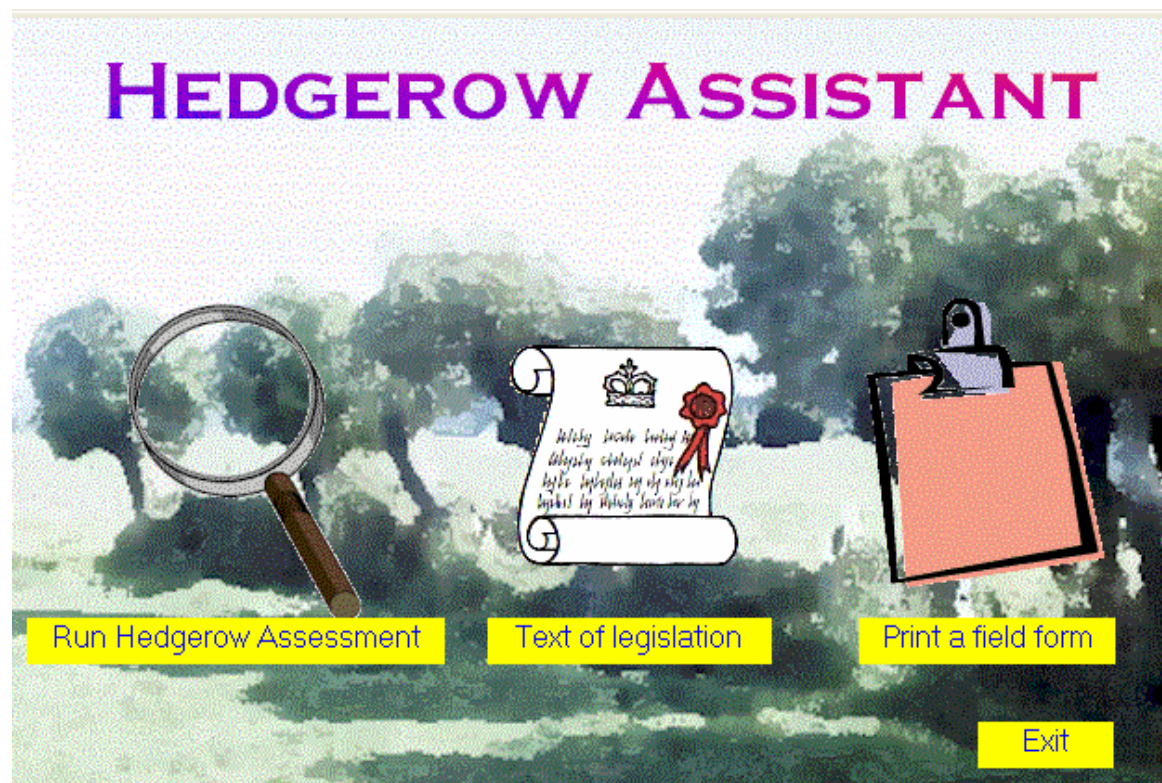
For more information visit our website [here](#).

Hedgerow Assistant provides the reliable way to survey hedgerows under the 1997 Hedgerows Regulations.

The 1997 Hedgerows Regulations provide standards by which hedgerows are evaluated and protected. The rules are complex and hard to interpret. Hedgerow Assistant guides you through the survey procedure, telling you which species to look for, which parts of the hedge to survey and which features to record. For environmental surveyors it saves time, work and money.

Features include:

- Step by step hedgerow evaluation.
- Lists of Red Data Book and Scheduled species.
- Full texts of relevant legislation.
- Advises on which parts of the hedge to survey.
- On-line help to interpret complex clauses.
- Prints field survey forms to guide data collection.



20.12 Simply Probit

Simply Probit

For more information visit our website [here](#).

Estimate lethal or environmental concentrations using toxicity data.

The program offers two types of analysis:

- Lethal dose calculations: a probit analysis of the survival of organisms exposed to a range of concentrations. Conforms to ISO/DIS 146609.
- Growth analysis: studies the inhibition in the growth of organisms exposed to different levels of a compound. Conforms to ISO 10253 .

Part



XXI

21 Lecture Resources

Lecture Resources

For more information visit our website [here](#).

Our current titles include:

- Seashore Ecology
- The Ecology of Freshwaters
- The Galapagos Islands from an Ecological Perspective - Peter Henderson & Scott Henderson
- An Introduction to Multivariate Statistics
- An Introduction to Diversity and Richness Methods

21.1 Seashore Ecology

Seashore Ecology

For more information visit our website [here](#).



This CD contains nine PowerPoint presentations showing the plants and animals of the seashore, and the techniques required to study and understand their ecology.

PowerPoint viewer software is included on the CD, so the lectures can be viewed on any PC.

Over 480 original photographs!



21.2 The Ecology of Freshwaters

The Ecology of Freshwaters

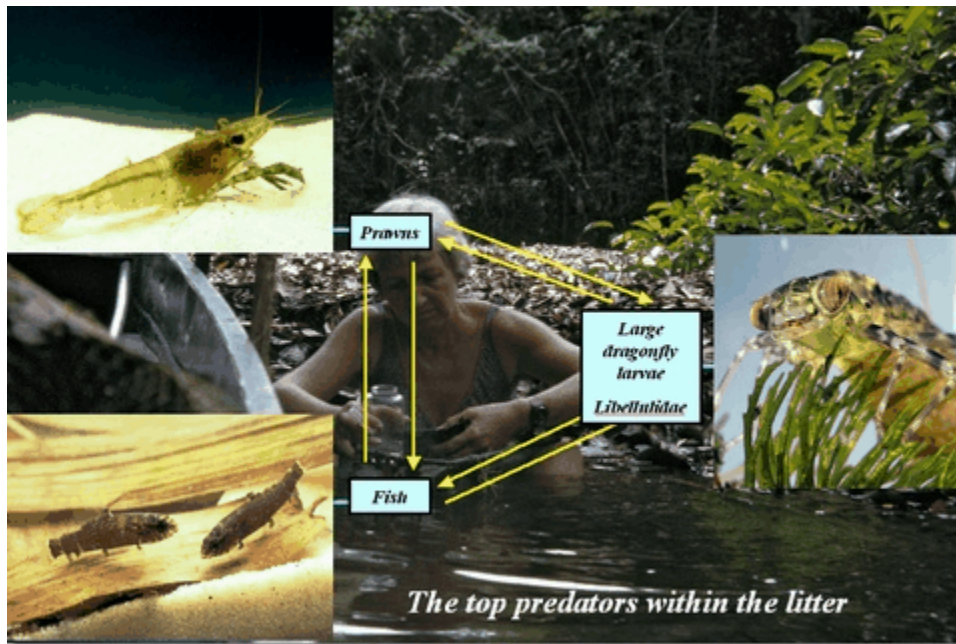
For more information visit our website [here](#).



This CD contains seven PowerPoint presentations showing the plants and animals of freshwaters around the world, and the techniques required to study and understand their ecology.

PowerPoint viewer software is included on the CD, so the lectures can be viewed on any PC.

Over 370 original photographs!



21.3 The Galapagos Islands

The Galapagos Islands from an Ecological Perspective

For more information visit our website [here](#).



This CD contains a PowerPoint presentation on the plants and animals of the islands, and exploring their habitats and ecology. Over 50 original photographs by senior ecologists and researchers at the Charles Darwin Research Station.



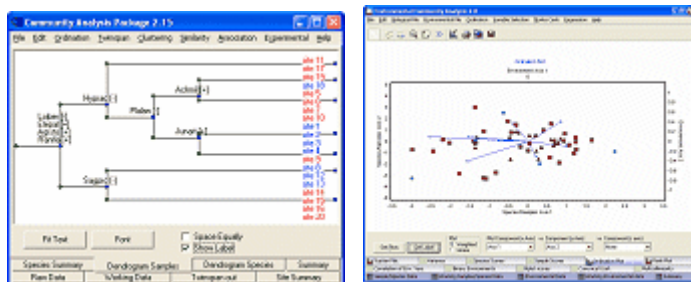
PowerPoint viewer software is included on the CD, so the lectures can be viewed on any PC.



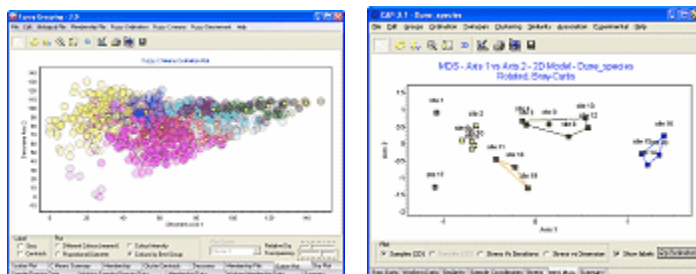
21.4 An Introduction to Multivariate Statistics

An Introduction to Multivariate Statistics

For more information visit our website [here](#).



This CD contains two PowerPoint presentations exploring the principles of multivariate statistical analysis, with guidance and example hints from biology, archaeology and geology. Lecture contents are designed to be accessible to senior undergraduate students, MSc students and postgraduates. They will, however, be useful to anyone seeking a better understanding of techniques of multivariate analysis.

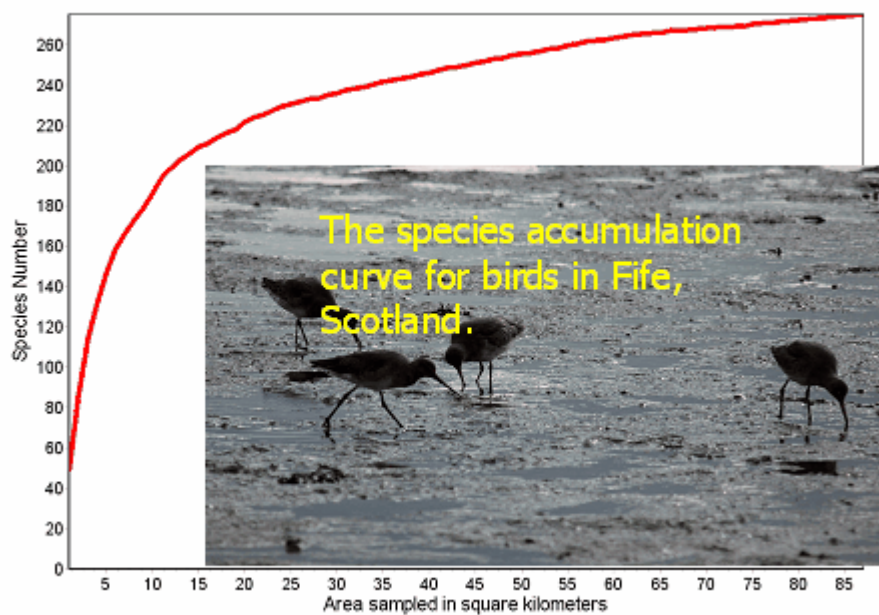


PowerPoint viewer software is included on the CD, so the lectures can be viewed on any PC.

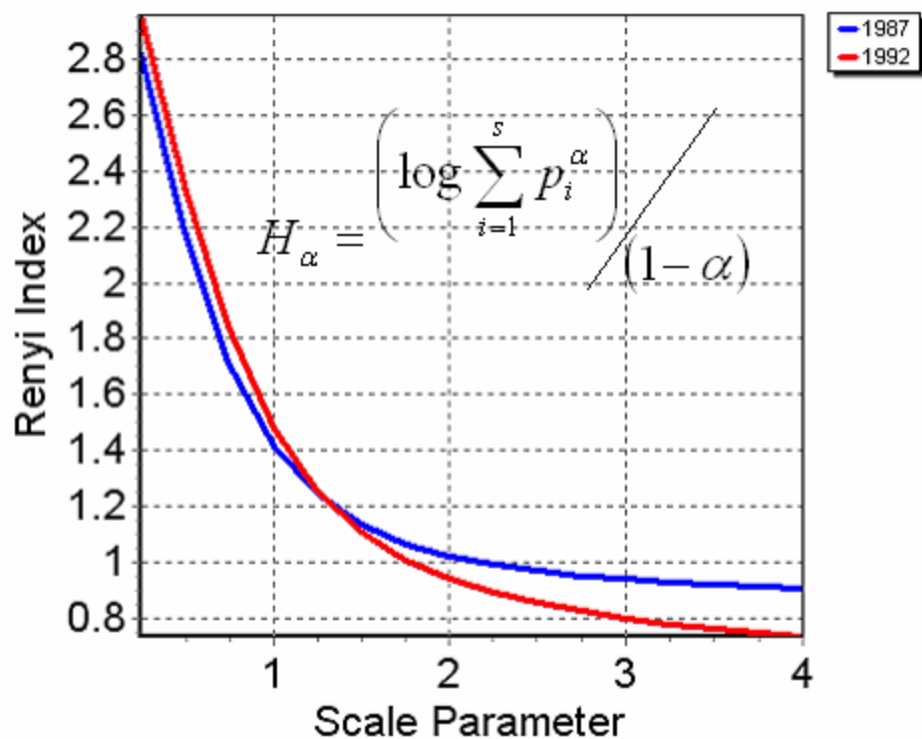
21.5 An Introduction to Species Diversity and Richness

An Introduction to Species Diversity and Richness

For more information visit our website [here](#).



This CD contains two PowerPoint presentations exploring the principles of Species Diversity and Richness, with guidance and example hints. Lecture contents are designed to be accessible to senior undergraduate students, MSc students and postgraduates. They will, however, be useful to anyone seeking a better understanding of techniques of Species Diversity and Richness.



PowerPoint viewer software is included on the CD, so the lectures can be viewed on any PC.

Part



22 E-books

E-Books

Our library of e-books is growing continuously.

Pisces e-books are supplied on CD as Adobe Acrobat PDF files. The PDF format is well known, extremely versatile, and will run on both Windows and Mac operating systems. The Acrobat Reader software to use them is available as a free download from Adobe - and we also supply the 2 latest Windows versions on the e-book CDs.

Many e-books are simply 'photographic' copies of the originals. This means that the pages are scanned images of the book. This can give a poor-quality image, with a huge file size, and means that the text can not be selected, or searched. Pisces e-books are carefully scanned from the originals, repeatedly proof-read, saved as fully-searchable text, and paginated exactly as the original, so that all page and index references remain correct. We do not make corrections or alterations to the original, except:

- in the case of obvious typos, where the correction is added in [square brackets], and
- where the original contains an 'Errata' section.

Quotes from customers about Pisces e-books:

" ... very well done for the service you are providing for naturalists and the quality of your products".

" ... Pisces are putting essential works back into the reach of many who needed them".

We now have e-books for those interested in the fields of entomology, botany, marine zoology and early exploration of the natural world.

For more information visit our website [here](#).

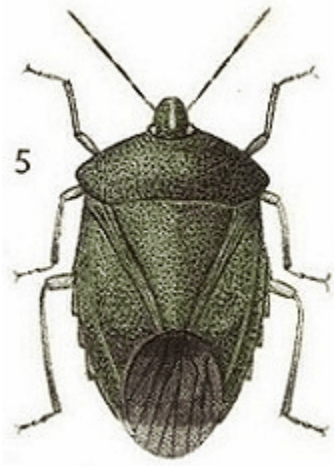
Our current titles include works on the following:

- [Entomology](#)
- [Aquatic Biology](#)
- [Botany](#)
- [Classics of Science and Exploration](#)
- [Pisces Thesis Series](#)

22.1 Entomology

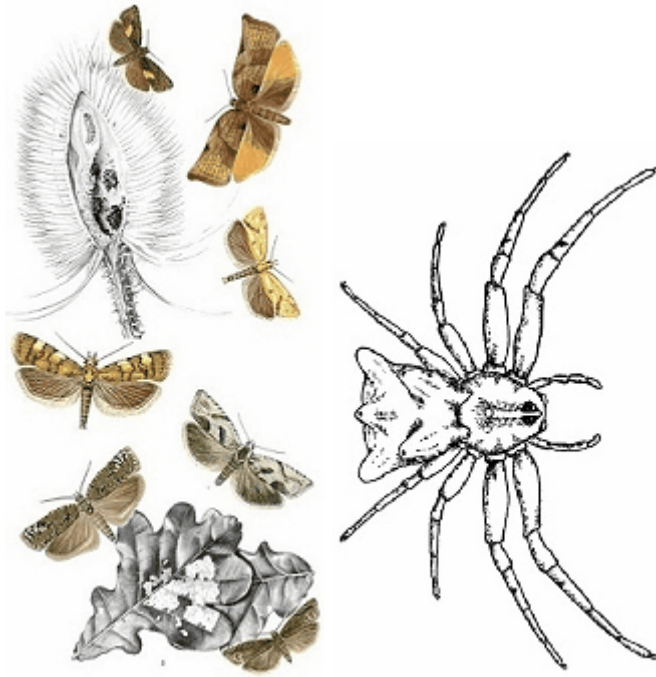
Entomology

For more information visit our website [here](#).



Our current titles include (ebooks unless stated):

- Land and Water Bugs of the British Isles - Southwood & Leston
- Land and Water Bugs - 2005 print edition
- British Tortricoid Moths - Bradley, Tremewan & Smith
- British Water Beetles - Balfour-Browne
- Centipedes of the British Isles - Eason
- Larvae of the British Butterflies and Moths (Wm. Buckler):
 - Vol. 1
 - Vols. 2 & 3
 - Vols. 4 to 6
 - Vols. 7 & 8
 - Vol. 9
- British Spiders (Vols 1-3) - Locket & Millidge
- Mantidae - Giglio-Tos
- The Comity of Spiders - W.S. Bristowe
- Collected Essays of J. Henri Fabre - Vol. 1
- The Genitalia of the Geometridae - Pierce
- The Genitalia of the Rhopalocera / Larger British Moths - Pierce
- The Genitalia of the Tortricidae - Pierce
- The Gentialia of the Pyrales, Deltoids and Plumes - Pierce



22.2 Aquatic Biology

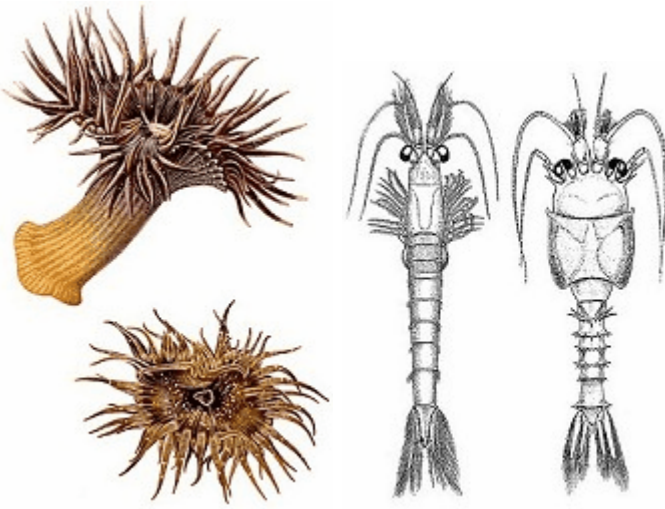
Aquatic Biology

For more information visit our website [here](#).



Our current titles include

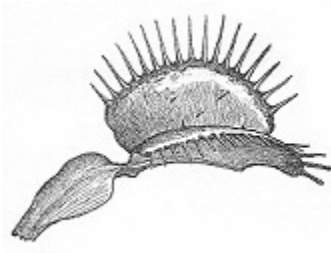
- British Sea Anemones - Stephenson
- British Mysidacea - Tattersall & Tattersall
- British Naked-eyed Medusae - Forbes
- The Tunicata - N.J. Berrill
- Freshwater Ostracods - Henderson
- Amazonian Fishes & Their Habitats - Henderson
- Crustacea of Norway, Vol. I: Amphipoda - G.O. Sars
- Crustacea of Norway, Vol. II: Isopoda - G.O. Sars
- Crustacea of Norway, Vol. III: Cumacea - G.O. Sars
- Crustacea of Norway, Vol. IV: Copepoda (Calanoida) - G.O. Sars
- Crustacea of Norway, Vol. IX: Ostracoda - G.O. Sars
- British Fresh-water Copepods, Vols. I - III - Robert Gurney
- British Bivalve Seashells - Norman Tebble
- Marine Plankton - A Practical Guide - Newell & Newell
- Marine Plankton - A Practical Guide - 2006 facsimile edition
- British Marine Amphipoda: Gammaridea - Roger J. Lincoln
- Biology of Intertidal Animals - Richard C. Newell



22.3 Botany

Botany

For more information visit our website [here](#).



Our current titles include :

- Insectivorous Plants - Darwin
- Brambles of the British Isles - Edees & Newton
- The Naturalist on the River Amazons - Bates
- Notes of a Botanist on the Amazon and Andes - Spruce
- Travels on the Amazon and Rio Negro - Wallace
- The Malay Archipelago - Wallace
- A Botanist in the Amazon Valley - Ruggles Gates



22.4 Classics of Science and Exploration

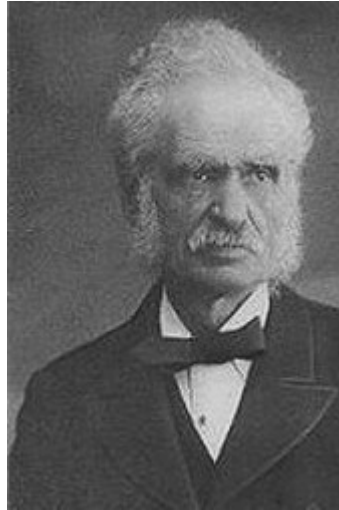
Classics of Science and Exploration

For more information visit our website [here](#).



Our current titles include:

- On the Origin of Species, and The Descent of Man - Darwin
- The Naturalist on the River Amazons - Bates
- Notes of a Botanist on the Amazon and Andes - Spruce
- Travels on the Amazon and Rio Negro - Wallace
- The Malay Archipelago - Wallace
- A Botanist in the Amazon Valley - Ruggles Gates
- The Natural History of Selborne - White
- Amazonian Fishes & Their Habitats - Henderson



Contemporary Science

- The River - A Journey to the Source of HIV and AIDS - Edward Hooper

22.5 Pisces Thesis Series

Pisces Thesis Series

For more information visit our website [here](#).



Our current titles include

- Satellite mapping of vegetation change - Canney
- Ground Ant Communities of Kakamega Forest - Espira
- Coral reef fish populations in Zanzibar, Tanzania - Tyler



Part

XII

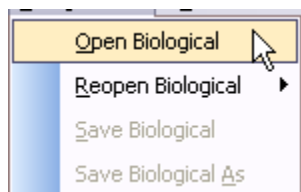
23 Worked Example

This Section gives an example of the use of ECOM. It is based on the well known Dune Meadow data set used by Jongman et al. (1995).

The objective of the analysis is *to determine if there are any gradients in the species data that can be explained by the environmental variables.*

The data files are supplied with the ECOM installation.

First open the species abundance file using the Biological File drop-down menu. This file holds the dependent variables and is named **Dune_species.csv**.



Now open the file of environmental variables using the Environmental File drop-down menu. This file holds the independent variables that may influence species abundance and is named **Dune_enviro.csv**.

We will now take you through the stages of a typical Canonical Correspondence Analysis. The purpose of this analysis is to identify the environmental variables that are most important in moulding the plant community. The stages of this process are described in turn below.

[Preliminary inspection and data transformation](#)
[Examine and Select Environmental Variables](#)
[Interpreting the results](#)

23.1 Preliminary inspection and data transformation

Having opened the data sets the species and environmental data will be presented in grids. The raw species and environmental data is viewed by clicking on the **Sample/Species Data** and **Environmental Data** tabs respectively. The data sets can be changed and transformed from the working data tabs.

Changes made to the working data will not change the raw data.

We will not undertake any transformations on the Dune Meadow data set. This is because the numbers in both the environmental and species data sets are both with a single order of magnitude.

We can find out about the range in our data using the **Summary** tab. Below it can be seen that the largest maximum environmental variable value is 11.5 and the smallest is 1. A similar inspection should be undertaken on the species data.

Summary of Data										
Row	Mean	Median	Max	Min	Zeros	Non-zeros	% zeros	Sum	Sum Sqr	Variance
A1	4.850000	4.200000	11.500000	2.800000	0	20	0.000000	97.000000	560.680000	4.748947
Moisture	2.900000	2.000000	5.000000	1.000000	0	20	0.000000	58.000000	230.000000	3.252632
Manure	1.750000	2.000000	4.000000	0.000000	6	14	30.000000	35.000000	103.000000	2.197368
Hayfield	0.350000	0.000000	1.000000	0.000000	13	7	65.000000	7.000000	7.000000	0.239474
Pasture	0.250000	0.000000	1.000000	0.000000	15	5	75.000000	5.000000	5.000000	0.197368
SF	0.300000	0.000000	1.000000	0.000000	14	6	70.000000	6.000000	6.000000	0.221053
BF	0.150000	0.000000	1.000000	0.000000	17	3	85.000000	3.000000	3.000000	0.134211
NM	0.300000	0.000000	1.000000	0.000000	14	6	70.000000	6.000000	6.000000	0.221053

Use

☒ Working Data
 ☐ Raw Data

Statistics

☐ General
 ☒ Row
 ☐ Column

Data Type

☒ Environmental
 ☐ Biological

☒ Sample/Species Data
 ☒ Working Samples/Species Data
 ☒ Environmental Data
 ☒ Working Environmental data
 ☒ Summary
 ☐ Scatter Plot

It is often the case that species vary greatly in abundance, over possibly 5 or more orders of magnitude. When there is great variation in abundance consider making a square root transformation. A log transformation might also be useful but most species data sets include many zeros so you would need to use a log $x+1$ transformation and this can be misleading if there are many low abundance species.

During this preliminary stage you should also check for zero sum rows and columns in your data sets as these will likely cause division by zero problems later. This is done using the **Handling zeros** radio button on the Working data tabs.

Brohor

0

4

0

3

Type of Adjustment

☐ Transform
 ☐ Relative
 ☒ Handling zeros
 ☐ Transpose

Deselect Column

Deselect Row

Delete 0 Columns

Delete 0 Rows

Remove Sparse Columns

Remove Sparse Rows

☒ Sample/Species Data
 ☒ Working Samples/Species Data
 ☐ Env

Having completed these initial inspections we must now Examine and Select Environmental Variables.

[Interpreting the results](#)

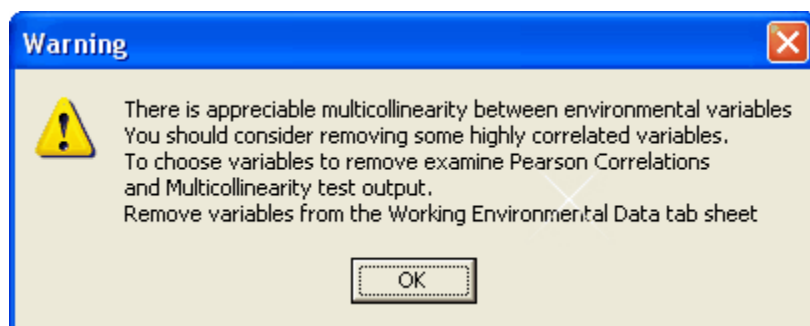
23.2 Examine and Select Environmental Variables

It is unlikely that all of the environmental variables will contribute usefully to the eventual model and there are often problems with [multicollinearity](#).

The first stage is to simply run a canonical correspondence analysis off the ordination tab using all the environmental variables.

1. Redundancy Analysis - distance biplot
2. Redundancy Analysis - correlation biplot
3. Correspondence Analysis - sites at species centroids
4. Correspondence Analysis - species at site centroids
5. Test for Multicollinearity

With the Dune Meadow data you will then read the following warning:



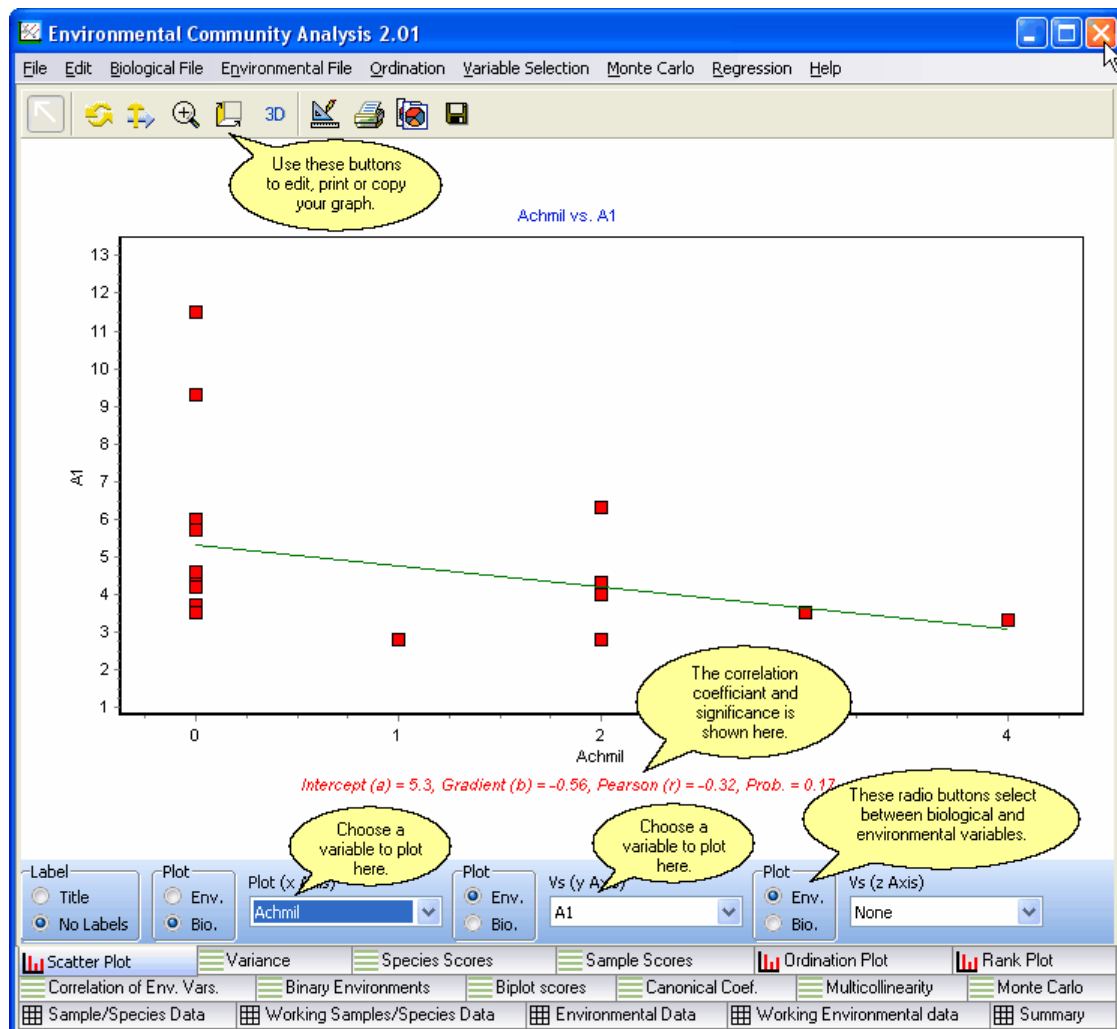
Click OK and ECOM will complete the analysis and open the output at the **Multicollinearity** tab.

Dependent variable	R-squared	VIF
A1	0.555667	2.250565
Moisture	0.457282	1.842579
Manure	0.903249	10.335815
Hayfield	0.714825	3.506623
Pasture	0.415338	1.710390
SF	0.694218	3.270308
BF	0.558008	2.262482
NM	0.790565	4.774750

Note the Variance Inflation Factor (VIF) values. A VIF above 10 indicates appreciable multicollinearity. In the above table it is clear that the Manure variable is highly correlated with a combination of other variables. Manure should probably be removed from the analysis. But first we

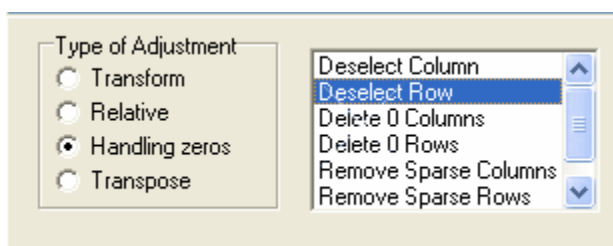
can see if Manure is highly correlated with any other single environmental variable.

Click on the **Scatter Plot** tab. Select Manure in Plot (x Axis) and each of the other environmental variables in turn in the y axis menu box. If you activate the box you can use the up and down arrows to swiftly move through the variables. Make sure you have selected Env. in the **Plot** radio button or you will only be offered the species!



The scatter plot of Manure against SF indicates a Pearson correlation between these variables of 0.72. Evidently the standard farming variable and Manure are related. This increases our suspicion that Manure or possibly SF should be removed from the final analysis.

We will not remove any variables from the present analysis but for future reference this is done using the Deselect Row in the **Working Environmental data** tab.



It is often useful to identify the top 1, 2 or 3 environmental variables for inclusion in the final

analysis. This can be done using the [Variable Selection](#) drop-down menu.

The next stage is to [interpret the results](#).

23.3 Interpreting the results

When a Canonical Correspondence Analysis is undertaken a range of tabbed sheets of output will be produced. In this section we will use the Dune Meadow data to take you through the output.

Having undertaken a CCA from the drop-down:

1. Redundancy Analysis - distance biplot
2. Redundancy Analysis - correlation biplot
3. Correspondence Analysis - sites at species centroids
4. Correspondence Analysis - species at site centroids
5. Test for Multicollinearity

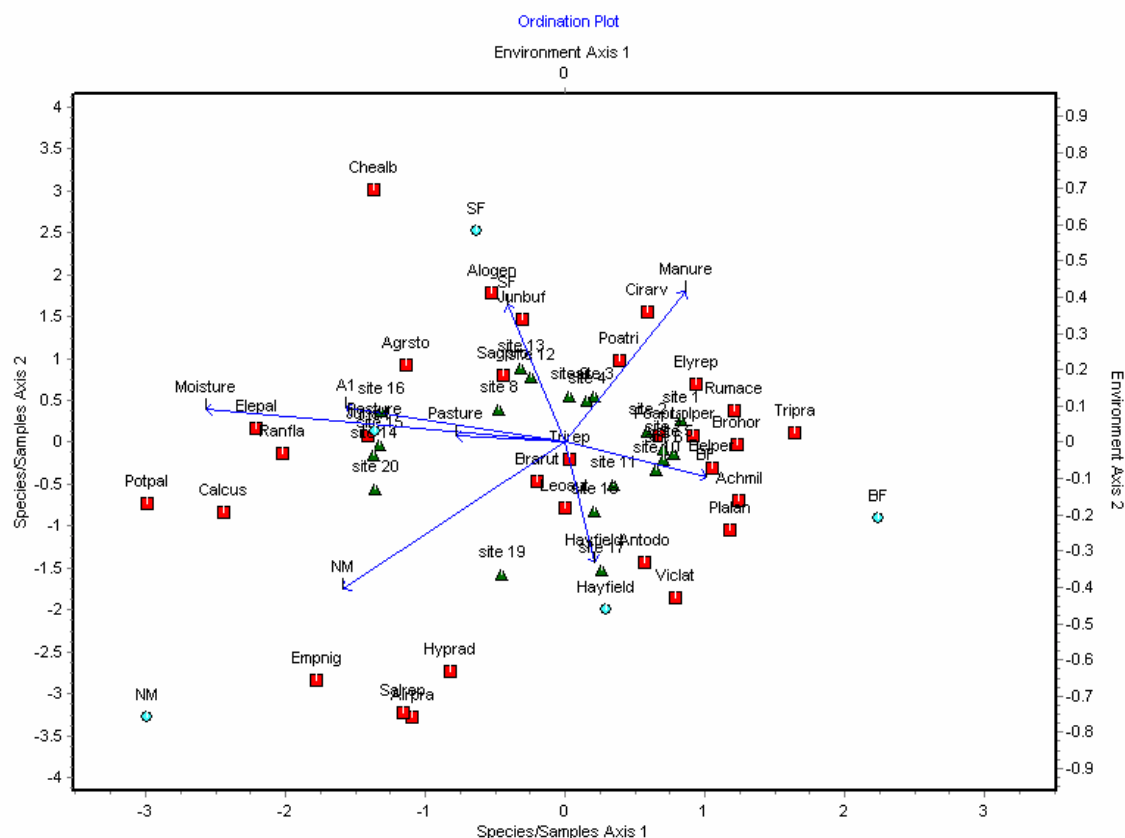
we first examine the **Variance** tab. The multiple correlation species/environment scores indicate how much of the variation in species composition is explicable from the environmental variables. The figure of 0.9579 for the first axis indicates that we can account for most of the variation in species composition by taking account of the environmental variables. The maximum and minimum values for this coefficient are 1.0 and 0, and a value close to 1 indicates that the environmental variables are having an appreciable effect. This correlation should be viewed with great caution and can be highly misleading - see [Problems with Species-Environment Correlation](#). Note also that the first canonical axis explains about 21% of the total variability. This is quite a large amount but may not be significant. A [Monte Carlo test](#) must be used to test for significance.

Canonical Eigen value =		0.461211	0.298078	0.160054	0.133690
% variance explained =		21.803949	14.091755	7.566618	6.320260
Cumulative % variance =		21.803949	35.895704	43.462322	49.782582
Multiple correlation species/environment scores =		0.957900	0.913556	0.871335	0.873106
Kendal rank correlation of species/environment scores =		0.810526	0.768421	0.705263	0.684211
Non-canonical Eigen value =					

The results of the Monte Carlo test with 1000 replicates are shown below. This indicates that the amount of variability explained by the environmental variables is significant for both the first and second canonical axes (Probability < 0.05 - less than a 5% chance that the observed relationship could occur by chance).

Axis	1	2	3
Actual Eigenvalues	0.461211	0.298078	0.160054
Eigenvalue results from simulation			
Mean	0.295348	0.186423	0.124931
Maximum	0.482257	0.330007	0.206152
Minimum	0.151126	0.108245	0.076928
Probability	0.002997	0.003996	0.071928
Number of trials	1000		

The sample and species scores that will be plotted are presented in grids in the **Sample Score** and **Species Score** tabs. Generally you will be far more interested in the plot of the results which is viewed from the **Ordination Plot** tab is shown below. This is a plot of the 1st and 2nd canonical axes, which we have shown above are the only statistically significant axes. Thus, there is not much point in examining the other axes.



The lengths and positions of the arrows provide information about the relationship between the environmental variables and the derived axes. Arrows that are parallel to an axis (e.g. moisture and axis 1) indicate a correlation, the length of the arrow tells us about the strength of that correlation. Thus, pasture is related to axis 1 but not as strongly as moisture. Neither of these is related to axis 2.

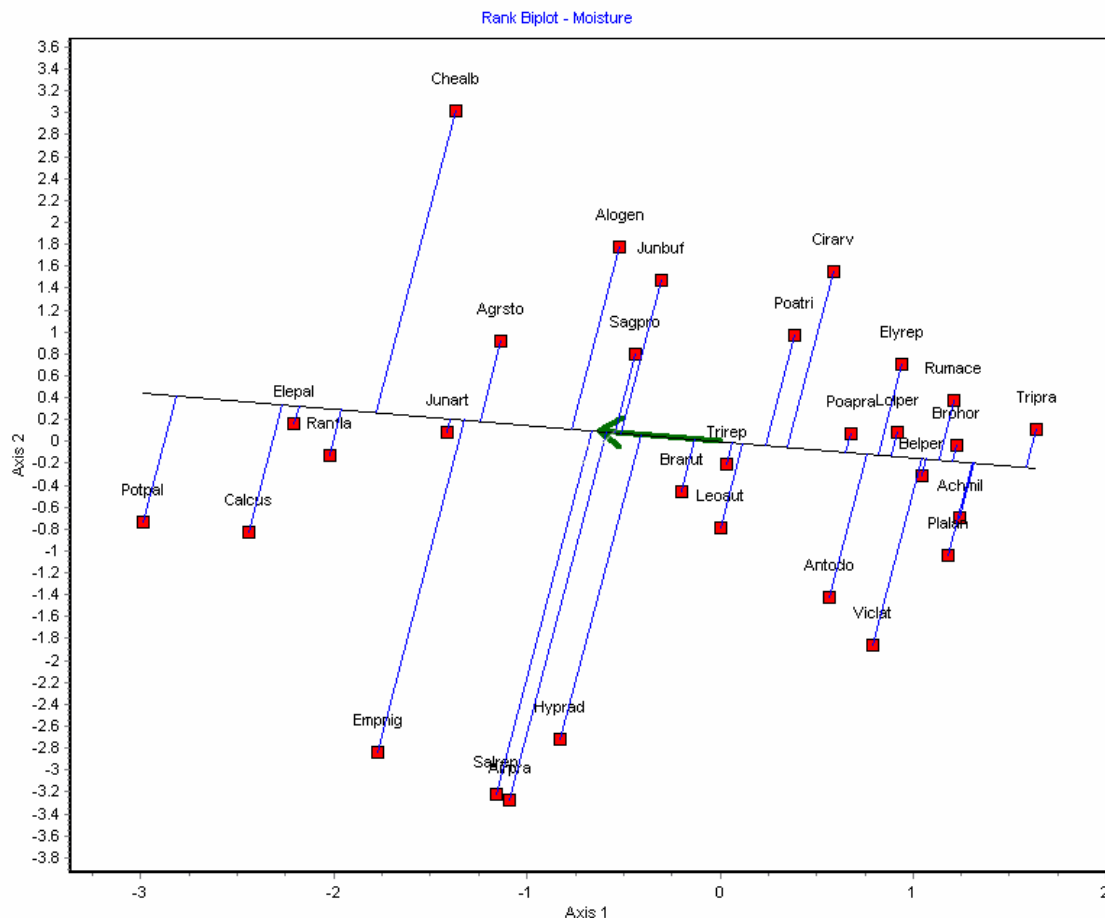
The above plot is rather confusing and cluttered because we have plotted the environmental vectors or centroids, the site (sample) scores and the species (dependent variable) scores all on the same graph. Further, all the labels have been included. The [graph plotting options](#) allow many

variations to be plotted to make the output clearer.

The position of the sites indicates their status with respect to the environmental variables. Thus site 19 has a negative score on axis 2 and axis 2 has a high manure component. This indicates that we would predict there has been little manuring on site 19.

In similar fashion the positions of the species indicates their response to particular environmental variables. For example Potpal, Elepal and Ranfla are associated with moist sites.

We can see the inferred relationship between the species and the different environmental variables far more clearly using the **Rank Plot** tab as shown for moisture below.



The position of the species Potpal, Elepal and Ranfla at the extreme high moisture content end of the distribution is now quite clear.

[Back to beginning](#)

23.3.1 Problems with Species-Environment Correlation

The following words of caution are from

Influence of noisy environmental data on canonical correspondence analysis

by Bruce McCune

Ecology, Dec 1997.

As the number of variables in the second matrix increases to near the maximum possible (one less than the number of sample units), the species-environment correlation always converges on 1.0. At that point, the LC scores (sites scores calculated as a linear combination of the environmental variables) and WA scores (sites scores calculated as a weighted average of the species abundances) are identical because the second matrix exerts no influence over the results (ter Braak 1994), since the large number of variables in the second matrix can support any pattern found by the weighted averaging step with the species matrix.

Using a wide variety of real data sets, it is clear that the species-environment correlation is almost always high. It is usually > 0.6 , and often much higher. This seems to be true regardless of other criteria for performance of the ordination, such as interpretability or proportion of variance in the species matrix that is explained.

For these reasons, I conclude that (1) the species-environment correlation is a poor criterion for evaluating the success of an ordination, (2) the species-environment correlation should not be interpreted literally as a measure of the strength of the relationship between species and the environment, and (3) the statistical significance of the species-environment correlation, even when it appears very high, should always be checked with randomization tests.

The problems caused by large numbers of noisy environmental variables cannot be alleviated by using a stepwise selection procedure, similar to that commonly used in multiple regression. If the environmental data are noisy, stepwise selection will simply pick out the best random variables, and the species-environment correlation will still be misleadingly high. As with multiple regression, the parsimony of the procedure is set by the size of the pool of independent variables, not the number of independent variables selected.

I recommend that the species-environment correlation not be reported as such. If it is reported at all, a more appropriate name would be the "LC-WA correlation" to avoid misleading the reader. If it is reported, it should be accompanied by a randomization test for statistical significance. A better (but still imperfect) measure of the strength of the relationship between species and environmental matrices is the proportion of variance in the species data that is captured by the environmental data.

Part



24 Obtaining help

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.

If you have problems using the program or entering data which you cannot solve then contact Pisces Conservation Ltd by e-mail - pisces@pisces-conservation.com - or by phone to +44 (0)1590 674000 during office hours (09.00 to 17.00 BST/GMT).

You might also like to consult our web site at www.pisces-conservation.com where you will find answers for many common problems.

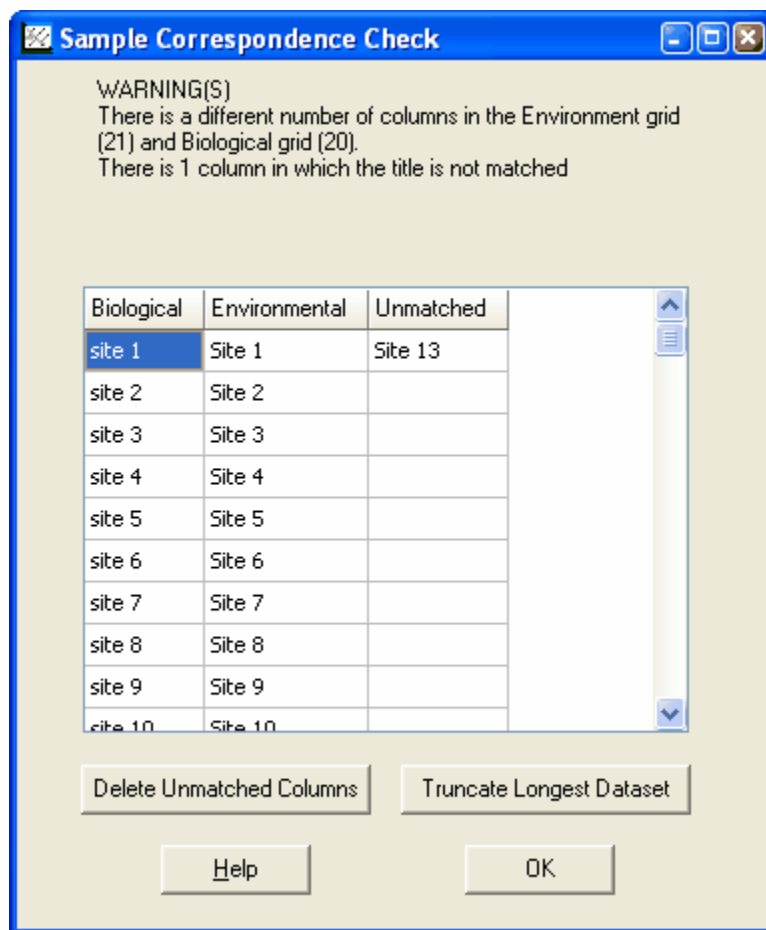
24.1 Common errors and problems

No matter how much care is taken with data preparation, it is not unusual to encounter problems when initially loading and running analyses on a data set. These are usually simple to rectify. Common error messages and their solutions are shown below

1. **" is not a valid floating point value.** This will occur if the raw data set holds blank columns or rows - ones that sum to zero. Remove blank columns and rows by using '[Handling zeros](#)' in the working data window. It may also occur if the raw data holds a blank cell. In some cases ECOM will identify the problem cell which should be edited. Normally it is because the data has been prepared in a spreadsheet using blanks to represent zero values.

Occasionally, this error can occur because a blank space has been accidentally entered into a cell outside the data matrix when it was being prepared in a spreadsheet program. 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.

2. **Data matrix disparity.** If ECOM detects that the biological and environmental data sets have different numbers of columns, then the following screen will appear, showing where the disparity lies. In this example, Site 2 is missing from the biological file. The options are to Delete the Unmatched Column, or to Truncate the Longest Dataset. This will, of course result in a mismatching of sites throughout the datasets, but it may be useful if dealing with errant entries at the end of either dataset.



3. I/O error 32 - access denied. This will occur if the data file you are trying to open is currently being used by another program - normally the spreadsheet which was used to organise the data.

Close the file in other programs and try again.

4. When a data file is opened, all the data are dumped into the first cell, rather than opening in the grid properly

This is usually because there are one or more blank cells in the second row of data. If the blanks are replaced by zeros, or another row with no blanks is put in the second position, the problem should disappear. Row 2 is the crucial one; blank cells are tolerated elsewhere.

5. 'I entered the data in a spreadsheet program, and saved it as a csv file. When I try to open it in ECOM, the numbers are separated by ';' not ',' so the analysis will not run.'

This is because some non-British systems use ';' instead of ',' to separate values in a csv file. Open Windows Explorer, and change the file extension from csv to txt (i.e. Filename.csv becomes Filename.txt). Open the txt file in Word or another word-processing program. Use the Find/Replace function to replace every ';' with a ',' then save the file again. Change the txt file extension back to csv. It should now open and run perfectly in ECOM.

6. Invalid floating point operation

This often indicates that there are rows or columns with no data in them. Use the Handling Zeros button on the working data tab to find and remove the offending row or column.

7. My files do not appear to be saved

This may be caused by the lack of the correct file extension. The program only shows files with the .csv extension. To see other files in the open file dialog type *.* into the file name box and press return. - WARNING this will show all the files in the directory - whether they are compatible with the program or not. If you have saved a file without an extension either add the extension outside the program or open the file and save with the correct extension using 'Save as...'

8. When I press F1, no Help screen appears.

Ensure that the window on which you are seeking help is the active one.

24.2 References

Jongman, R. H. G., Ter Braak, C. F. J. & van Tongeren (eds) 1987. Data Analysis in Community and Landscape Ecology. Cambridge University Press, England.

Legendre, P & Legendre, L. 1998. Numerical Ecology 2nd Ed. Elsevier, Amsterdam, 853pp.

Palmer, M. W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. Ecology 74:2215-30.

24.3 Demonstration version

The Pisces website contains a downloadable demonstration version of ECOM. It is the full version of the program with the following limitations.:

1. It is supplied pre-loaded with a single pair of [demo data sets](#) (Dune_species.csv & Dune_enviro.csv).
2. The data screens are not editable.
3. It is not possible to Copy/Paste from/to the data screens.

In all other respects the demonstration program is fully functional and will give you plenty of opportunity to explore its methods, features and graphical output quality.

Download it at <http://www.pisces-conservation.com/freestuff.html>, our page that contains all our program demo downloads, program manuals and e-book samples.

24.4 Citation

For publication purposes this product should be cited as follows:

Environmental Community Analysis Version 2.1, 2007, Pisces Conservation Ltd. Lymington, UK (www.pisces-conservation.com)

or alternatively, if you prefer...

Henderson, P.A. & Seaby, R. M. H., 2007, Environmental Community Analysis 2.1. Pisces Conservation Ltd, Lymington, UK.

24.5 Contacting PISCES for help

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-mail - pisces@pisces-conservation.com - or by phone to +44 (0)1590 674000 during office hours (09.00 to 17.00 BST/GMT).

You might also like to consult our web site at www.pisces-conservation.com where you will find answers for many common problems.

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

Index

- 2 -

26 Similarity and distance measures 92

- A -

adjust 29
adjustments 29
Agglomerative clustering 92
allocate 40
allocating 40
Analysis of Similarity 92
ANOSIM 92
arcsine 29
aspect 50
assigning 40
assistance 131
Association analysis 92
authors 133
average 37

- B -

base10 29
Bibliography 133
binary 48, 56
binary variables 48
Biological variables 37
Biplot 55, 57
bitmap 78
by mean 30

- C -

canonical 46
Canonical coefficients 59
Canonical Correspondence Analysis 46, 93
categorical variables 48
CCA 46, 93
chart 79
chart export 78
chart format 78
charts 78, 79, 85
choice 37
circular 50

Circular Statistics 97
coefficients 59
colinearity 48
Co-linearity 48
Collinearity 48, 49
column statistics 37
columns 31, 32
common 131
compass 50
constant 29
Contact 133
contacting 133
copy 84
copying 78
correlation 56, 59
correspondence 36, 46
counting zeros 37

- D -

data 17, 24, 25, 27, 36, 73
Data checks 36
data relativisations 30
data summary 34
decimal places 12
DECORANA 92
delete 31, 36
demos 17
Density from Distance 102
Depth 84
Detrended Correspondence Analysis 92
discriminant analysis 98
display options 12
Divisive cluster analysis 92
dummy 48
dummy variables 48
Dynamica 105

- E -

Ecom 93
Edit chart 84
editing 27
Eigenvalues 51
enhanced metafile 78
environment 59
environmental 59, 93
Environmental data 49
Environmental variables 50
Environmental variables 37, 47, 56

Errors 131
Excel 15
Explanatory variables 56
exporting 78

- F -

File 19
File Menu 19
Fuzzy Grouping 98

- G -

getting started 6, 9, 12
Graph 54
graph export 78
Graphical output 54
graphics 64, 78, 79, 84, 85
graphs 12, 64, 78, 79, 84, 85
grouping 40
groups 40

- H -

Hedgerow Assistant 106
help 131, 133
Hinkley 17

- I -

image format 78
Improvements 4
input 6, 21, 24, 25, 27
install 12
installation 12
instructions 12, 21
Introduction 2

- J -

Jolly-Seber 100
jpeg 78

- K -

K means 98
kurtosis 35, 37

- L -

labels 12
large array 27
Large data sets 27
large files 27
large matrix 27
linear 62
linear regression 62
loading from Excel 15
log 29
logarithmic transformation 49
logs 49

- M -

main form 6
maximum 27, 37
maximum value 30
mean 30, 35, 37
median 35, 37
Memory 12
menu 19
menus 6
metafile 78
methods 12, 46
Microsoft Excel 15
minimum 37
Monte Carlo 58
Move 84
MR 93
MS Excel 15
Multiple Regression 62, 93

- N -

natural 29
NMDS 92
Non-metric Multidimensional Scaling 92

- O -

operating system 12
options 12, 79
Ordination 54
Other programs 91
output 34, 51, 53, 55, 57, 59, 64, 78, 79
Output screens 34

- P -

panning 85
PCA 92
Pearson 56
physical 47
physical data 49
physical variables 56, 59
Pisces 133
plot 54, 59, 64
plots 85
plotting 54
power 29
predictors 47, 56
preferences 12
Principal Components Analysis 92
print 79, 84, 88
printing 6, 78, 79, 88
printing grids 88
printing output 88
problems 131

- Q -

quick guide 9

- R -

Randomisation 58
Randomization 58
Rank 55, 57
RDA 93
Reciprocal Averaging 92
Redundancy Analysis 93
References 133
Regression 62
Relative 30
Removal Sampling 101
removing zeros 31
results 34, 51, 53, 54, 55, 57, 59
Rotate 84
row statistics 37
rows 31, 32

- S -

sample number 37
samples 73
save 78

saving 27, 78
saving grids 88
saving output 88
scores 53, 55, 57
SDR 95
selection 47
sets 40
Setup 12
Similarity Percentages 92
SIMPER 92
Simply Growth 103
Simply Probit 107
Simply Tagging 100
singular 48
size 24
skewness 35, 37
software 91
species 73
Species Diversity & Richness 95
Species Filtering 92
species number 37
speed 27
square root 29
standard deviation 30
statistics 34, 35
step by step 9
Stepwise Regression (SR) 93
Student t 62
summary 34, 37
sums of squares 37
system requirements 12

- T -

t value 62
TeeReader 78
tests 58
transformation 49
transformations 29
transposing 32
truncate 36
TWINSPAN 92
Two-way INDicator Species Analysis 92

- U -

user preferences 12

- V -

variance 35, 37, 51

vectors 59

- W -

what's new 4

working 73

working data 29

- Y -

year 50

- Z -

zero 31, 37

zoom 84, 85