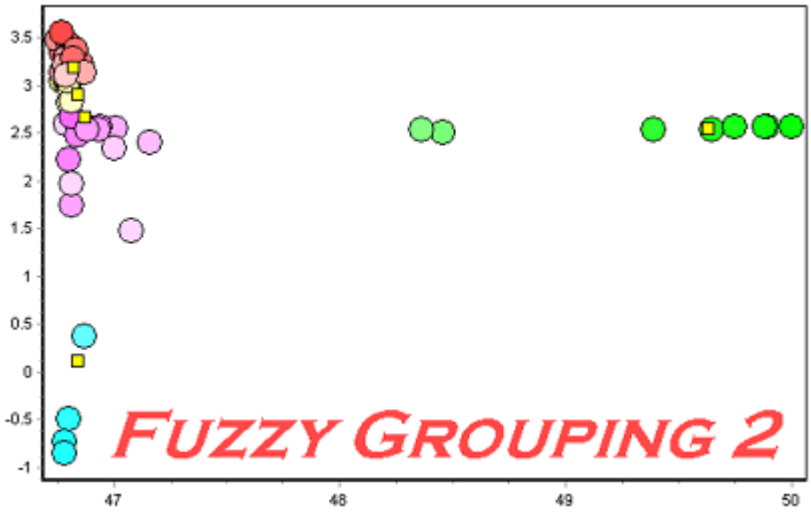


Fuzzy Grouping 2.0

Fuzzy Grouping 2.0

by R Seaby, P Henderson

Fuzzy Grouping

Printed: September 2014 in Lymington, UK.

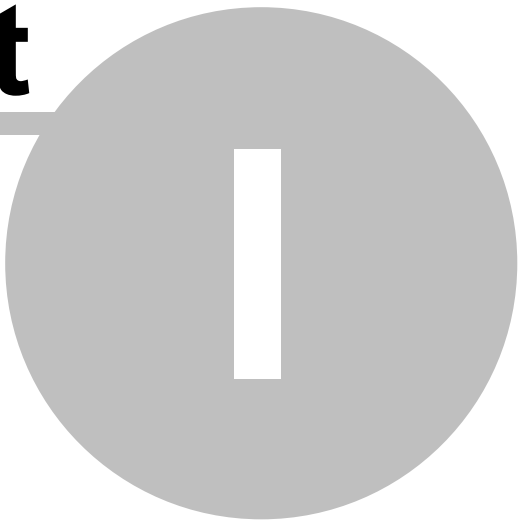
Table of Contents

Licence Agreement

PISCES LICENSE AGREEMENT
This is a legal agreement betw een 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 w ritten 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 softw are product acquired w ith 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 w hen 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 netw ork server for the sole purpose of distribution to other computers is not 'in use'. If the anticipated number of users of the SOFTWARE w ill 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 netw ork 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 ow ned by PISCES or its suppliers and is protected by all applicable national law s. Therefore, you must treat the SOFTWARE like any other copyrighted material (e.g. a book) except that if the softw are 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 w ritten 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 w ritten 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 w arrants that (a) the SOFTWARE w ill perform substantially in accordance w ith the accompanying Product Manual(s) for a period of 90 days from the date of receipt; and (b) any PISCES supplied hardw are accompanying the SOFTWARE w ill be free from defects in materials and w orkmanship under nominal use and service for a period of one year from the date of receipt. Any implied w arranties on the SOFTWARE and hardw are are limited to 90 days and one (1) year respectively or the shortest period permitted by applicable law , w hichever 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 hardw are that does not meet PISCES'S Limited Warranty, and w hich is returned to PISCES w ith a copy of your receipt. This Limited Warrant is void if failure of the SOFTWARE or hardw are has resulted from accident, abuse or misapplication. Any replacement SOFTWARE w ill be w arranted for the remainder of the original w arranty period or 30 days, w hichever 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 law s 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 w rite:
PISCES Conservation Ltd
IRC House, The Square
Pennington, Lymington
Hampshire, England
SO41 8GN
Tel 01590 674000

Part



1 Contents

Part II

2 Introduction

2.1 Introduction

Fuzzy Grouping is a Windows® program that offers fuzzy clustering methods that are now becoming used by community ecologists. Researchers in other fields such as palaeontology, archaeology and the social sciences may also use these methods. Easy to use programs for the PC to carry out these techniques and aimed to meet the needs of ecologists are not presently available, and this has undoubtedly slowed the introduction of these techniques into ecological analysis. 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 wide range of scientifically 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 hypothesised 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, fuzzy c-means, fuzzy linear discriminant analysis and fuzzy ordination. 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 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. The third, fuzzy ordination, is appropriate when you have information on possible group membership. For example, the samples might be taken along a transect up a mountain slope and the altitude can be used as a measure of likely membership of a sample to the high altitude community group. The program also undertakes DECORANA - Detrended Correspondence Analysis - primarily as a means of allowing the fuzzy clusters to be visualised in a 2-dimensional space. However, the program can also be used to produce a standard ordination using this powerful method. A final feature that can be useful is the production of summary statistics on each row (species) and column (sample).

Data can be organised using standard Windows programs such as Excel, and the output from Fuzzy Grouping is displayed, exported and printed using standard Windows techniques. This results in a program that is easily 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 their data within a familiar Windows setting. The input data set is arranged as a two dimensional array. In ecology, it is usual for the samples, which are normally collected from set localities and may be called, for example, quadrates or stations, to form the columns. The variables for each sample are the rows and usually comprise the numbers of each species or other taxon observed. The fuzzy ordination

analysis requires the user to input observations of one or more variables reflecting the physical or biological change along the transect or possibly your assumed membership. For example, if the results were for quadrates taken along a transect up a hill, the physical variable could be the altitude.

The program will run on Mac computers with Windows emulation software. Fuzzy Grouping uses the same data structure as Species Diversity and Richness II, Community Analysis Package (CAP) and ECOM, which produce a wide range of diversity and species richness measures, multivariate ordinations and Canonical Correspondence Analyses respectively. Together, these four programs offer an extensive range of methods for the analysis of ecological communities.

To help users to understand how to use Fuzzy Grouping, the instructions are written from an ecological viewpoint. Below, we generally refer to columns as samples, sites or quadrates, and rows as species. This is the way an ecological data set should be organised to study both the relationship and similarity between samples and the pattern of association between species.

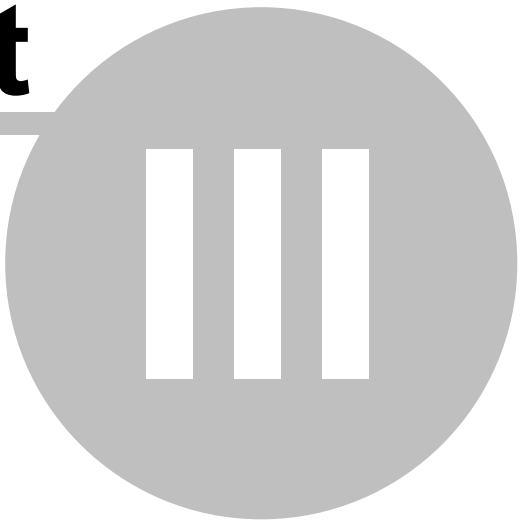
How to reference Fuzzy Grouping.

2.2 How to Reference Fuzzy Grouping

This program should be referenced thus:

Henderson, P.A., Seaby, R.M.H., and Somes, J.R. (2014). Fuzzy Grouping, version 2. Pisces Conservation Ltd., Lymington, Hampshire, UK. http://www.pisces-conservation.com

Part



3 Installation

System requirements:

1. A PC running Windows XP or later
2. 5 MB of spare hard disk space

Installation:

1. Place the Fuzzy Grouping 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 click file named Setup.exe in the root directory.
3. When installation is complete, there will be a Fuzzy Grouping entry under Start: Programs.
4. An uninstall program will also be created, if you wish to remove Fuzzy Grouping.

Part IV

4 Quick guide to running a data set

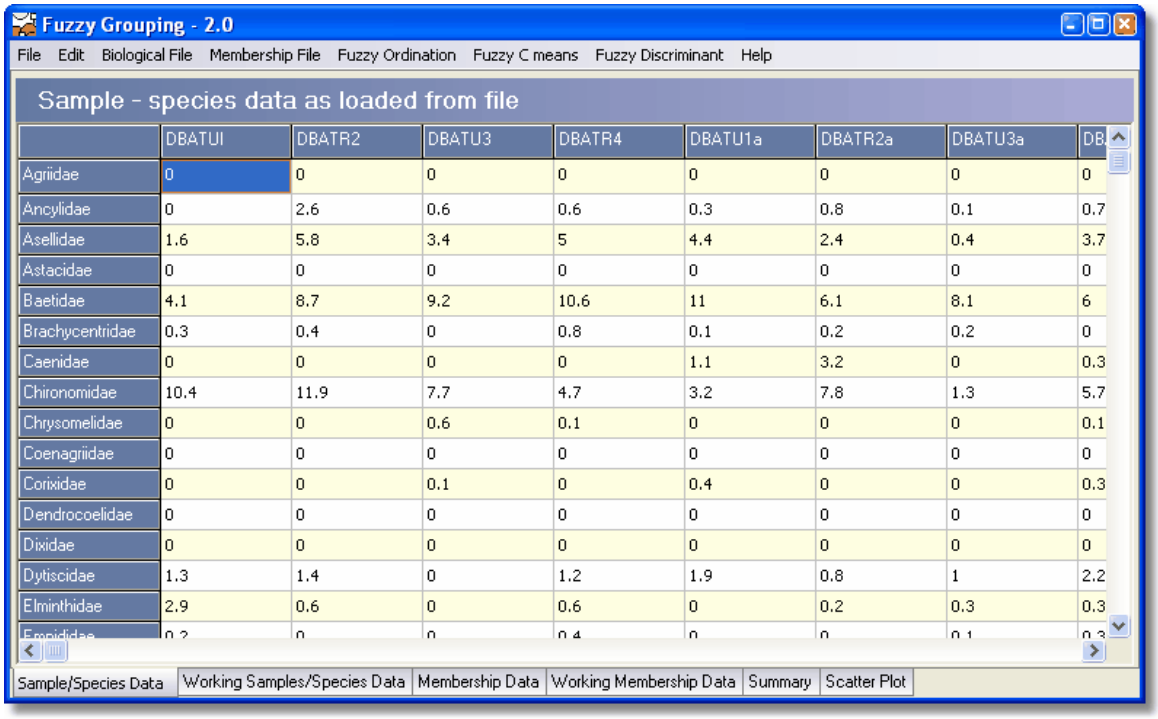
1. Prepare your species and environmental data sets using a spreadsheet such as Excel. The data are best arranged with samples (quadrates, or stations) as columns and species (variables) as the rows. Give a title to each sample and species. Values can be real or integer, place 0 (a zero) in all localities where a species was not recorded. This can be automated by searching for blanks and replacing them with 0. Make sure you use a zero and not a capital o!

2. Save these arrays as CSV (comma delimited text files) from the Save As menu. Take care if using Excel that the work sheet does not hold any other data or this will be included in the CSV files. It can be helpful to select the first 10 or so rows below, and columns to the right of, the block of data in your data set, and use 'Delete' to remove any blank spaces or other unwanted data accidentally entered in cells outside the range of the data set.

3. Close the spreadsheet data file before opening the data in Fuzzy Grouping, otherwise you will get an 'I/O error 32 - access denied' error message. This is because Excel 'locks' any file it has open, to prevent it being modified by another program while it is in use.

4. Start Fuzzy Grouping and open your data files from the 'Biological File' and 'Membership File' menus.

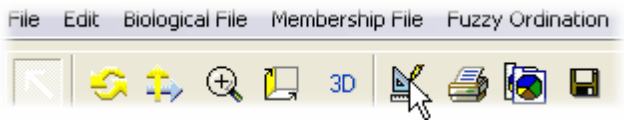
5. The data will now open in the raw data windows (see image below). Before choosing a method it is wise to check your data for columns that sum to zero (samples with no species) and rows that sum to zero (species not found in any sample). First click on the Working Data tabs. To remove zero sum rows or columns, select Zero Issues - 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 adjusted data sets for analysis.



	DBATUI	DBATR2	DBATU3	DBATR4	DBATU1a	DBATR2a	DBATU3a	DB
Agriidae	0	0	0	0	0	0	0	0
Ancylidae	0	2.6	0.6	0.6	0.3	0.8	0.1	0.7
Asellidae	1.6	5.8	3.4	5	4.4	2.4	0.4	3.7
Astacidae	0	0	0	0	0	0	0	0
Baetidae	4.1	8.7	9.2	10.6	11	6.1	8.1	6
Brachycentridae	0.3	0.4	0	0.8	0.1	0.2	0.2	0
Caenidae	0	0	0	0	1.1	3.2	0	0.3
Chironomidae	10.4	11.9	7.7	4.7	3.2	7.8	1.3	5.7
Chrysomelidae	0	0	0.6	0.1	0	0	0	0.1
Coenagriidae	0	0	0	0	0	0	0	0
Corixidae	0	0	0.1	0	0.4	0	0	0.3
Dendrocoelidae	0	0	0	0	0	0	0	0
Dixidae	0	0	0	0	0	0	0	0
Dytiscidae	1.3	1.4	0	1.2	1.9	0.8	1	2.2
Elminthidae	2.9	0.6	0	0.6	0	0.2	0.3	0.3

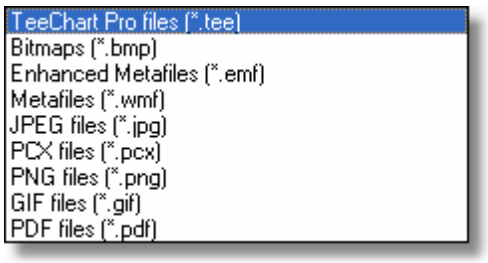
6. Undertake any transformations or relativisations you may wish on the Working data tabs. Note that any changes undertaken on the working data will not change the raw data.

7. Select a method from the Fuzzy Ordination, Fuzzy C means or Fuzzy Discriminant menus, and use the tabbed windows to examine the output. All aspects of the charts and plots can be altered using the buttons on the toolbar above the chart (image below). For more instructions on editing charts, see 'Printing, editing and exporting your results'.



8. Use File:Print or the Printer button on the toolbar to print the charts and Edit: Copy or the Copy button to copy it to the clipboard for pasting into another application.

9. Use the Save button on the toolbar to save the chart as an image file for use in another application, or as a TeeChart Pro (*.tee) file. If you intend to paste the image into an application (such as Microsoft Word), the Enhanced Metafile is the most useful format, since it can be resized by dragging, without distortion or loss of detail. TeeChart Pro files (*.tee) include the data series which make up the chart, and can be manipulated using the free TeeReader software supplied on the installation CD, so that you can save and alter the chart at a later date.



10. See the example application.

Part V

5 Worked Example

To help the user understand the full capabilities of the program and its methods, we present below a worked example, using two data files installed with the program in C:\Program Files\Fuzzy Grouping, AllRiversBiological.csv and AllRiversMembership.csv.

The data were collected in 2000 as part of an investigation into the effects of river restoration on biodiversity in several chalk streams in southern England. The report on the investigation, Effects of physical restructuring of channels on the flora and fauna of three Wessex rivers, can be downloaded in Adobe Acrobat PDF format from our website at http://www.irchouse.demon.co.uk/latestreports.html

Concentrating on macroinvertebrates, the biological data comprise percentage composition by family from restored and unrestored reaches of the 3 rivers and 2 tributaries. There are 61 families (rows) across 49 sample sites (columns), with 7 environmental/membership variables. The variables are:

Binary membership of restored/unrestored status
A code designating the river the sample was taken from
Maximum river depth at sample site
Standard deviation of depth
Mean width
Level of shading by overhanging trees
% of riverbed covered by Water-crowfoot, Ranunculus spp.

The site names are coded with the initial letter of the river and 'U' or 'R' for unrestored or restored. The rivers are the Wylye, its tributary the Till, the Piddle, its tributary the Devil's Brook, and the River Avon. So for instance, the unrestored stretch at Hyams Farm on the Avon is coded Ahyam U, and the restored site at Great Wishford on the Wylye is coded Wgwish R.

For this example, our hypothesis is that, if assumptions on changes in biodiversity attributable to river restoration are correct, sites can be grouped according to their membership of restored or unrestored status.

A. Fuzzy C-Means analysis

To begin the example, open AllRiversBiological.csv and AllRiversMembership.csv from the Biological File and Membership File menus.

1. On the Working Samples/Species Data and Working Membership Data pages, use Select: 0 issues: Delete 0 Rows and Delete 0 Columns. If any 0-sum rows or columns are found, a message will say so.

2. Since you have no firm a priori evidence of group membership you should

calculate the optimum number of clusters for grouping the sites. Start Fuzzy C-means: Compare Partition Coefficients, set the minimum number of clusters to 2, and the maximum to 5; click OK. Click on the 'C Means Summary' tab; the coefficients for the range of cluster numbers are displayed, as below:

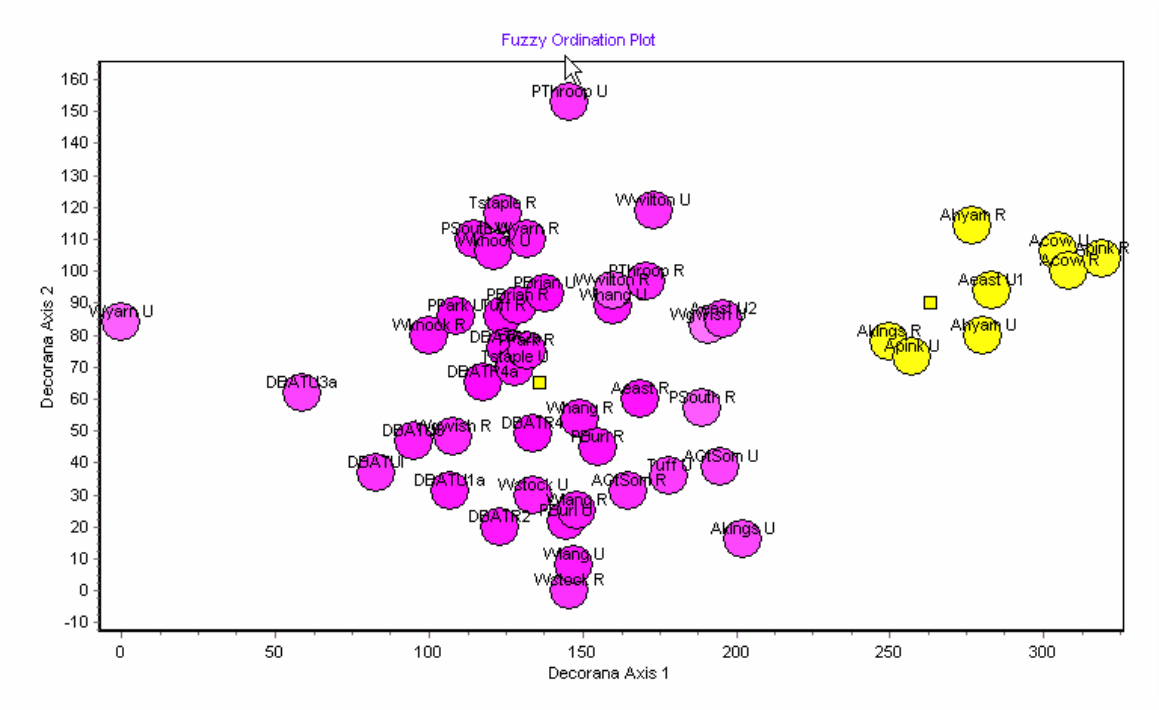
Summary information partition coefficient

Partition coefficient 2 Group	0.5892
Partition coefficient 3 Group	0.3513
Partition coefficient 4 Group	0.3185
Partition coefficient 5 Group	0.2628

The largest coefficient, 0.5892, is for 2 groups, indicating that this is the optimum cluster number.

3. Perform the Fuzzy C-Means calculation: click Fuzzy C-Means: Analysis for selected cluster number, in the Options dialog set the number of clusters to 2, and click OK.

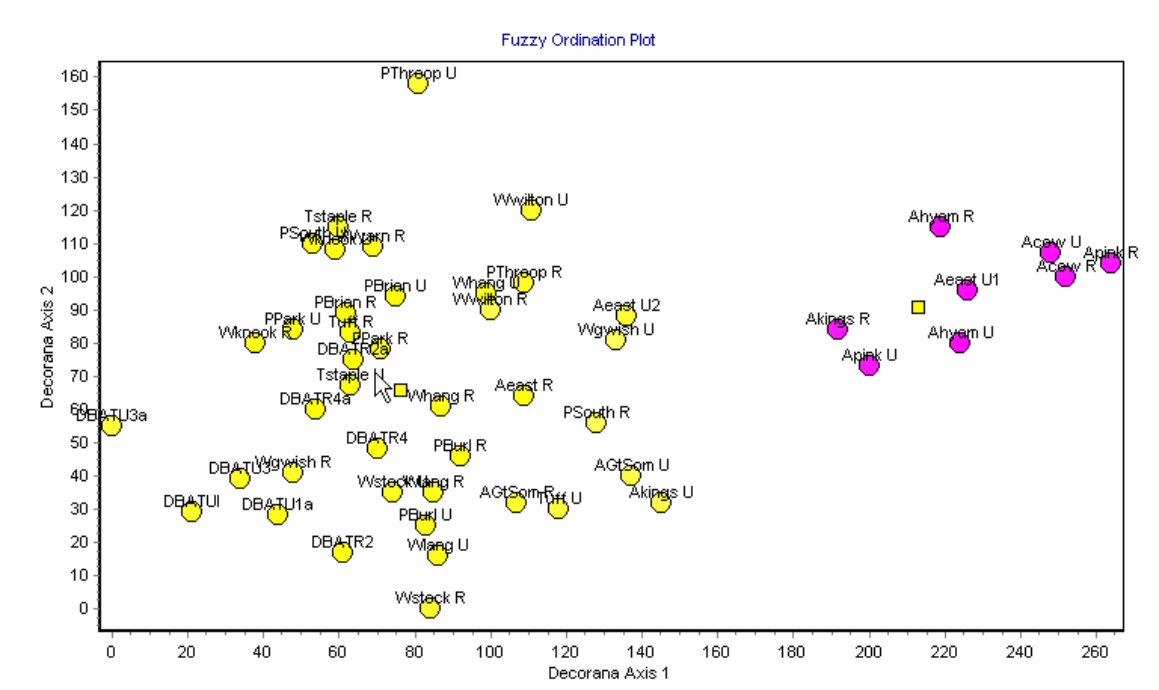
4. The results. From a first glance at the ordination plot (below), you can see that one site is positioned away from the others, to the left of the plot:



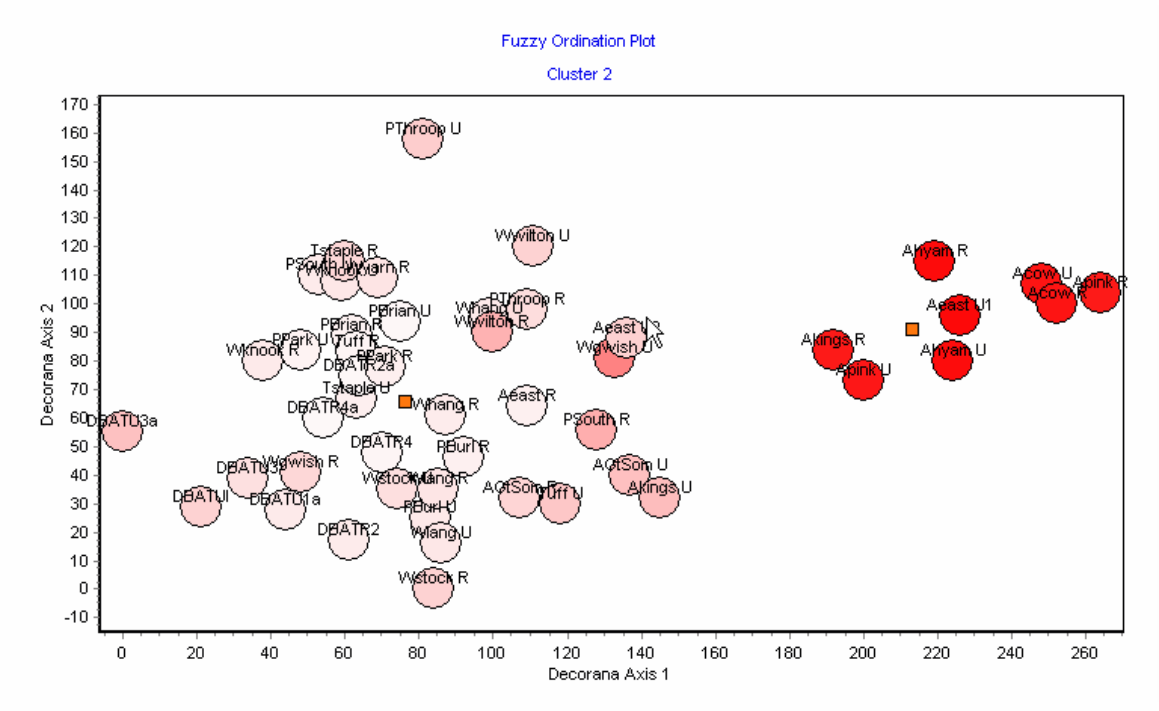
Predictably, the site is radically different from the other sites sampled, in that 90% of its individuals are of one family, the Simuliidae. It is therefore the least diverse site, and the most heavily dominated by one family, and this makes it unlikely to be easily clustered with the other sites. Removing this site (Wyarn U) from the analysis would

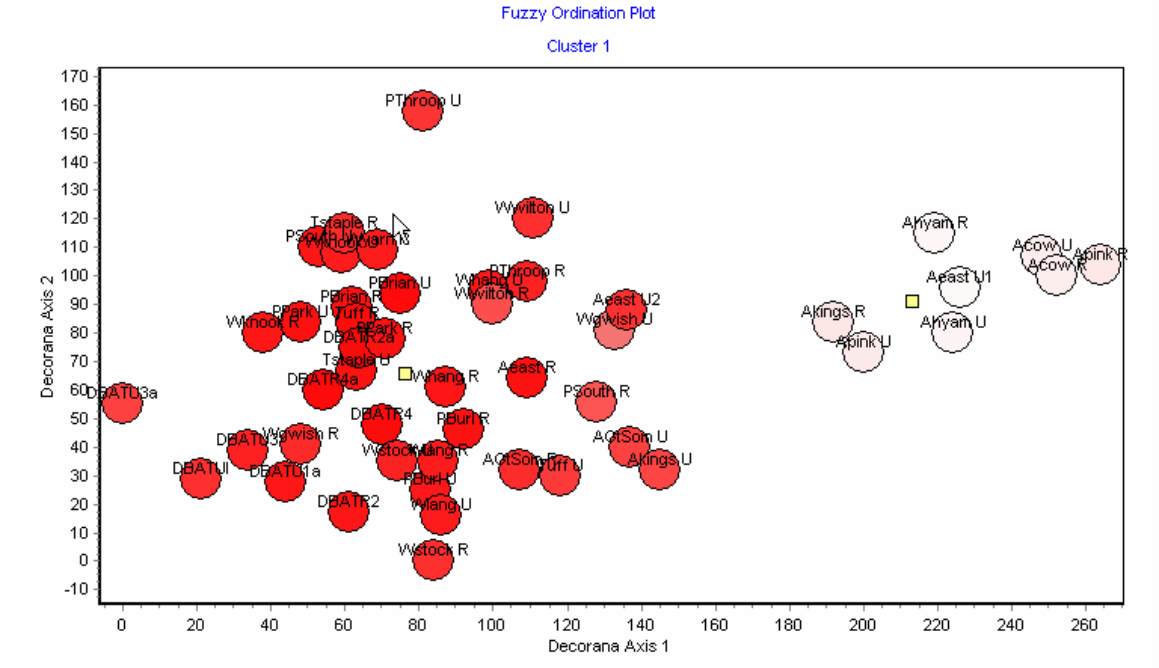
improve the spread of the rest of the points in the plot, since the range of the axes would be smaller. On the Working Samples/Species Data and Working Membership Data pages, use Select/0 issues: Deselect Column to remove column Wyarn U from the working data sets, and repeat step 3.

Having removed the rogue site, you can see from the plot below that the spread of the points is much larger, allowing for easier interpretation. Adjust the relative size of the points with the Relative Diameter slider. The plot below shows the 'Different colours' option, which makes it easier to distinguish clusters (particularly if there are more than 2), but does not show 'fuzziness', or the likelihood of belonging to one cluster or another.



If you select the 'Colour Intensity' option, you can see from the two plots below that the likelihood of membership is much more easily observed, since it is depicted by the intensity of the colour fill of the circle:





Switching the view back and forth between the two clusters, you can see that one site (Wgwish U) has a roughly equal likelihood of belonging to either cluster.

When you switch on the Site Labels, it is apparent that the cluster on the right of the

plot comprises 8 sites from one river, the Avon, with the other cluster holding all the other sites. Clearly, in this particular Fuzzy C-means analysis, sites cannot be reliably grouped according to their restored/unrestored status, and in the case of most of the sites on the River Avon, river membership is considerably more significant. To conclude, there are two major clusters the sites can be grouped into, and these appear to be related to the river, rather than the restoration status.

B. Fuzzy Ordination

The objective of the analysis is to compare the actual ordination of the sites with respect to some environmental variable or classification with the ordination generated from the observed species communities.

The relationship between the biological community and the environmental variable may be based on actual physical readings, or might be hypothetical. Sites could, for example, be assigned values that represent their relative position along a gradient. For example, for our rivers, the gradient might be the depth of water, or the width of the river. Alternatively, sites could simply be given a value between 0 and 1 that denoted their membership of a set, for example restored/unrestored.

Using the same data set as for the Fuzzy C-Means calculation above, we will now perform the Fuzzy Ordination analysis. As above, open the two data files, using the Working Samples/Species Data and Working Membership Data pages, use Select/0 issues: Delete 0 Rows and Delete 0 Columns to remove any 0-sum rows. In addition, on the Working Samples/Species Data and Working Membership Data pages, use Select/0 issues: Deselect Column to remove column Wyarn U, the aberrant sample, from the working data sets.

To show the method we will undertake two ordinations, the first using restored/ unrestored classification, the second using the percentage cover of Water-crowfoot, Ranunculus spp.

1. Analysis: restored/unrestored status.

Click Fuzzy Ordination to begin the analysis; on the Similarity dialog box, choose Memb Unrest (membership unrestored) as the environmental variable to use for the observed ordination, and then select a similarity measure. Since we are using quantitative species data, then one of the two Quantitative measures is probably the best to use – for this example we have used Steinhaus Quantitative. The results are shown on the Fuzzy Ordination tab, illustrated below:

Actual and apparent scores

Site	DBATUI	DBATR2	DBATU3	DBATR4	DBATU1a	DBATR2a	DBATU3a	DBATR4a	PBurl U	PBurl R	PPark U	PF
Actual	1	0	1	0	1	0	1	0	1	0	1	0
Actual (Relative)	1	0	1	0	1	0	1	0	1	0	1	0
Apparent	0.6889	0.601	0.416	0.2695	0.3501	0.4492	0.7592	0.3565	0.2256	0.2787	0.3358	0.

The results, comparing the actual (relative) and apparent scores, demonstrate that the restoration status is a poor predictor of the species community. The lack of correlation can also be seen from the Ordination Plot.

1. Analysis: % Ranunculus cover.

Begin the Fuzzy Ordination, choosing '% Ranunc cover' as your environmental variable, and undertake the analysis as described above.

In this case, there is a clear relationship between the environmental score and the apparent score, as can be seen from the Ordination Plot below:

Actual and apparent scores

Site	DBATUI	DBATR2	DBATU3	DBATR4	DBATU1a	DBATR2a	DBATU3a	DBATR4a	PBurl U	PBurl R	PPark U	PP
Actual	80	80	80	80	90	80	90	80	30	60	60	70
Actual (Relative)	0.8889	0.8889	0.8889	0.8889	1	0.8889	1	0.8889	0.3333	0.6667	0.6667	0.7
Apparent	0.961	0.9319	0.9576	0.8872	0.9652	0.8976	1	0.908	0.8762	0.8061	0.9251	0.8

The correlation between Actual and Apparent scores can be seen on the Fuzzy Ordination tab. These results suggest that the degree of Ranunculus cover is influencing the invertebrate species community. You may now wish to perform Fuzzy Ordination analyses using the other environmental variables to examine the level of their influence on the species community.

C. Fuzzy Discriminant Analysis

This method is useful to produce the best plot showing the split between the groups.

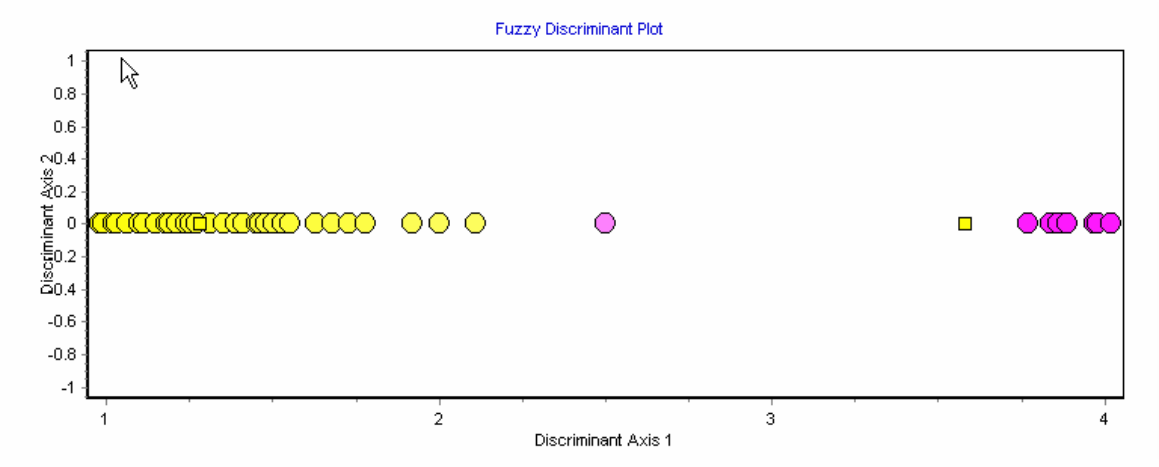
1. Select Fuzzy Discriminant and in the options window input for Number of Clusters
2. Remember that the Fuzzy c-means analysis above indicated that 2 groups gave the best split.

2. As there are only 2 groups the fuzzy discriminant analysis produces a 1-dimensional output. The coordinate of each site is given in the Discriminant coordinates grid shown below.

Discriminant Coordina

	Root 1
DBATUI	1.46
DBATR2	1.15
DBATU3	1.27
DBATR4	1.03
DBATU1a	1.15
DBATR2a	0.99
DBATU3a	1.68
DBATR4a	0.98
PBurl U	1.19
PBurl R	1.10
PPark U	1.03
PPark R	1.10

and the plot showing the clear separation of the two groups is shown using the cluster plot tab.



Part VI

6 Methods offered by Fuzzy Grouping

Part VII

7 Obtaining help

For most active windows context-sensitive help can be obtained by pressing F1, clicking on the Help button or selecting the Help dropdown menu, or clicking on the right-hand mouse button and choosing Help from the pop-up menu.

If you have problems using the program or entering data which you cannot solve then contact Pisces Conservation by e-mailing pisces@irchouse.demon.co.uk or by phone to England 44 (0)1590 676622 during office hours (09.00 to 17.00).

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

Telephone 44 (0) 1590 676622
Fax 44 (0) 1590 675599

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

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

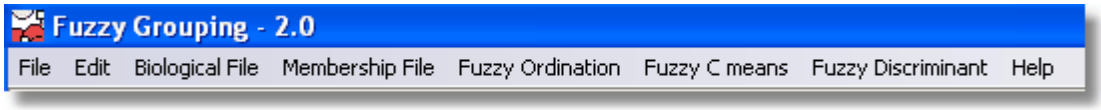
Part VIII

8 General Instructions

8.1 General instructions

Start Fuzzy Grouping in the normal Windows fashion either by clicking on the program icon or from the start button.
Along the top bar are a number of pull-down menus. These work in the same way as standard Windows programs.

File: To open and save data sets.
Edit: To cut copy and paste to and from the active window.
Biological File: To select and open the Species data file.
Membership File: To select and open the environmental/membership data file.
Fuzzy Ordination. - to undertake a fuzzy analysis of samples ordered along a gradient.
Fuzzy C means. - to undertake an analysis to allocate each sample to a fuzzy group and plot the results.
Fuzzy Discriminant. - to undertake a fuzzy discriminant analysis.
Help: to enter the Help system and show details about the program version.



When the program is started, you will be presented with a blank working area; first, open your Biological and Environmental data files. The raw data will be displayed on the 'Sample/Species data' and 'Membership data' pages; each has a corresponding 'Working data' page where you can manipulate and transform the data. 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 form will display the adjusted data.

The analyses will use the adjusted data set by default, changes made to the Working data will not be applied to the Raw data, so that you can always return to the raw data if you wish.

8.2 Maximum size of the data set and computation speed

Fuzzy Grouping version 2 has no fixed maximum size for the data sets it can handle. Limitations will be set by the speed and memory of your computer. In practice, the program has been found to perform well with 1000 species (variables) and 1000 samples. It is difficult to test the output of datasets much larger than 1000 x 1000 and the plotted output may become difficult to present given the resolution of computer screens. In general, the more sites used, and the greater number of clusters selected to assign those sites to, the slower the computation will be.

The number of colours on Cluster plots is limited to 14, to prevent the graphical output from being visually too confusing. If more than 14 clusters are specified to assign samples to, some of these when plotted will be displayed in the same colour if the "Different Colours (nearest)" or "Colours by Best Group" plots are displayed. The full range of clusters can still be distinguished by using "Proportional Diameter" or "Colour Intensity" to examine membership of one cluster at a time.

When using Fuzzy C-means or Fuzzy Discriminant, the Fuzziness factor should normally be left at its default value of 2, for normal sized data sets. Large data sets may benefit from having Fuzziness reduced to roughly 1.5. It is unwise to set Fuzziness much above, since this can give rise to erroneous results, where every site/sample has a reasonably high probability of membership of every group.

8.3 Preparing large data sets in a spreadsheet program

You do not need to create data sets within Fuzzy Grouping; 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.

The basic structure of the data is as follows.

A biological data set:

A membership (environmental) data set:

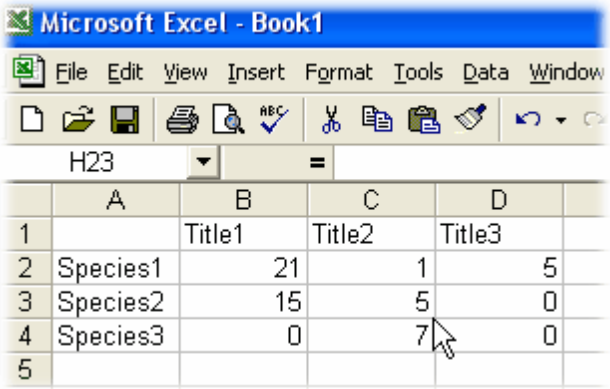
	Title1	Title2	Title3
Membership 1	1	1	0
Membership 2	0.9	0.5	0.1
Altitude	300	400	1000
Slope	0.25	0.4	0.7
Moisture	50	30	10

Although the environmental data set can contain more than one membership or environmental variable, only one is used at any one time.

The normal arrangement of community data within FUZZY GROUPING is to have the samples (quadrates) as columns and the species or membership variables as the rows. The data can be arranged with the species forming the columns, as the Transpose option within FUZZY GROUPING can be used to switch columns and rows. This can be useful if you have a very large number of samples, since Excel can only accept spreadsheets with fewer than 255 columns - in this case arrange your data set with Species in the columns, Sites/samples in the rows, then use Transpose in Fuzzy Grouping to switch the data. Alternatively, you can use our List Combiner utility, which is specifically designed for combining large data sets and exporting the data in the .csv format. More details are available on our website.

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. Note both the order of the variables and their type. 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. Cell A,1 should be left blank, otherwise put in zeros rather than leaving cells blank. The species names/ membership classes or environmental variables are input from row 2 in column 1.



	A	B	C	D
1		Title1	Title2	Title3
2	Species1	21	1	5
3	Species2	15	5	0
4	Species3	0	7	0
5				

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

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

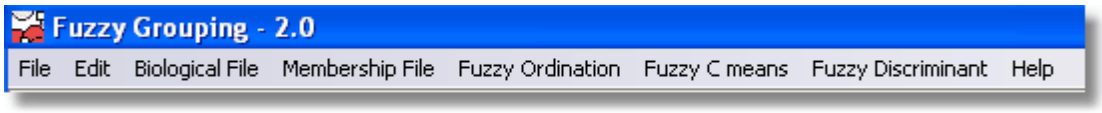
The csv file can also be opened and edited in a text editor (e.g. Notepad) or word processor (Microsoft Word), in which case it will appear like this:

,Title1,Title2,Title3
Species1,21,1,5
Species2,15,5,0
Species3,0,7,0
Species4,1,9,0
Species5,0,0,8
Note the leading comma on the first row which will make the first cell blank.

see also:
Opening a dataset ;
Creating a new data set within Fuzzy Grouping

8.4 Opening a data set

Note that a Biological data set is always required; the Fuzzy Ordination analysis also requires membership data.
Use Biological File and Membership File to select and open data files for analysis.



The default extension for Fuzzy Grouping files is *.csv , comma-separated text files.

Fuzzy Grouping comes with several sets of demonstration data.

These files can be found in the Fuzzy Grouping folder in C:\Program Files.

8.5 Creating and editing a data set

Preparing large data sets in a spreadsheet program
Data entry from within Fuzzy Grouping
Editing existing data
Saving edited data or creating a new data file

8.6 Data entry from within Fuzzy Grouping

Data sets can be created and edited within Fuzzy Grouping. Creating Species and Membership data sets is done in exactly the same manner, described below.

To create a new data set select File: New from the pull-down menus. After the new file is opened you are presented with a dialog where you can specify the number of rows and columns in the grids. The light 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, cluster or variable names. Leave the upper left-hand cell free. To input text or data, click on the cell and type a number. Numbers can be either integer or real; some methods may require integers, but 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 and when pressed with the cursor in the lower right hand column will add a new column. To add a new row, press the Insert key. A row will be inserted below that currently selected. Please note, if the text within a box is selected (it will usually appear dark blue) insert will not work. Simply click on the left-hand border of the cell again to select the entire cell and insert will be activated. To remove a row, click in any cell in the row, then press the Delete key.

8.7 Editing existing data

The raw data grids 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 Biological File : Save Biological or Membership File : Save Membership is selected.

Fuzzy Grouping 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 grids can also be edited. Changes made to the working grids 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.

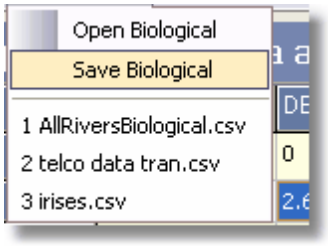
8.8 Saving edited data or creating a new data file

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.

8.9 Sample/Species data

This screen shows the raw data as opened from the species data file. If you wish to change the data, this is the place to do so. Just click on a cell in the grid and type in the new number.

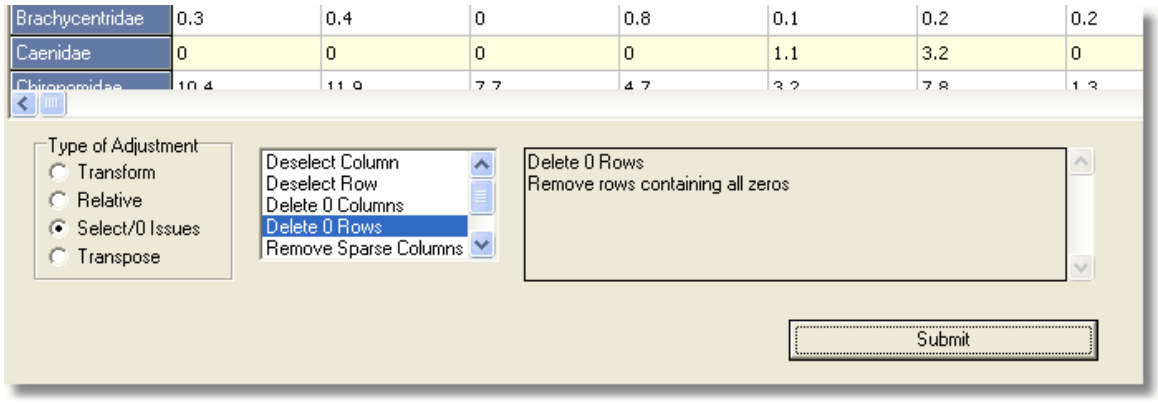
To save your changes use the Biological File: Save Biological dropdown menu.



If you do change data, you will then need to click the 'Submit' button on the Working Samples/Species Data page, to ensure that the analyses use the amended data set.

8.10 Working Samples/Species Data screen

This screen allows the user to make a variety of changes to the raw 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.



Data transformations
Relative adjustments
Dealing with zeros
Transposing data

8.11 Membership data

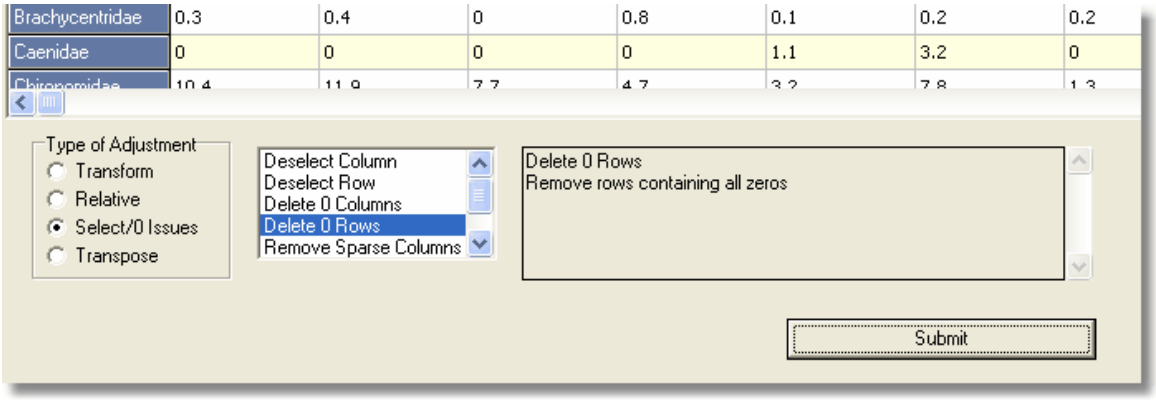
This screen shows the raw data as opened from the membership/environmental variable data file. If you wish to change the data, this is the place to do so. Just click on a cell in the grid and type in the new number.

To save your changes use the Membership File: Save Membership dropdown menu.

If you do change the data, you will then need to click the 'Submit' button on the Working Membership Data page, to ensure that the analyses use the amended data set.

8.12 Working Membership Data

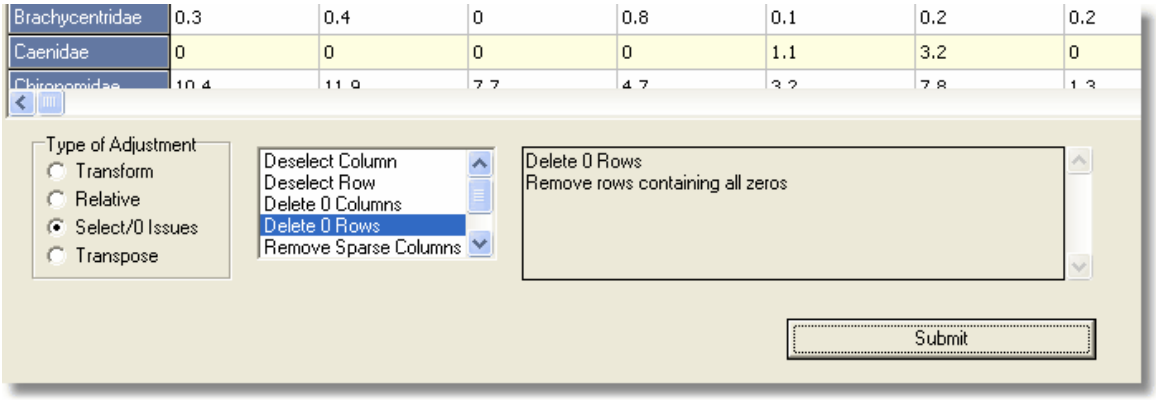
This screen allows the user to make a variety of changes to the raw membership data prior to undertaking an analysis. Initially, the user will be presented with a grid filled with the raw data - this can be adjusted using the options in the panel below the data grid.



Data transformations
Relative adjustments
Dealing with zeros
Transposing data

8.13 Data transformations

The values in the Working Data grid 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 Fuzzy Grouping are given below. Select the transformation to be made and click on Submit to make the change. The transformation options within Fuzzy Grouping are itemised below.

Revert to raw - This will cause the working data to revert to the raw data.

Log(10) - Each value is transformed to the log to base 10. This cannot be done for numbers <= 0.

Log10(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.

Log e (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 xa, 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.

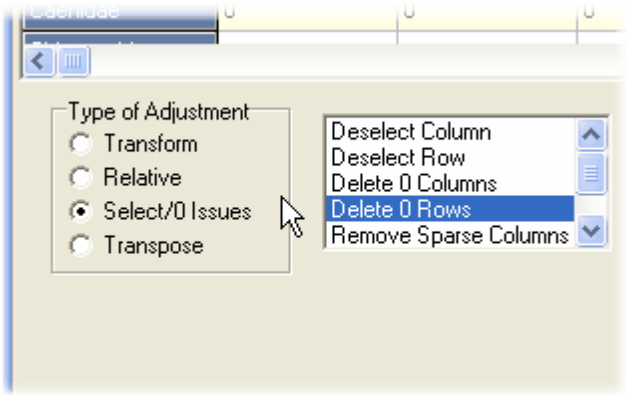
Multiply by constant - Each value is multiplied by a constant value chosen by the user.

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

8.14 Dealing with zeros

For some methods it is important that you do not have data rows or columns that sum to zero, since this can lead to division by zero errors.

Rows or columns in the working data holding zero values can be removed. On the Working Data page select 0 issues 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.

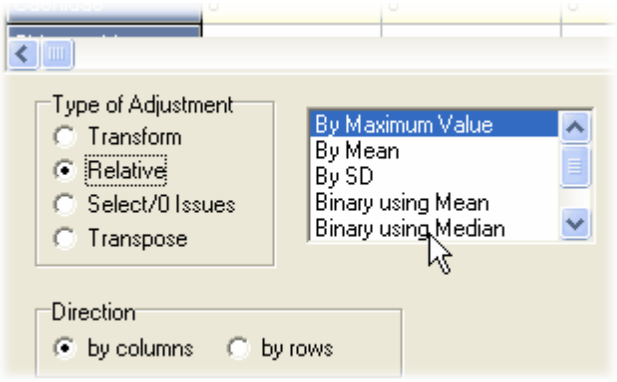
Beals - Beals smoothing is used on sparse data sets where many of the samples hold a small proportion of the total species list. As the transformation only uses the presence/absence of species, it should not be applied to good quality quantitative data as you will be discarding much of your information. 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, eij, 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, Ni is the number of species in column (sample) i, Bothij is the number of samples holding both species j and k, Samplesk is the number of samples holding species k and maxS is the total number of columns (samples).

8.15 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 Fuzzy Grouping are given below. In each case where by row or column is not stated the user 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 SD - 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 Column - This allows the user to define a general relativisation to be applied to each column. Each value is divided by

$$\left(\sum_{j=1}^{j=n} x_j^a\right)^{1/a}$$

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

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

$$\left(\sum_{i=1}^{i=n} x_i^a\right)^{1/a}$$

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

8.15.1 Summary of data

8.15.1.1 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 data set. This can be changed using the options below the grid.

General dataset statistics
Statistics for columns and rows

8.15.1.2 Statistics radio buttons

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.

8.15.1.3 Data set statistics

Summary statistics for both the raw and working data sets are displayed by clicking on the summary tab. When first activated the data grid will display the following general statistics for the working data.

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

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

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

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

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

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

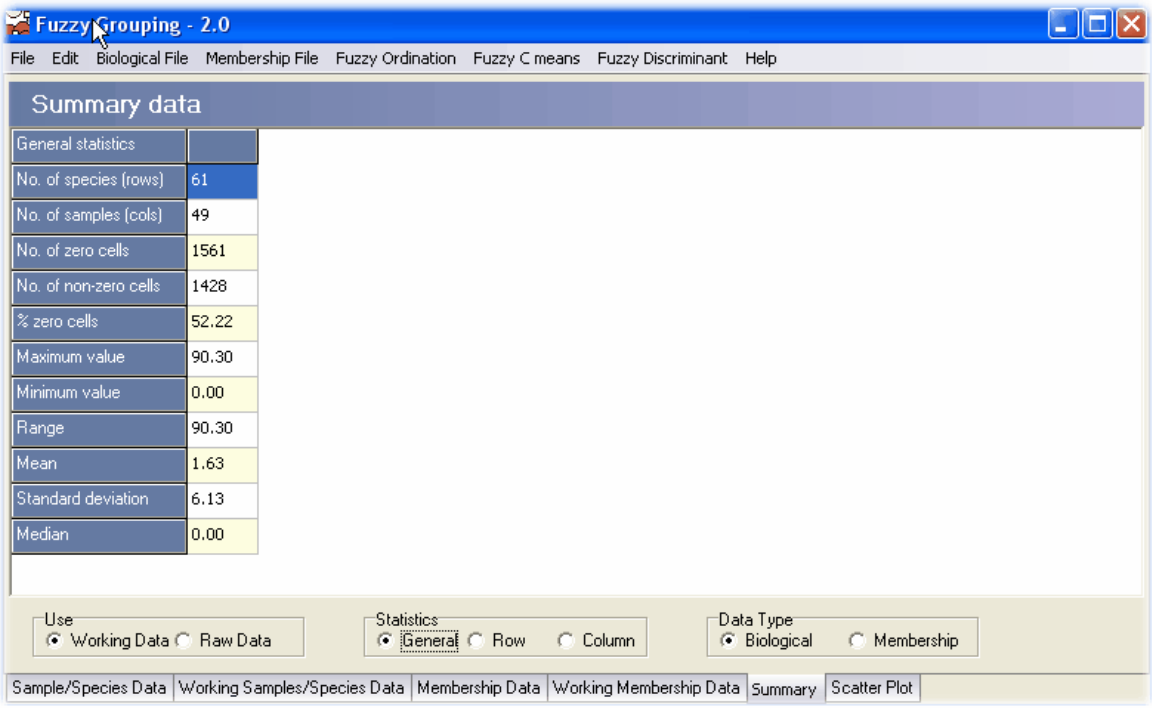
Mean - This is the mean of all the values in the data matrix.

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

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

To obtain general statistics on the raw data use the radio buttons in the Use panel situated below the grid to select raw data.

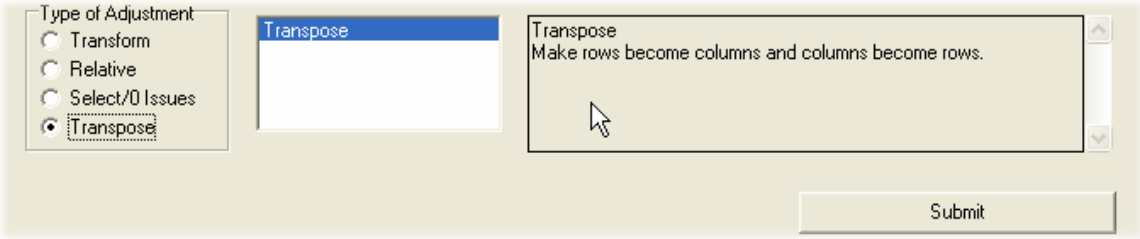
Statistics for the individual rows and columns of either the raw or working data matrices are selected using the Statistics radio buttons situated below the grid.



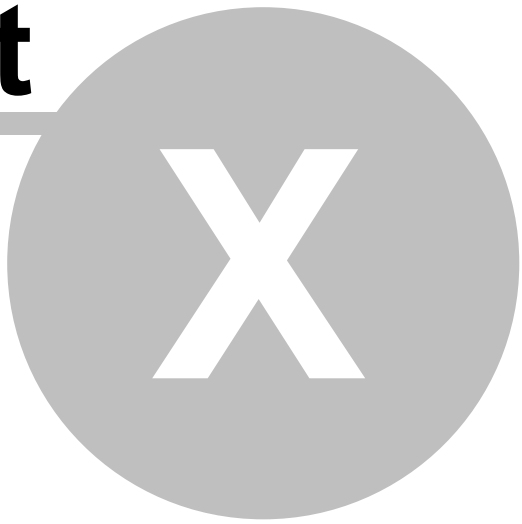
Part IX

9 Transposing data

Use this option to switch the rows and columns of the data set. Like all the other adjustments it is applied to the working data set. Select Transpose in the Type of Adjustment radio box and click on the Submit button.



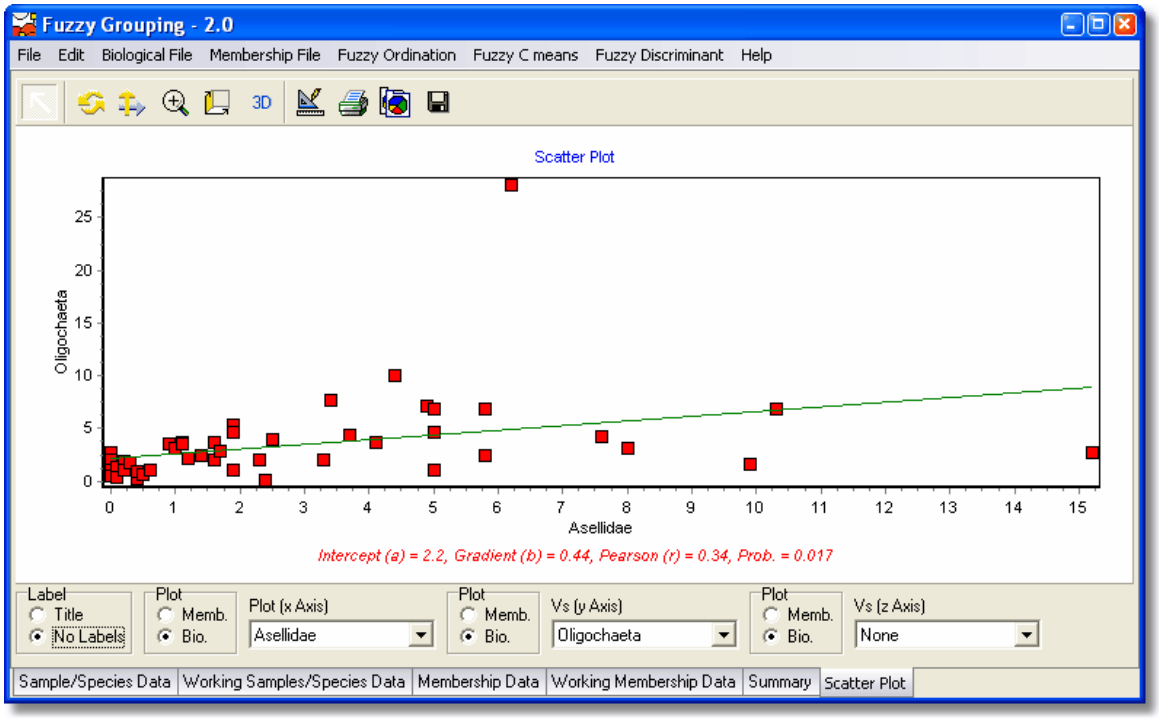
Part



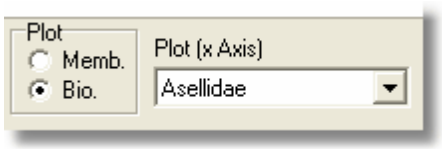
10 Scatter Plot

Click on the scatter plot tab to plot any two variables. This will allow you to search for relationships between variables or variables and their pre-defined group membership.

The plot also includes a straight line fitted by linear regression. Printed in red below the graph there are details of the regression including the intercept, gradient, the correlation coefficient (r) and the significance of the observed correlation (p). A significant relationship probably exists if p is less than 0.05.



Select Memb. if a variable from the membership array is to be plotted and Bio. if a variable from the biological file is to be plotted. Choose the variable to be plotted from the drop-down menus. You can activate the drop-down menus and then quickly scroll through the variables using the up and down arrow keys.



Part XI

11 Printing, editing and exporting your results

Output, both graphical and text, can be exported as a file, copied to the clipboard or printed. Fuzzy Grouping also offers a wide variety of options for editing and designing your graphs, including pre-defined themes. See topics below for further details:

Editing

Almost every aspect of your graphs can be edited.

The graph option buttons on the Chart Toolbar are described in order from left to right below. A hint will pop up if you hover over a button.

Pointer - the standard cursor symbol.

Edit - 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.

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 between the default setting, coloured squares, and triangles and rotated squares.

Add grid - Add graph grid lines to the plot.

Add legend - Switch the chart legend on or off. N.B. This legend can be used to select groups for plotting.

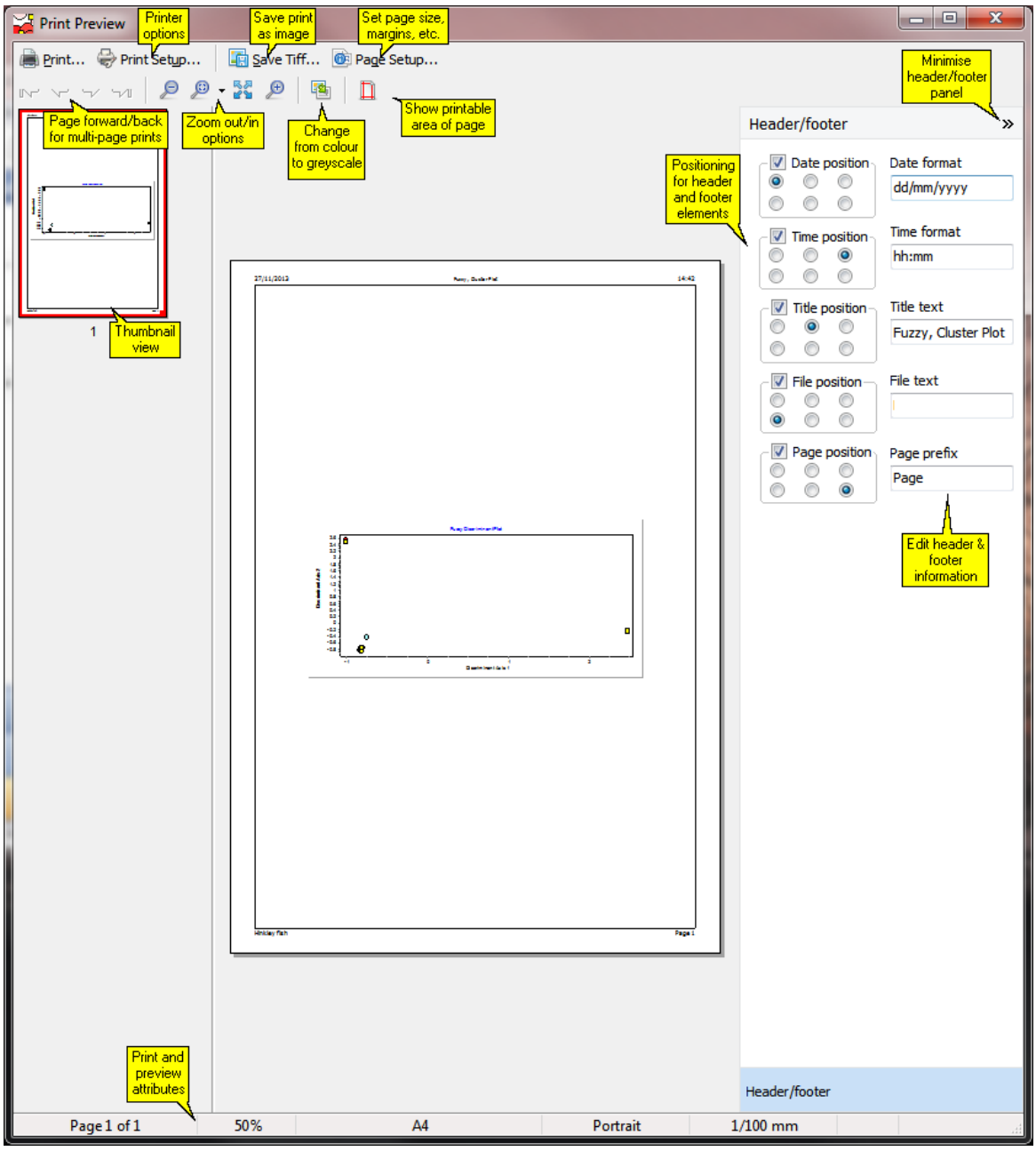
Add stalks - Add stalks to points - useful for 3D plots.

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

Printing

Charts can be printed by using the **Print** button on the graphics toolbar:

or by clicking **File: Print**. In both cases, this will bring up the Print Preview dialog, where you can alter many attributes of the printed page.



Thumbnail view

The Thumbnail view is useful when you have zoomed in to view a portion of the print; it shows which portion of the page is displayed. Click and drag on the red box to show a different portion of the page, or click/drag the square at the bottom right hand corner of the red box to zoom in or out further. (You can also click and drag the main preview page to move it around).

Colour/greyscale printing

When you switch from Colour to Greyscale view using the Colour/greyscale button, the panel at the bottom of the page allows you to alter the brightness and contrast of the greyscale print:

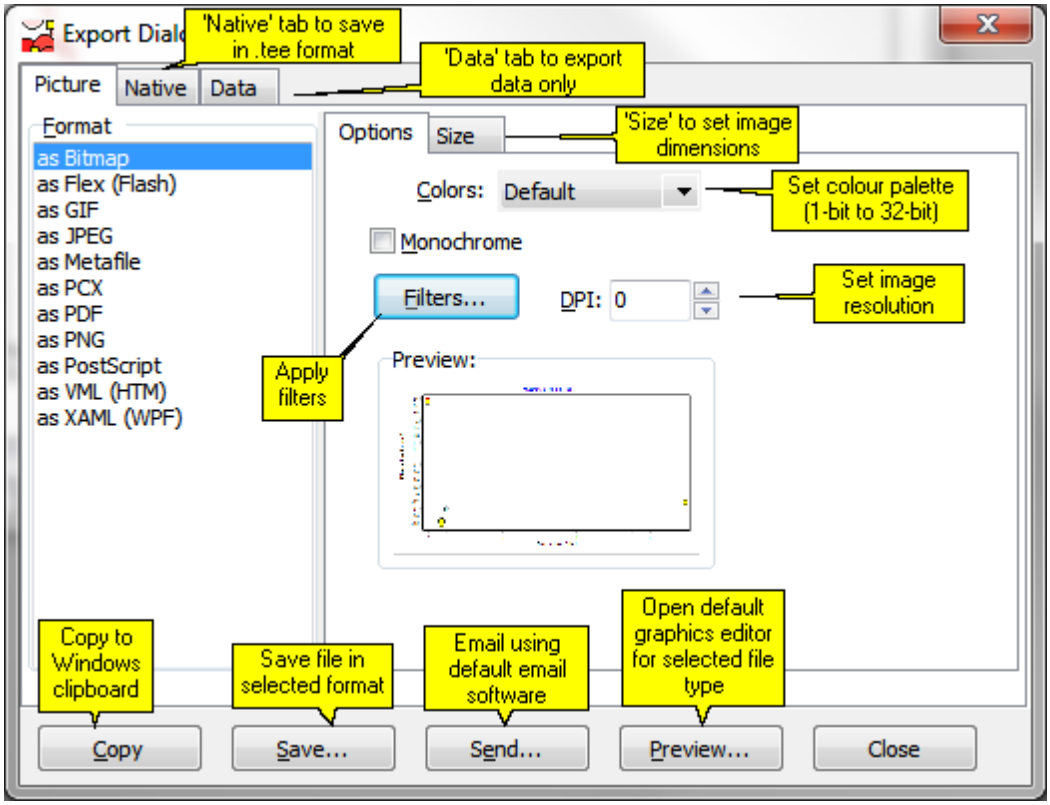
This panel disappears again when you switch back to a colour print.

Copying charts

Charts can also be copied to the Windows clipboard using **File: Export**, **Edit: Copy** or **Ctrl-C** on your keyboard, and then pasted in the normal Windows fashion into a Word document or other suitable document for subsequent printing.

Exporting charts

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

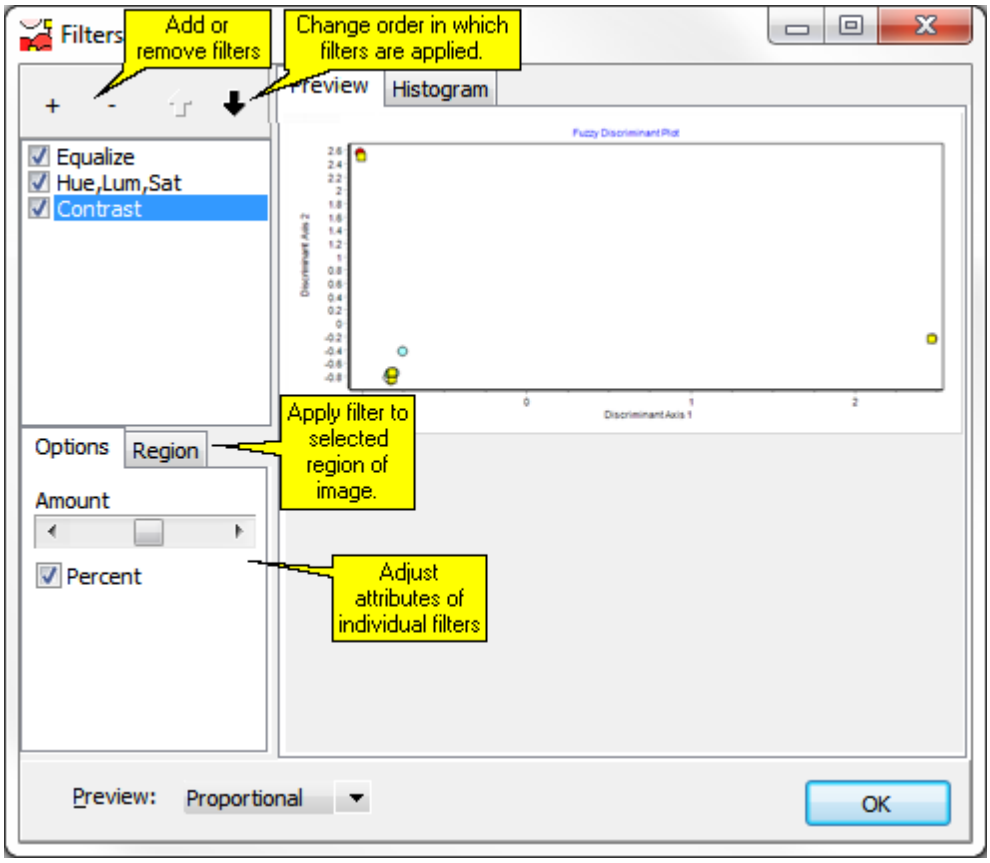


The chart can be saved in a wide range of different file formats, including Enhanced Metafile (*.emf), Bitmap (*.bmp), JPEG (*.jpg), PCX, PNG, GIF or Native (*.tee). Each file format has advantages and disadvantages. Enhanced Metafiles, when pasted into a Word document, can be resized by dragging, without losing resolution. Bitmaps are a lossless method of saving; the stored file will not lose any of the original's detail; however, the file size will be much larger than compressed files such as Enhanced Metafiles or JPEGs. JPEGs are file formats which can be compressed

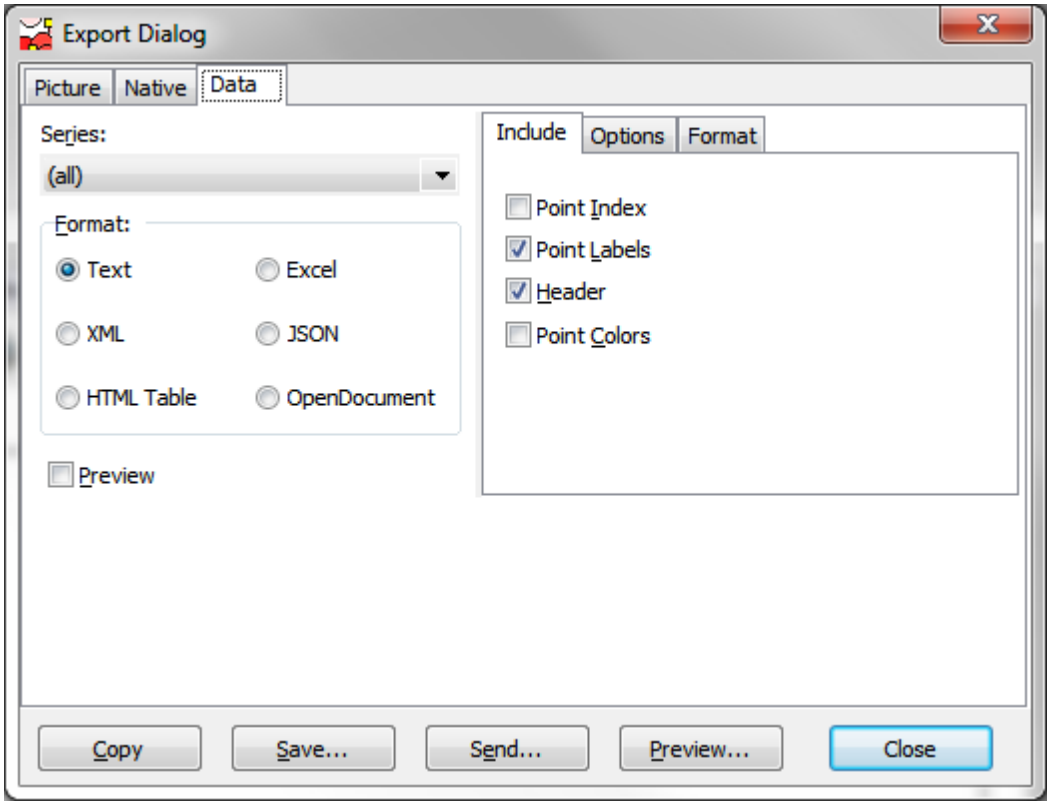
to take up less space - useful if you wish to send one by email, put it on a website, or paste it in to a document. If they are compressed too heavily, they can lose resolution and detail, and spoil colours. GIFs are also compressed files useful for web sites; they have a considerably smaller colour 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 supplied on the installation CD. Tee files tend to be very small indeed; often less than 1KB.

To save the chart as a PDF document, use the **File: Export** facility and select PDF as the required file type. Alternatively, 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 Adobe PDF from the list of available printers in the Print dialog box.

If you select the Bitmap (.bmp) format, you can apply a range of filters to the image:



This export dialog can also be used to export the data shown on the chart, using the 'Data' tab:



Part XII

12 Zooming and panning on graphs

Tight clusters of points which cannot be differentiated can sometimes occur. To zoom in on an area, move to the top right corner of the area to be enlarged then hold the left hand mouse button and drag to the lower left hand corner and release the button. An enlarged view of the selected area will be displayed. To return to the original view, hold down the left hand mouse button and move upwards and to the left and release. To pan the graph hold down the right hand mouse button and move the mouse.

Part XIII

13 Demonstration data sets

Fuzzy Grouping comes with several sets of demonstration data:

AllRiversBiological.csv paired with AllRiversMembership.csv – see Worked Example
irises.csv
telco data tran.csv

These files can be found in the Fuzzy Grouping folder in folder you installed Fuzzy Grouping to. The files allow the user to test the program and by opening them in Excel or another spreadsheet, see how the data is organised. The files can be deleted without affecting the program.

Part XIV

14 Common errors and problems

Common error messages and their solution are shown below

1. '' is not a valid floating point value. This will occur if the raw data holds blank columns or rows - ones that sum to zero. Remove blank columns and rows by using 0 Issues in the working data window. It may also occur if the raw data holds a blank cell. In some cases Fuzzy Grouping 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. 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.

Part XV

15 Fuzzy C-Means

15.1 Fuzzy C-Means

This method was developed by Dunn (1974) and Bezdek (1974, 1981 & 1987). The method is based on the minimization of the within-group sums of squares and is a generalisation of the hard c-means clustering algorithm. A single parameter, m, termed 'fuzziness' determines the degree of uncertainty about membership. If m = 1 the equation would produce a hard clustering and as the value of m increases above this value the more fuzzy the partitions produced.

In practice you cannot use a value of m = 1 as this would produce a division by zero error and it has been found that a value of about 2 is best.

The program works by iteratively searching for a solution working from an initial cluster membership array. These initial memberships can be either given by the user or generated by random by the program. There is a possibility with such iterative methods that the result will depend on the starting conditions. It is therefore advisable to run the analysis a number of times to ensure that the result is stable.

After opening your data and undertaking any transformations and adjustments in the working data panel, the Fuzzy c-means clustering is started by selecting Fuzzy C means from the dropdown menu at the top of the program window. You will be presented with two choices, Analysis for selected cluster number or Compare partition coefficients. The Fuzzy c-means algorithm works on a pre-selected number of clusters between which the data is to be divided. You should therefore select Analysis for selected cluster number if you know the number of clusters you would like the samples to be allocated to. If you are not sure which number of clusters is appropriate you can select Compare partition coefficients to see which cluster number gives the largest partition coefficient and then use this number of clusters in a subsequent run.

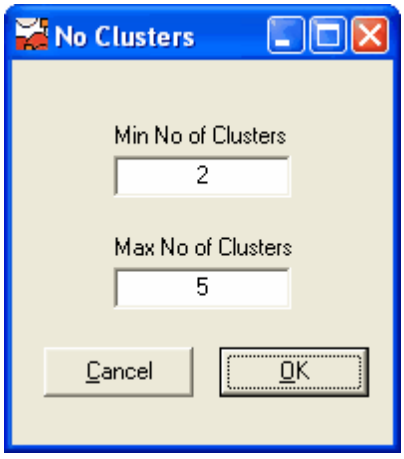
A. Analysis for selected cluster number:
C means options window
Presentation of results

B. Analysis to determine the cluster number:
Compare partition coefficients

15.2 Compare partition coefficients

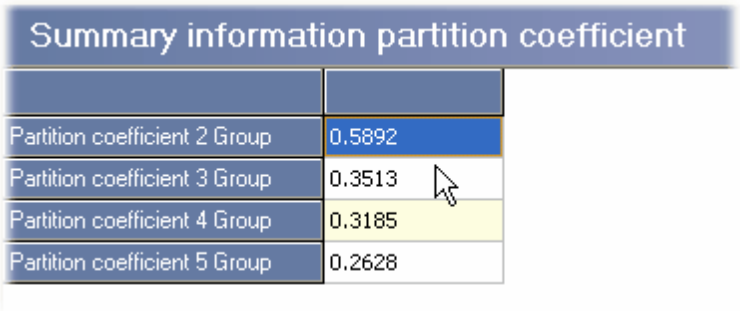
This option is used to examine how the partition coefficient changes with the number of clusters.

1. No. Clusters Window:



This window allows you to select the upper and lower values for the number of clusters that will be considered by the program. When you have chosen suitable values click OK.

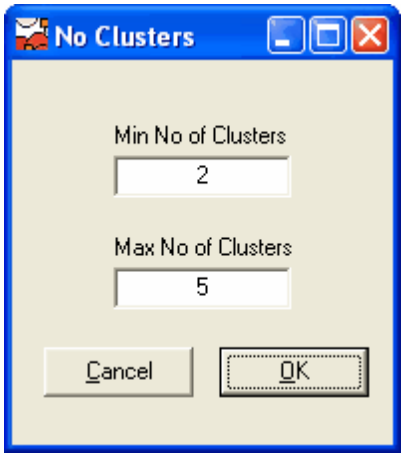
2. Presentation of results



The results of the analysis are given in tabulated form by clicking on the C-means summary tab. The closer the partition coefficient is to 1 the better the chosen group number is able to partition the data.

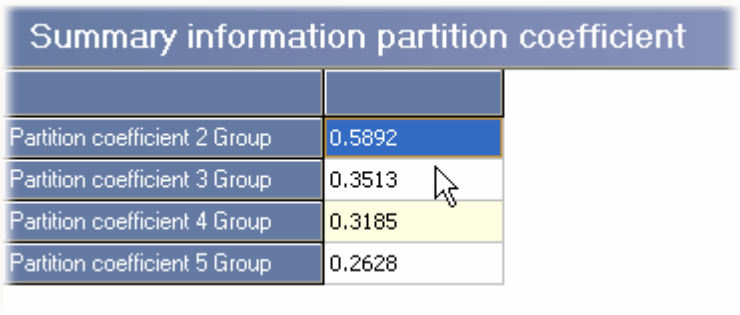
15.3 No. Clusters window

This window allows you to select the upper and lower values for the number of clusters that will be considered by the program. When you have chosen suitable values click OK.



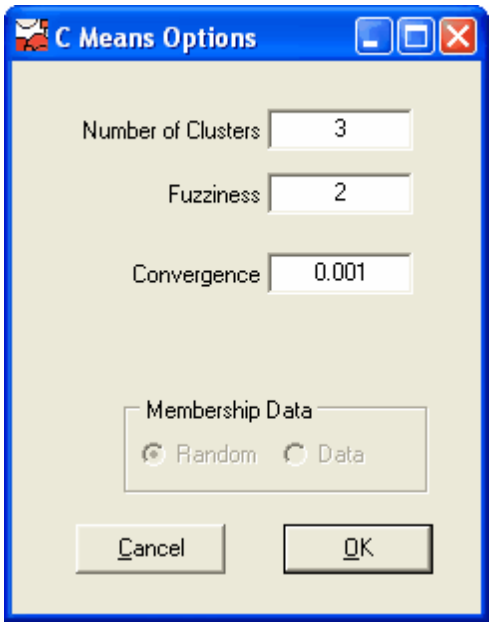
15.4 Summary Information Partition Coefficient

This is presented in tabulated form by clicking on the C-means summary tab.



15.5 C-Means options window

Before a c-means fuzzy cluster analysis is undertaken you are presented with a window that offers a series of options as follows.



Number of clusters. This is the only option that you are likely to need to change. Choose an integer > 1. A large number would not be sensible.

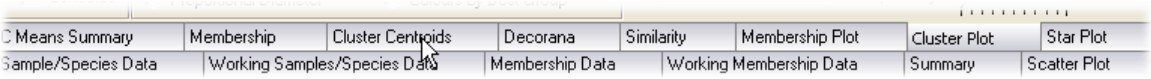
Fuzziness. This is the value of the fuzziness parameter, m. The larger the value the more fuzzy the membership. Choose a real value > 1. Generally, the default value of 2 has been found to be suitable for normal sized data sets. Large data sets may benefit from having Fuzziness reduced to roughly 1.5. It is unwise to set Fuzziness much above 2, since this can give rise to erroneous results, where every site/sample has a reasonably high probability of membership of every group.

Convergence. This is the convergence criterion for determining when the program should stop iterating towards a solution. The smaller this number, the larger the number of iterations and possibly the more accurate the eventual solution produced. In practice, there is little to be gained from making this number much smaller than the default.

When you have chosen your options click OK to continue.

15.6 Presentation of results

The results are presented over a number of tabbed sheets, each of which will be described in turn.



a) C-Means summary

This sheet gives the input options chosen and the resulting partition coefficient. The results can be selected and copied to the clipboard from where they can be pasted into other programs.

b) Membership

This sheet gives the probability of membership of each sample (or site or quadrate) to each cluster. The results can be selected and copied to the clipboard from where they can be pasted into other programs.

c) Cluster centroids

This sheet gives the position of the centre of gravity of each species within each cluster. These centroids are used in the MDS and then plotted. The results can be selected and copied to the clipboard from where they can be pasted into other programs.

d) Decorana

This sheet gives the position of each sample projected into a 2 dimensional space together with the position of the cluster centroids within this space. The results can be selected and copied to the clipboard from where they can be pasted into other programs.

f) Membership plot
This sheet plots the probability of membership of each site for each cluster.

g) Cluster plot
This sheet allows the resulting fuzzy clustering to be visualized in a variety in ways as a 2-dimensional plot. This is achieved by using DECORANA to produce an ordination of the sites within a 2-D space.

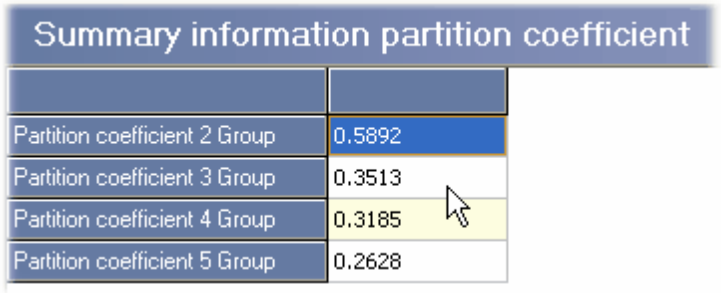
h) Star plot

15.7 C-means summary

This sheet gives the resulting partition coefficient for different numbers of clusters chosen; the closer the coefficient is to 1, the more optimal that number of clusters. In the output shown below the best group number is 2.

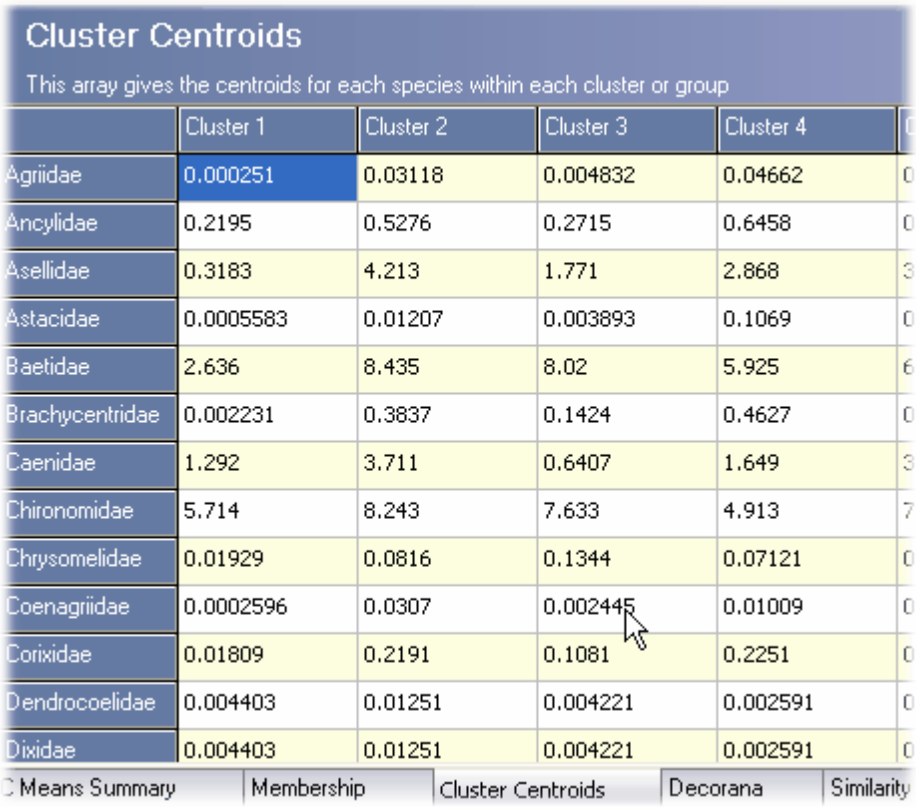
This output will only be visible if you have selected Fuzzy c-means : Compare partition coefficients has been selected.

The results can be selected and copied to the clipboard from where they can be pasted into other programs.



15.8 Cluster centroids

This sheet gives the position of the centre of gravity of each species within each cluster or group. These centroids are also plotted. The results can be selected and copied to the clipboard from where they can be pasted into other programs.



Cluster Centroids

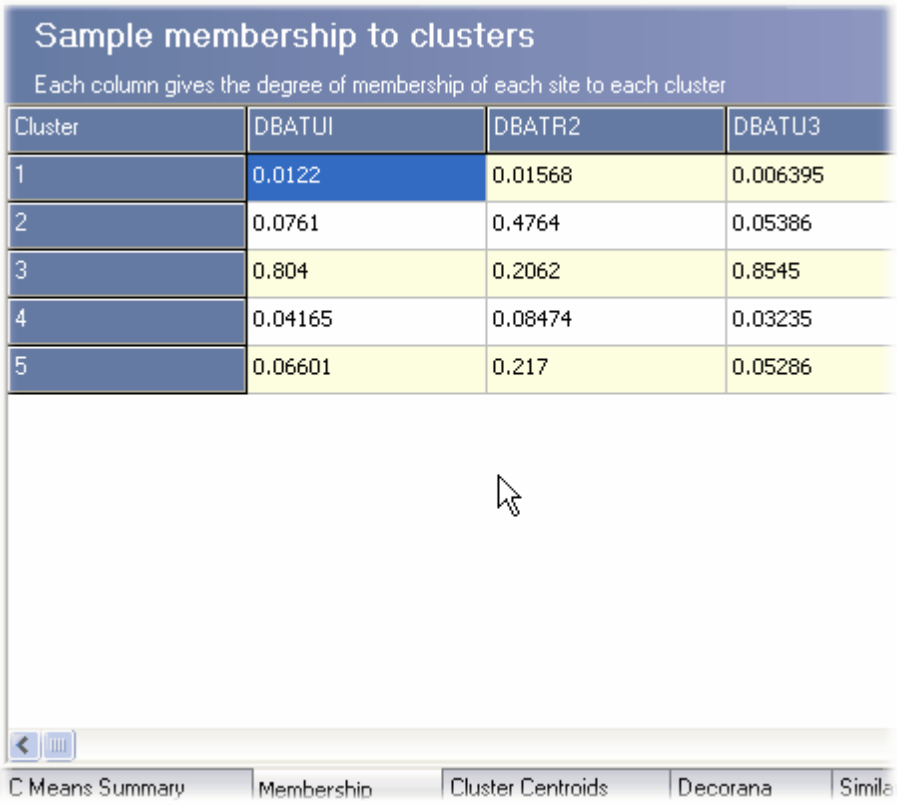
This array gives the centroids for each species within each cluster or group

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
Agriidae	0.000251	0.03118	0.004832	0.04662	0
Ancylidae	0.2195	0.5276	0.2715	0.6458	0
Asellidae	0.3183	4.213	1.771	2.868	3
Astacidae	0.0005583	0.01207	0.003893	0.1069	0
Baetidae	2.636	8.435	8.02	5.925	6
Brachycentridae	0.002231	0.3837	0.1424	0.4627	0
Caenidae	1.292	3.711	0.6407	1.649	3
Chironomidae	5.714	8.243	7.633	4.913	7
Chrysomelidae	0.01929	0.0816	0.1344	0.07121	0
Coenagriidae	0.0002596	0.0307	0.002445	0.01009	0
Corixidae	0.01809	0.2191	0.1081	0.2251	0
Dendrocoelidae	0.004403	0.01251	0.004221	0.002591	0
Dixidae	0.004403	0.01251	0.004221	0.002591	0

Means Summary | Membership | Cluster Centroids | Decorana | Similarity

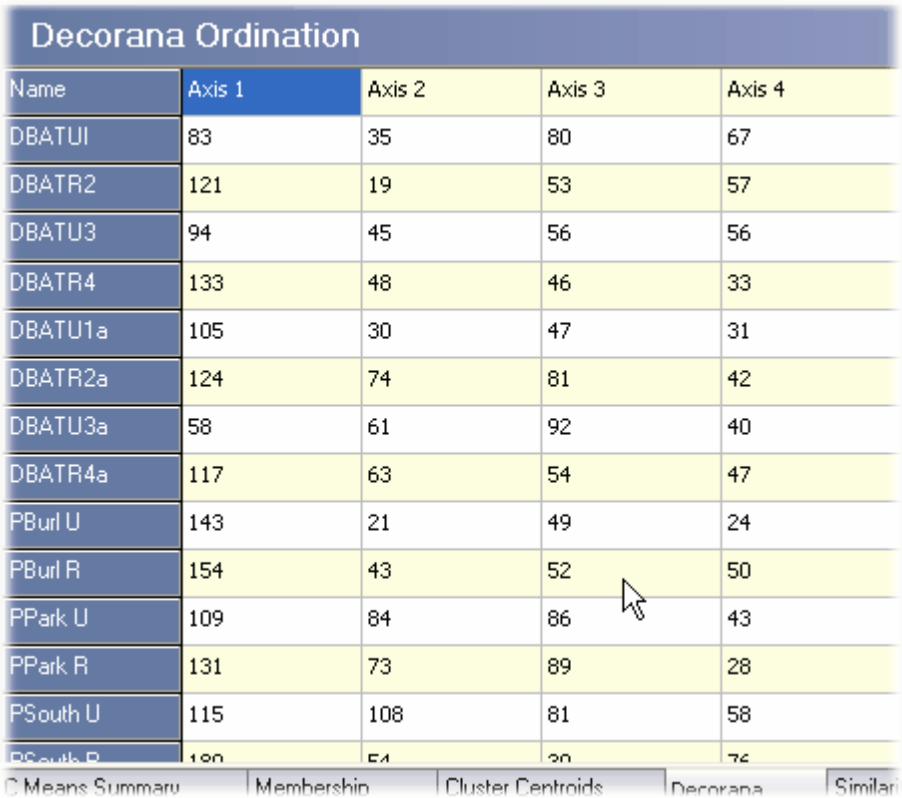
15.9 Membership

This sheet gives the probability of membership of each sample (or site or quadrate) to each cluster or group. The results can be selected and copied to the clipboard from where they can be pasted into other programs.



15.10 Decorana

This grid presents the co-ordinates of the samples (sites or quadrates) in the ordination space calculated by DECORANA.



Decorana Ordination

Name	Axis 1	Axis 2	Axis 3	Axis 4
DBATUI	83	35	80	67
DBATR2	121	19	53	57
DBATU3	94	45	56	56
DBATR4	133	48	46	33
DBATU1a	105	30	47	31
DBATR2a	124	74	81	42
DBATU3a	58	61	92	40
DBATR4a	117	63	54	47
PBurl U	143	21	49	24
PBurl R	154	43	52	50
PPark U	109	84	86	43
PPark R	131	73	89	28
PSouth U	115	108	81	58

Means Summary | Membership | Cluster Centroids | Decorana | Similar

The grid presents the final coordinates for each site (column). The site identifier is given in the first column, in the second and third columns are listed the coordinates of each site for the two dimensional solution.

The results can be selected and copied to the clipboard from where they can be pasted into other programs.

15.11 Cluster plot

This sheet allows the resulting fuzzy clustering to be visualized in a variety in ways as a 2-dimensional plot. This is achieved by using DECORANA for a standard fuzzy c-means analysis or using discriminant space if fuzzy discriminant analysis is selected to produce an ordination of the sites within a 2-D space. The various types of plot are selected using the radio button panel labeled Plot.

The number of colours on Cluster plots is limited to 14, to prevent the graphical output from being visually too confusing. If more than 14 clusters are specified to assign samples to, some of these when plotted will be displayed in the same colour if the "Different Colours (nearest)" or "Colours by Best Group" plots are displayed.

The full range of clusters can still be distinguished by using "Proportional Diameter" or "Colour Intensity" to examine membership of one cluster at a time.

The plot options are:

Different colours (nearest): using different coloured circles to the cluster centroid the sample is closest to in space.

Proportional diameter: in which the diameter of Circle X represents the likelihood of its belonging to Cluster Y.

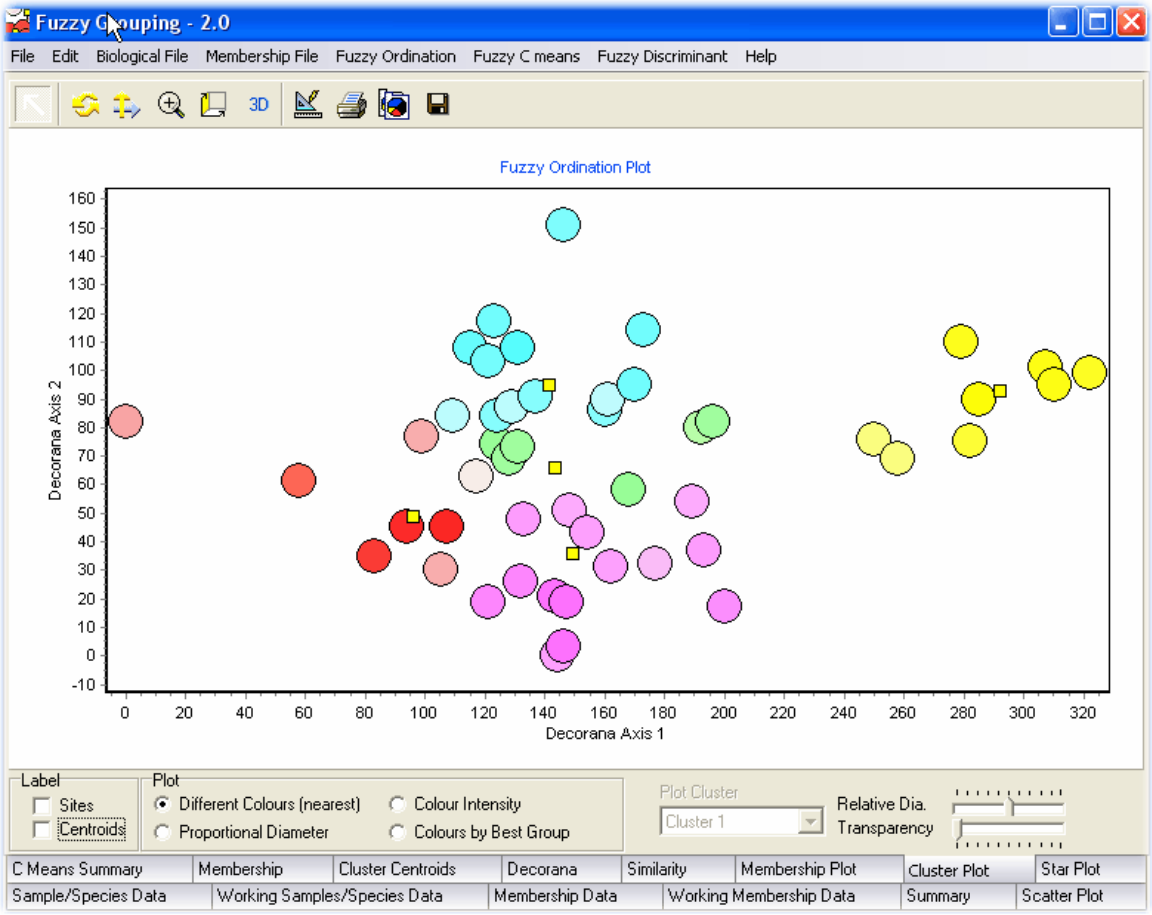
Intensity: where the intensity of the shading of Circle X represents the likelihood of its belonging to Cluster Y.

Colours by Best group: using different coloured circles to represent the most likely membership using the degree of membership .

There are two slider controls to adjust your plot

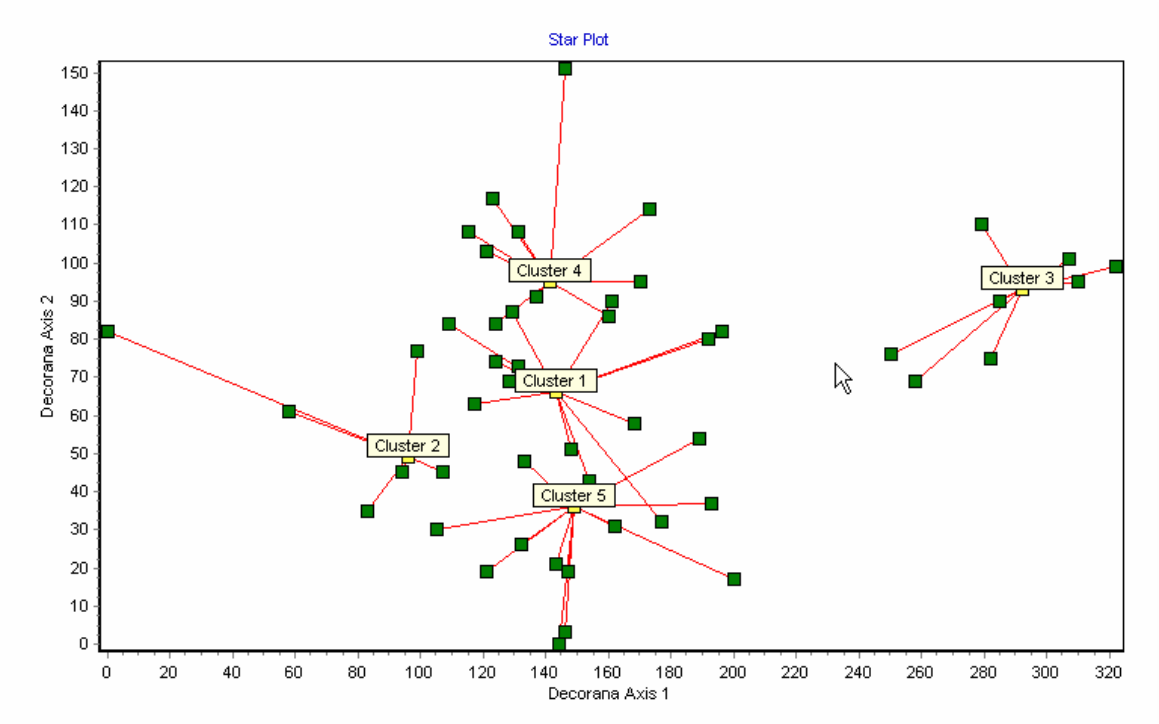
Relative Dia. Changes the diameter of the circles

Transparency Adjusts circle transparency so that you can see underlying samples.



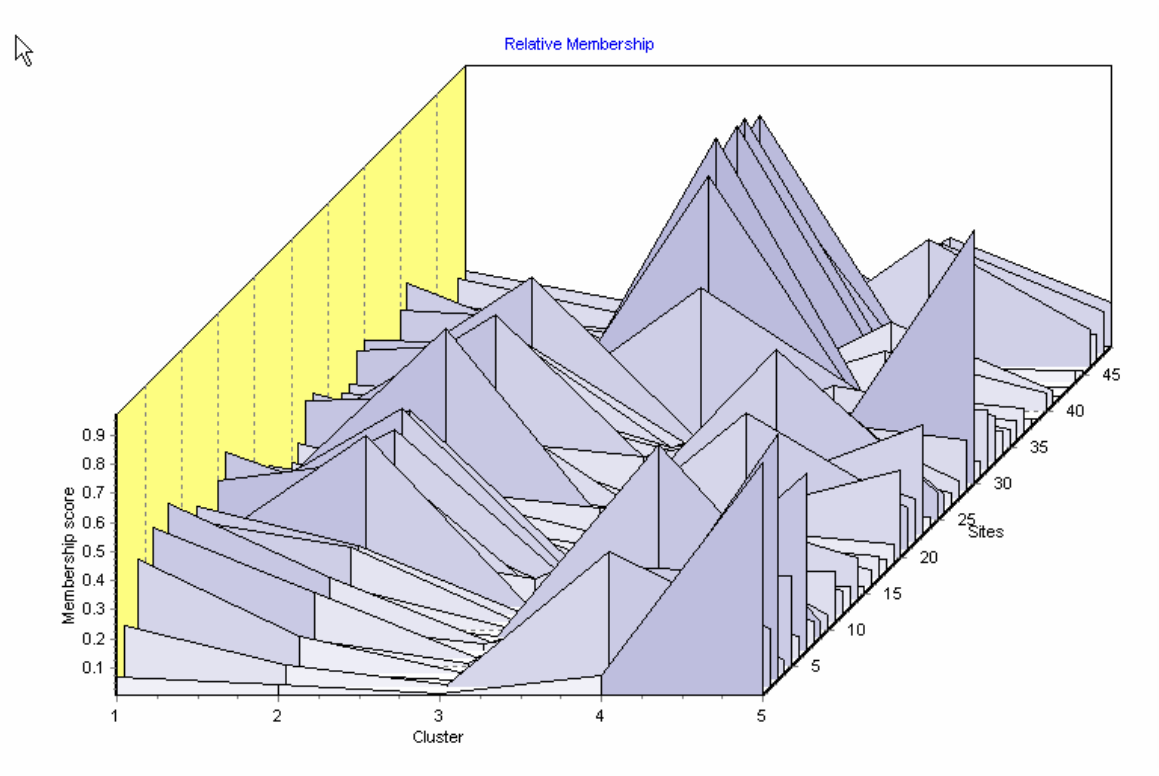
15.12 Star plot

Shows each site/species linked by a straight line to its closest centroid. The sample coordinates have been calculated using DECORANA for a standard fuzzy c-means analysis or using discriminant space if fuzzy discriminant analysis is selected.



15.13 Membership plot

This sheet plots the probability of membership of each site for each cluster as a 3 dimensional graph.



15.13.1 DECORANA

15.13.1.1 DECORANA

Fuzzy c-means does not output co-ordinates for each sample that can be used to produce a plot in which similar samples are grouped together.

Fuzzy Grouping 2 produces these co-ordinates using Detrended Correspondence analysis (DECORANA).

Detrended correspondence analysis was devised by Hill (1979) as an attempt to improve upon Reciprocal Averaging (RA). Two problems that occur with reciprocal averaging are termed the 'arch effect' and 'end point compression'. When the first and second axes produced by RA are plotted it is often observed that the points are arranged in an arch, because of the quadratic relationship between the axes, rather than any ecological relationship. DECORANA removes this arch by a technique termed detrending. The tendency for points at each end of the first axis to be closer together than those in the middle is removed by segmenting the axis and expanding the terminal segments and compressing those towards the centre. Whereas RA scales the axes between 0 and 100 in relation to the magnitude of the eigenvalue, DECORANA scales in units of average standard deviation of species turnover. Therefore a change of 50% in species composition occurs in about 1 standard deviation.

15.13.1.2 Reciprocal Averaging (RA)

This method, also called Correspondence Analysis, is a method of showing the relationship between both species and samples (quadrates) in a reduced space. Originally proposed by Hirschfeld (1935) and Fischer (1940) it was first used by ecologists in the 1960s (Roux & Roux, 1967; Benzécri, 1967) - see Kent & Coker (1992) for more details. The method is described by (Hill 1973) and a non-mathematical introduction to the technique is given in Kent & Coker (1992). RA uses chi-squared distance values and this results in low abundance species having a possibly disproportionately large effect on the ordination produced, and can over emphasise the difference in samples containing several infrequently recorded species. RA performs best for analysing samples that were collected along an environmental gradient. If there are no clear environmental gradients in the habitat under study, or the gradients are short, then PCA may give better results. RA can be applied to both presence/absence and quantitative data.

Part XVI

16 Fuzzy Ordination Analysis

16.1 Fuzzy Ordination Analysis

This method is a type of Fuzzy Ordination Analysis and was introduced by Roberts (1986) as an alternative to traditional ordination techniques. Unlike these latter techniques, an investigator using fuzzy set ordination must give a relationship between the environment and the biological community before performing the ordination. Therefore unlike the fuzzy c-means method you will need to supply a vector of membership (environmental) data for each site as well as the array of biological data. Although the environmental data set can contain more than one membership or environmental variable, only one is used at any one time.

The relationship between the biological community and the environmental variable may be based on actual physical readings or might be hypothetical. Sites could, for example, be assigned values that represent their relative position along a gradient. For example, if a transect up a mountain side is sampled, then each quadrate could be defined by its altitude above sea level. Alternatively, sites could simply be given a value between 0 and 1 that denoted their membership of the set. For example, all high altitude sites might be given the value 1 and low altitude sites the value 0. Sites of intermediate altitude would be given a value between 0 and 1.

The method used here is based on the algorithm given by Rick Boyce on the Fuzzy Set Ordination web site - http://www.nku.edu/~boycer/fso/

1. Demonstration data

The program is supplied with demonstration data set to show you how to organize your data to undertake a Fuzzy Ordination Analysis. The species data to be opened as the Biological File from the drop-down menu is called AllRiversBiological.csv. The membership data file giving membership and physical data on the sites is opened by selecting Membership File from the drop-down menu and selecting AllRiversMembership.csv. These data included membership of restored or unrestored stretches of river, membership of individual river groups, and physical data such as depth, width, canopy cover and % cover by Ranunculus spp.

2. Running a fuzzy ordination

Once the data has been loaded, choose Fuzzy Ordination; the Similarity menu box will appear. You will be presented with an options menu from which you can select the similarity measure and whether or not to apply step across. Although the environmental data set can contain more than one membership or environmental variable, only one is used at any one time.

Select the variable you require from the drop-down menu at the top of the dialog box. Select an appropriate similarity measure by clicking on the radio button. A large number of similarity measures are offered because there is as yet no clear consensus as to the best one to use in all circumstances. Some of the similarity measures are suitable for binary (presence/absence data represented by 0 and 1) while others are suitable for quantitative data when the abundance of each species has been recorded. The various measures are discussed in Similarity measures.

If desired, select Step across by clicking the tick box. Step across is a technique that reduces the tendency of some data to produce a "horseshoe" effect when results are plotted because of the problems of comparing the similarity of sites at the ends of the range. This option is generally to be preferred, however, it can extend considerably the time taken to analyze a large data set.

The results of the ordination are displayed on a number of tabbed pages, described below.

a) Similarity

This simply presents a table of the similarity between samples calculated using the index selected from the Similarity menu.

b) Fuzzy Ordination

This presents the result of the ordination. The first row is the predefined membership or actual value for the membership/environmental variable. The second row is this value of this variable relativised to range between 0 and 1. The third row of the grid gives the apparent ordination based on this species assemblage in each sample.

c) Ordination Plot

This plots the membership as given by the chosen environmental variable (x axis) against the apparent membership produced by the ordination (y axis). If the ordination produces the same result as the predefined membership, the plot will be a diagonal straight line. The plot can be used to identify samples that differ greatly between their actual and apparent membership.

Ordination Plot

16.2 Similarity

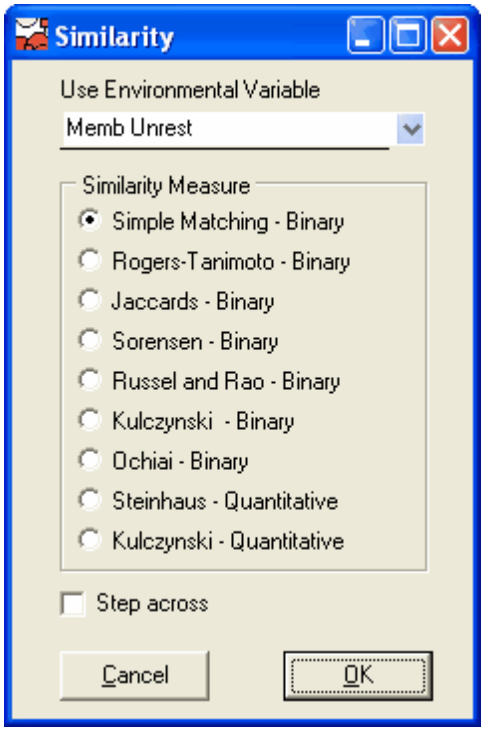
Gives the similarity between sites. This grid is only shown if a Fuzzy ordination has been undertaken.

16.3 Similarity choices

This window allows the choice of the both the environmental variable to be used for the ordination and the similarity measure to be used by the program.

Choose the environmental or membership variable from the dropdown menu at the top of the window.

The similarity measure is chosen by clicking on a radio button.



Additional information about similarity measures

16.4 Actual and apparent scores

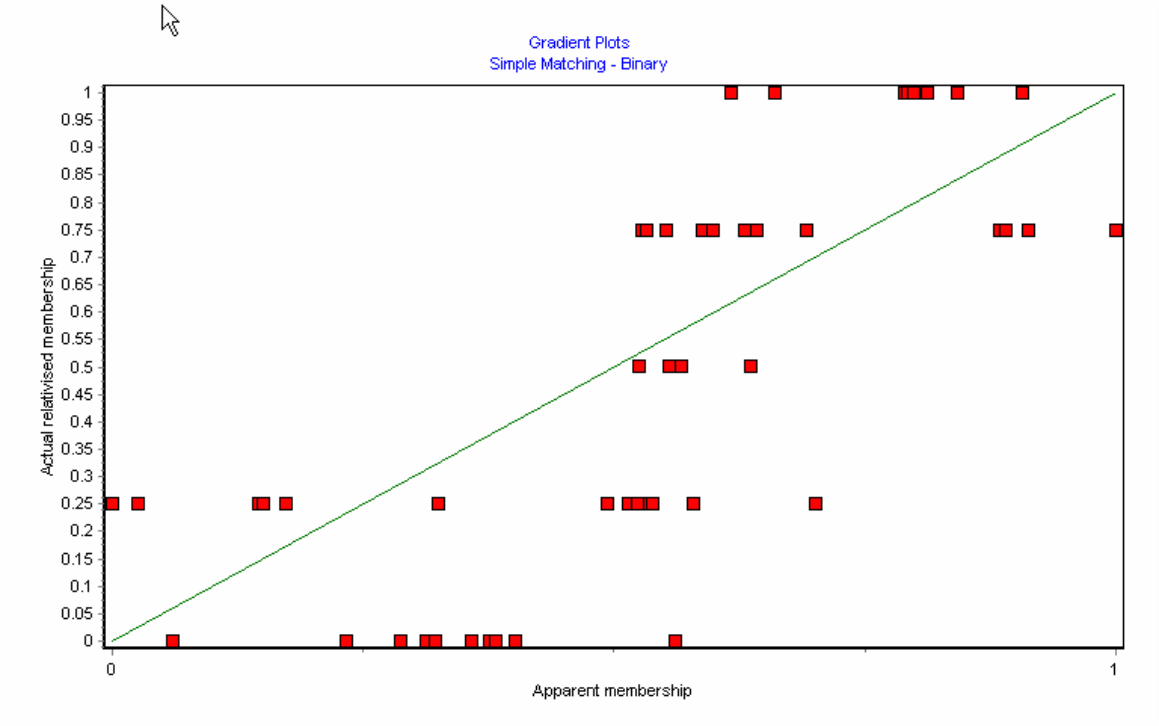
Fuzzy ordination produces a comparison between the actual score for an environmental variable or group membership and the apparent score produced by the fuzzy ordination. If the actual score is a very good predictor of the sample compositions, then these two scores should show a high correspondence. This tabbed grid shows both the actual and apparent scores.

Site	DBATUl	DBATR2	DBATU3	DBATR4	DBATU1a	DBATR2a	DBATU3a	DBATR4a	PBurl U	PBurl R	PPark U	PPar
Actual	1	0	1	0	1	0	1	0	1	0	1	0
Actual (Relative)	1	0	1	0	1	0	1	0	1	0	1	0
Apparent	0.3797	0.6051	0.4121	0.6113	0.594	0.6935	0.4119	0.6535	0.2202	0.4824	0.1684	0.565

16.5 Ordination Plot

This output is only produced by Fuzzy Ordination.

It is a plot of the membership as given by the chosen environmental variable (y axis) against the apparent membership produced by the ordination (x axis). If the ordination produces the same result as the predefined membership, the plot will be a diagonal straight line. The plot can be used to identify samples that differ greatly between their actual and apparent membership.



16.6 Similarity measures

16.6.1 Similarity measures

These are simple measures of either the extent to which two habitats have species in common (Q analysis) or species have habitats in common (R analysis). Binary similarity coefficients use presence-absence data, following the introduction of computers more complex quantitative coefficients became practicable. Both groups of indices can be further divided to those which take account of the absence from both communities (double zero methods) and those which do not. In most ecological applications it is unwise to use double-zero methods as they assign a high level of similarity to localities which both lack many species; a problem which becomes particularly acute in habitats which have a potentially extremely large species list, such as the marine benthos. A good account of similarity and distance measures is given in Legendre & Legendre (1983). Because of division by zero problems, for some data sets not all measures can be calculated. When a division by zero error would occur Fuzzy Grouping gives an index of -99.

For measures of similarity between samples based on species presence-absence the observations can be summarised in a simple frequency table:

		Sample 1 Species Present	Species Absent
Sample 2	Species Present	a	c
	Species Absent	b	d

where the number of species present in both samples is a, the number of species present in sample 1 but missing from sample 2 is b, the number of species missing in sample 1 but present in sample 2 is c and the number of species missing from both samples is d. The total number of species, N, is therefore a+b+c+d.

Binary - double zeros
Simple matching
Rogers_Tanimoto

Binary - no double zeros
Jaccards
Sorensen
Russel & Rao
Kulczynski

Ochiai

Quantitative measures
Steinhaus
Kulczynski-Quantitative

16.6.2 Sorensen

For measures of similarity between samples based on species presence-absence the observations can be summarised in a simple frequency table:

		Sample 1 Species Present	Species Absent
Sample 2	Species Present	a	c
	Species Absent	b	d

where the number of species present in both samples is a, the number of species present in sample 1 but missing from sample 2 is b, the number of species missing in sample 1 but present in sample 2 is c and the number of species missing from both samples is d. The total number of species, N, is therefore a+b+c+d.

This measure is:

$$\frac{2a}{2a + b + c}$$

16.6.3 Jaccards

For measures of similarity between samples based on species presence-absence the observations can be summarised in a simple frequency table:

		Sample 1 Species Present	Species Absent
Sample 2	Species Present	a	c
	Species Absent	b	d

where the number of species present in both samples is a, the number of species present in sample 1 but missing from sample 2 is b, the number of species missing in sample 1 but present in sample 2 is c and the number of species missing from both samples is d. The total number of species, N, is therefore a+b+c+d.

This measure is:

$$\frac{a}{a + b + c}$$

16.6.4 Rogers_Tanimoto

For measures of similarity between samples based on species presence-absence the observations can be summarised in a simple frequency table:

		Sample 1 Species Present	Species Absent
Sample 2	Species Present	a	c
	Species Absent	b	d

where the number of species present in both samples is a, the number of species

present in sample 1 but missing from sample 2 is b, the number of species missing in sample 1 but present in sample 2 is c and the number of species missing from both samples is d. The total number of species, N, is therefore a+b+c+d.

This measure is:

$$\frac{a+d}{a+2b+2c+d}$$

16.6.5 Russel & Rao

For measures of similarity between samples based on species presence-absence the observations can be summarised in a simple frequency table:

		Sample 1 Species Present	Species Absent
Sample 2	Species Present	a	c
	Species Absent	b	d

where the number of species present in both samples is a, the number of species present in sample 1 but missing from sample 2 is b, the number of species missing in sample 1 but present in sample 2 is c and the number of species missing from both samples is d. The total number of species, N, is therefore a+b+c+d.

This measure is:

$$\frac{a}{N}$$

16.6.6 Kulczynski

For measures of similarity between samples based on species presence-absence the observations can be summarised in a simple frequency table:

		Sample 1 Species Present	Species Absent
Sample 2	Species Present	a	c
	Species Absent	b	d

where the number of species present in both samples is a, the number of species present in sample 1 but missing from sample 2 is b, the number of species missing in sample 1 but present in sample 2 is c and the number of species missing from both samples is d. The total number of species, N, is therefore a+b+c+d.

This measure is:

$$\frac{b}{b+c}$$

16.6.7 Ochiai

For measures of similarity between samples based on species presence-absence the observations can be summarised in a simple frequency table:

		Sample 1	
		Species Present	Species Absent
Sample 2	Species Present	a	c
	Species Absent	b	d

where the number of species present in both samples is a, the number of species present in sample 1 but missing from sample 2 is b, the number of species missing in sample 1 but present in sample 2 is c and the number of species missing from both samples is d. The total number of species, N, is therefore a+b+c+d.

This measure is:

$$\frac{a}{\sqrt{(a+b)(a+c)}}$$

16.6.8 Steinhaus

If W is the sum of the minimum abundance for each species in two samples, A and B are the sum of the abundance of all the species in each sample this similarity coefficient is given by:

$$\frac{W}{A+B}$$

16.6.9 Czekanowski distance

There is some confusion in the literature as to the correct name for this distance measure. It is attributed to Steinhaus by Motyka (See Legendre & Legendre, 1983).

For two samples the distance is given by:

2w/A+B

where w is the sum of the minimum abundances of the species in the two samples,

A is the sum of species abundance in sample 1 and B is the sum of species abundance in sample 2.

16.6.10 Kulczynski-Quantitative

Using the same nomenclature as for the Steinhaus coefficient this measure is given by:

$$\frac{1}{2}\left(\frac{W}{A}+\frac{W}{B}\right)$$

16.6.11 Simple matching

For measures of similarity between samples based on species presence-absence the observations can be summarised in a simple frequency table:

		Sample 1	
		Species Present	Species Absent
Sample 2	Species Present	a	c
	Species Absent	b	d

where the number of species present in both samples is a, the number of species present in sample 1 but missing from sample 2 is b, the number of species missing in sample 1 but present in sample 2 is c and the number of species missing from both samples is d. The total number of species, N, is therefore a+b+c+d.

Simple matching is (a+d) / N. Note that this measure gives equal weighting to double zeros and species which are present in both samples. This is rarely useful in ecological studies.

Part XVI

17 Fuzzy Discriminant Analysis

17.1 Fuzzy Discriminant Analysis

Linear discriminant analysis (Fisher, 1936) is used to investigate the differences between groups or classes of objects defined by a number of variables such as species.

There are several reasons for undertaking a discriminant analysis including

· To classify cases into groups using a discriminant prediction equation.

· To investigate independent variable mean differences between groups formed by the dependent variable.

· To determine the percent of variance in the dependent variable explained by the independents.

· To determine the percent of variance in the dependent variable explained by the independents over and above the variance accounted for by control variables, using sequential discriminant analysis.

· To assess the relative importance of the independent variables in classifying the dependent variable.

· To discard variables which are little related to group distinctions.

· To test theory by observing whether cases are classified as predicted.

However, the ones which are most important when Fuzzy Discriminant Analysis is undertaken are;

· the projection of the samples into a space where they can be plotted to best illustrate the differences between the groups and

· the use of the fuzzy Wilks' Lambda to identify the most appropriate number of groups.

In Fuzzy Discriminant Analysis the following calculations are undertaken.

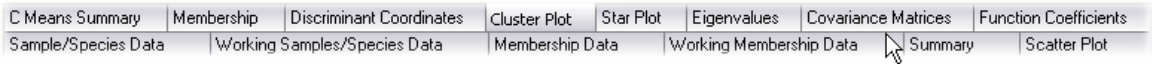
1. The fuzzy membership of each sample is calculated for the selected number of groups using the fuzzy c-means algorithm.
2. The sums of squares and products (SSP) within-classes matrix W (also called within-classes fuzzy scatter matrix) is calculated.
3. The fuzzy sum of squares and products (SSP) between-classes matrix Bf is calculated.
4. The total SSP matrix Tf is calculated as Tf = Bf + Wf .
5. The ratio of the determinant of the within-classes to the total SSP matrix is the fuzzy Wilks' Lambda. The fuzzy Wilks' is a measure of the difference between classes.
6. The projection of a data vector x on the i th canonical axis is computed.
7. Similarly, the class means centroids of the j th class are calculated.

C means options window

Fuzzy discriminant analysis results

17.2 Fuzzy Discriminant Analysis Results

The results are presented over a number of tabbed sheets, each of which will be described in turn.



a) C-Means summary

This sheet gives the input options chosen and the resulting partition coefficient. The results can be selected and copied to the clipboard from where they can be pasted into other programs.

b) Membership

This sheet gives the probability of membership of each sample (or site or quadrate) to each cluster. The results can be selected and copied to the clipboard from where they can be pasted into other programs.

c) Discriminant Coordinates

d) Cluster plot

This sheet allows the resulting fuzzy clustering to be visualized in a variety in ways as a 2-dimensional plot.

e) Star plot

This sheet shows the relationship of the individual samples to their group centroid.

f) Eigenvalues

g) Covariance Matrices

h) Function Coefficients

17.3 Discriminant Coordinates

This grid gives the coordinates of each of the samples within the discriminant axes and is a tabulation of the data plotted in the cluster plot.

The actual number of axes is equal to the number of unique eigenvalues.

Discriminant Coordinates

	Root 1	Root 2	Root 3	Root 4
BATUI	46.80	-0.50	-1.98	0.69
BATR2	46.80	2.22	0.18	-0.23
BATU3	46.78	-0.73	-2.09	0.73
BATR4	46.78	3.04	-0.05	1.61
BATU1a	46.81	1.75	-0.49	0.37
BATR2a	46.77	3.04	-1.14	2.69
BATU3a	46.87	0.37	-1.72	0.68
BATR4a	46.77	3.05	-1.25	3.13

17.4 Eigenvalues

This grid gives the Eigenvalues and fuzzy Wilks' Lambda for the fuzzy discriminant analysis. The eigenvalues are equal to the value of the objective function for each of the discriminant functions. By examining the relative magnitudes of the Eigenvalues you can decide which discriminant function does the best job at discriminating between the groups. This grid also presents significance tests for the Eigenvalues using the standard method used for non-fuzzy discriminant analysis. These should be used with caution as we are unsure if they can be relied upon with a fuzzy analysis.

Fuzzy Grouping - 2.0

File Edit Biological File Membership File Fuzzy Ordination Fuzzy C means Fuzzy Discriminant Help

Eigenvalues and their significance

Number of roots extracted	1.00					
Percent of trace extracted	100.00					
Wilks Lambda	0.69					
Roots of the W inverse time B Matrix						
No.	Root	Proportion	Canonical R	Chi-Squared	D.F.	Prob.
Root 1	0.46	1.00	0.56	376.27	5.00	0.00

The word "Error" will appear under Wilks' Lambda if there is a very determinant for one of the sums of squares and cross products matrices.

17.5 Covariance Matrices

This grid presents the values for

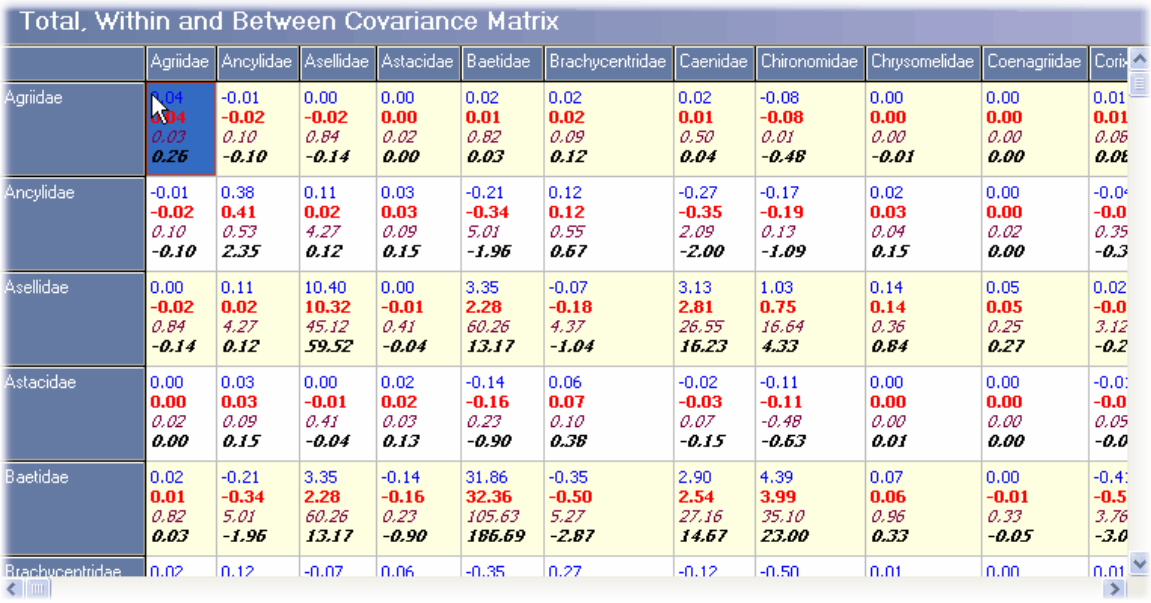
1. The total fuzzy sums of squares and cross products matrix in blue

2. The fuzzy pooled within group sums of squares and cross products matrix in red bold

3. The fuzzy between group sums of squares and cross products matrix in purple italic

4. The total fuzzy variance covariance matrix in black italic bold.

(This information is also displayed at the bottom of the grid).



	Agriidae	Ancylidae	Asellidae	Astacidae	Baetidae	Brachycentridae	Caenidae	Chironomidae	Chrysomelidae	Coenagriidae	Cori
Agriidae	0.04 0.04 0.03 0.25	-0.01 -0.02 0.10 -0.10	0.00 -0.02 0.84 -0.14	0.00 0.00 0.02 0.00	0.02 0.01 0.82 0.03	0.02 0.02 0.09 0.12	0.02 0.01 0.50 0.04	-0.08 -0.08 0.01 -0.48	0.00 0.00 0.00 -0.01	0.00 0.00 0.00 0.00	0.01 0.01 0.06 0.0
Ancylidae	-0.01 -0.02 0.10 -0.10	0.38 0.41 0.53 2.35	0.11 0.02 4.27 0.12	0.03 0.03 0.09 0.15	-0.21 -0.34 5.01 -1.96	0.12 0.12 0.55 0.67	-0.27 -0.35 2.09 -2.00	-0.17 -0.19 0.13 -1.09	0.02 0.03 0.04 0.15	0.00 0.00 0.02 0.00	-0.0 -0.0 0.35 -0.3
Asellidae	0.00 -0.02 0.84 -0.14	0.11 0.02 4.27 0.12	10.40 10.32 45.12 59.52	0.00 -0.01 0.41 -0.04	3.35 2.28 60.26 13.17	-0.07 -0.18 4.37 -1.04	3.13 2.81 26.55 16.23	1.03 0.75 16.64 4.33	0.14 0.14 0.36 0.84	0.05 0.05 0.25 0.27	0.02 -0.0 3.12 -0.2
Astacidae	0.00 0.00 0.02 0.00	0.03 0.03 0.09 0.15	0.00 -0.01 0.41 -0.04	0.02 0.02 0.03 0.13	-0.14 -0.16 0.23 -0.90	0.06 0.07 0.10 0.38	-0.02 -0.03 0.07 -0.15	-0.11 -0.11 -0.48 -0.63	0.00 0.00 0.00 0.01	0.00 0.00 0.00 0.00	-0.0 -0.0 0.05 -0.0
Baetidae	0.02 0.01 0.82 0.03	-0.21 -0.34 5.01 -1.96	3.35 2.28 60.26 13.17	-0.14 -0.16 0.23 -0.90	31.86 32.36 105.63 186.69	-0.35 -0.50 5.27 -2.87	2.90 2.54 27.16 14.67	4.39 3.99 35.10 23.00	0.07 0.06 0.96 0.33	0.00 -0.01 0.33 -0.05	-0.4 -0.5 3.76 -3.0
Brachycentridae	0.02	0.12	-0.07	0.06	-0.35	0.27	-0.12	-0.50	0.01	0.00	0.01

17.6 Discriminant Function Coefficients

This grid presents the discriminant function coefficients for each discriminant function. It is possible to identify the variables (e.g. species or higher taxa) which are the most important for discriminating between the groups by examining the relative size of their coefficients. For example, in the grid below the first discriminant function mostly reflects the abundance of Hydrobiidae.

Canonical Discriminant Function Coefficients

	Root 1	Root 2	Root 3	Root 4
Corixidae	-1.23	-0.13	0.32	0.89
Dendrocoelidae	-0.31	0.09	0.09	0.26
Dixidae	-0.31	0.09	0.09	0.26
Dytiscidae	1.01	1.39	0.42	0.26
Elminthidae	-0.83	-0.99	-0.07	-0.76
Empididae	-0.06	0.14	0.14	0.43
Ephemerellidae	-6.88	-2.76	2.42	2.03
Ephemeridae	-0.50	-0.36	-0.21	0.24
Erpobdellidae	-0.85	0.37	0.41	-0.45
Gammaridae	-10.39	-1.83	4.09	3.05
Glossiphoniidae	0.69	-0.26	0.20	0.41
Goeridae	-0.76	-0.49	-0.06	0.18
Gyrinidae	-0.19	-0.35	0.01	0.05
Haliplidae	-0.19	-0.35	0.01	0.05
Helodidae	-0.03	0.01	0.01	0.03
Heptageniidae	-0.96	-0.09	0.27	0.84
Hydrobiidae	-17.56	-4.27	7.99	6.32
Hydroporini	-0.65	0.35	0.27	0.29
Hydropsychidae	-0.53	-0.04	0.16	0.37
Hydroptilidae	-0.20	-0.15	0.24	-0.35
Lepidostomatidae	-2.09	-0.01	1.02	1.62
Leptoceridae	1.84	1.11	-0.31	-1.06

A

Part XVII

18 References

Boyce, R.L. 1998. Fuzzy set ordination along an elevation gradient on a mountain in Vermont, USA. J. Veg. Sci. 9: 191-200.

Bradfield, G.E. & Kenkel, N.C. 1987. Nonlinear ordination using flexible shortest path adjustment of ecological distances. Ecology 68: 750-753.

Bray, J. R. & Curtis, J.T. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. 27: 325-349.

Equihua, M. 1991. Fuzzy clustering of ecological data. J. Ecol. 78: 519-534.

Equihua, M. 1991. Application de la teoria de los conjuntos difusos al analisis de comunidades vegetales. Acta Botanica Mexicana 15: 1-16.

Fulton, M. R. 1996. An application of fuzzy set ordination to determine tree habitat suitability of sites from a regional data set. J. Veg. Sci. 7: 739-746.

Krebs, C.J. 1989. Ecological methodology. Harper & Row, New York.

Loucks, O.L. 1962. Ordinating forest communities by means of environmental scalars and phytosociological indices. Ecol. Monogr. 32: 137-166.

Minieka, E. 1978. Optimization algorithm for networks and graphs. Marcel Dekker, New York.

Palmer, M. W. 1994. Fuzzy sets or blurry artifacts? A comment on Zhang and Oxley. J. Veg. Sci. 5: 439-440.

Palmer, M. W. 1995. Blurry artifacts: a retraction. J. Veg. Sci. 6: 447.

Pielou, E. C. 1984. The interpretation of ecological data. John Wiley & Sons, New York, USA.

Roberts, D. W. 1986. Ordination on the basis of fuzzy set theory. Vegetatio 66: 123-131.

Williamson, M.H. 1978. The ordination of incidence data. J. Ecol. 66:911-920.

Zhang, J.-T. 1994. A combination of fuzzy set ordination with detrended correspondence analysis: One way to combine multi-environmental variables with vegetation data. Vegetatio 115: 115-121.

Zhang, J.-T. & Oxley, E.R.B. 1994. A comparison of three methods of multivariate analysis of upland grasslands in North Wales. J. Veg. Sci. 5: 71-76.

Part XIX

19 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 dropdown menu, or clicking on the right-hand mouse button and choosing Help from the pop-up menu.

If you have problems using the program or entering data which you cannot solve then contact Pisces Conservation by e-mailing pisces@irchouse.demon.co.uk or by phone to England 44 (0)1590 676622 during office hours (09.00 to 17.00).

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

Telephone 44 (0) 1590 676622
Fax 44 (0) 1590 675599

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

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

Part

20 Other Programs by Pisces Conservation Ltd

20.1 Other Programs by Pisces Conservation Ltd

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

Community Analysis Package
Searching for structure in community data, using: NMDS, DECORANA, TWINSPAN, Reciprocal Averaging, PCA, Cluster analysis, Similarity and distance measures.

Ecom
Analytical techniques to detect, visualise and order relationships from both species data and environmental variables. Principal methods: Canonical Correspondence Analysis (CCA), Redundancy Analysis (RDA) and Multiple Regression (MR).

Species Diversity and Richness II
A program to calculate and compare species diversity, estimate total species richness and study patterns of abundance

Population Estimation by Removal Sampling
A program to calculate the statistics from removal trapping experiments.

Simply Growth
To analyse growth curves from length and/or weight data.

Simply Probits
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 the user 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.

20.2 Species Diversity and Richness III

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

Species Diversity and Dichness is a Windows program designed for both the professional ecologist and students. The methods on offer range from the familiar, such as the calculation of various diversity indices and the fitting of common distributions to more recently developed techniques such diversity ordering and total species richness estimators. Together they provide a powerful suite of methods to explore, compare and analyse community diversity. 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
9 Alpha diversity indices including Shannon's, Simpson's, Fisher's and Q statistic.
Bootstrapping of confidence intervals.
Randomization tests to compare diversity between sites.
2 diversity ordering methods.
6 methods to estimate total species richness
Rank order and species accumulation curves.
Fitting of log series, log normal, geometric and broken stick abundance models
Ability to generate simulated data
4 beta diversity measures
Freshwater quality measures including BMWP, ASPT & Irish quality rating.

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.

20.3 Community Analysis Package

CAP searches for structure in ecological community data. It is easy to use, produces quality output and is affordable. A Windows program, it offers analytical techniques commonly used by community ecologists, although many are also widely used by researchers in other fields such as palaeontology, archaeology and the social sciences. Programs to carry out many of these techniques have long been available, but they are often difficult to use as they were written for mainframe computers or PCs using DOS, and frequently have little or no graphical output. In CAP, data can be organised using standard programs such as Excel, and the output is displayed, exported and printed using standard Windows techniques. 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.

Available methods:
Reciprocal Averaging
Non-Metric Multidimensional scaling
Detrended Correspondence Analysis (DECORANA)
Two-way Indicator Species Analysis (TWINSPAN)
7 Agglomerative cluster analysis methods Reciprocal Averaging
Principal Components Analysis (PCA)
26 Similarity and distance measures
Divisive cluster analysis
Association analysis
ANOSIM
SIMPER

20.4 Ecom

ECOM is a Windows® program with a range of analytical techniques to detect, visualise and order relationships in multivariate data where the researcher has information on both the species present and the most influential environmental variables.

The principal methods are:
Canonical Correspondence Analysis (CCA),
Redundancy Analysis (RDA)
Multiple Regression (MR).

Researchers in other fields such as palaeontology, archaeology and the social sciences also use many of these methods.

A particular feature of ECOM is the graphical output that enables the user to quickly visualise the complex relationships within their 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 user supplies two input data sets, in a 2-D array, holding the biological and physical data for the samples (sites) respectively. In ecology, it is usual for the samples, which are normally collected from set localities and may be called, for example, quadrates 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) and Species Diversity and Richness III (SDR), also produced by PISCES, which offer a range of techniques that only require biological data to undertake the analysis.

20.5 Simply Tagging

Simply Tagging is a Windows program to estimate population size for both closed and open populations using mark-recapture methods, designed to be used by undergraduates and allows them to simulate data and explore the accuracy and reliability of mark-recapture methods. Features include data simulation options to model mark-recapture experiments for open and closed populations. The user can vary sampling intensity, population size, animal behaviour and birth/death rates.

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.

20.6 Population Estimation by Removal Sampling

A Windows program to calculate the statistics from removal trapping experiments.

Contains three easy to use methods: constant & variable probability of capture

(Zippin's maximum likelihood, ML), & 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.

20.7 Density from Distance

DfD is a Windows program 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. DfD has been designed for ease of use on PCs running under Windows. 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.

20.8 Dynamica

Users can explore real data, and reach their own conclusions about how animal numbers are inter-related and change through time. The data consist of monthly abundance estimates of fish and crustaceans from January 1981 to January 1995 at Hinkley Point Power station, Severn Estuary, UK. Time series for more than one hundred species are available for study. Details about data collection and habitat are also given. Dynamica was designed as a research tool to help explore the complex inter-relationships between species and environmental factors such as temperature, salinity and tides.

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.

You can find more information on the Hinkley Point studies on our website at http://www.irchouse.demon.co.uk/consultproj1.html

20.9 Simply Growth

A Windows program to analyse growth curves from length and/or weight data.

The program will fit and plot von Bertalanffy, Gompertz and Logistic growth curves to your data. These curves are used to model the growth of a wide variety of organisms but are not simple to fit, as they require the estimation of a number of parameters by non-linear regression.

Particular care has been taken to allow data to be quickly and easily imported from spreadsheet programs such as Excel and the plots can be exported in a variety of forms .

20.10 Simply Probits

A Windows program to 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.

Index

- " -

- A -

- B -

- C -

- D -

- E -

- F -

- G -

- H -

- I -

- J -

- K -

- L -

- M -

- N -

- O -

- P -

- Q -

- R -

- S -

- T -

- U -

- V -

- W -

- Z -