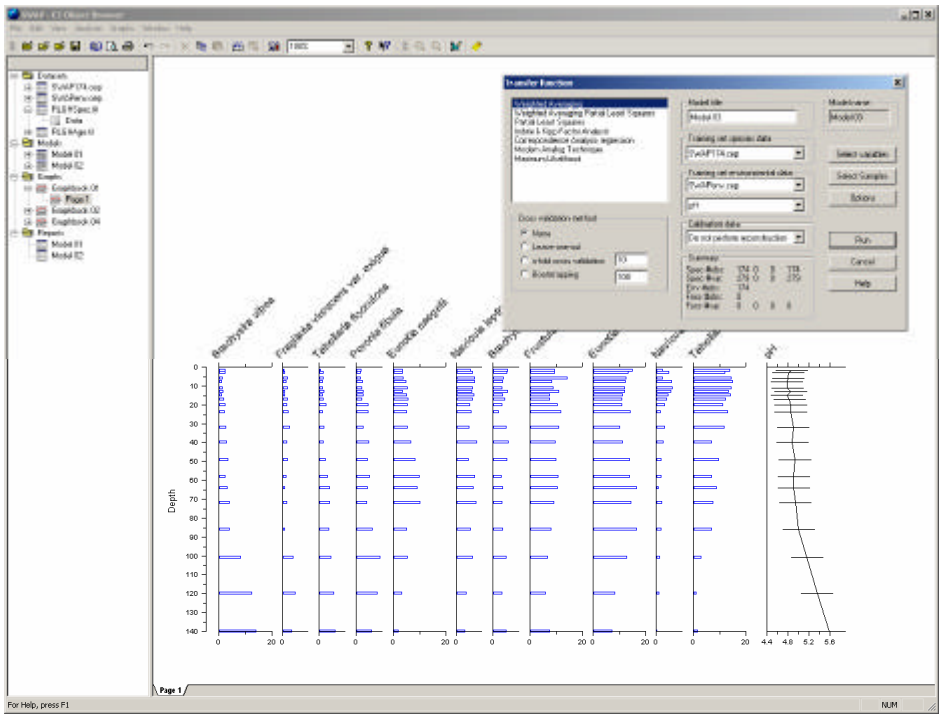
User Guide

C2

Software for ecological and palaeoecological data analysis and visualisation

User guide Version 1.5



Please send comments / bug reports to: Stephen.Juggins@ncl.ac.uk

Steve Juggins
School of Geography, Politics & Sociology
Newcastle University
Newcastle upon Tyne
NE1 7RU
http://www.staff.ncl.ac.uk/stephen.juggins

University of Newcastle © 2003-2007
Steve Juggins © 1991-2007

Contents

Introduction 7

Welcome to C2 7
Program limits 8
Contact 8
Suggested citation 8

Getting started 9

The C2 workspace 9
Importing data 9
Selecting the worksheet / table from an Excel spreadsheet or Database file 10
Duplicate row or column labels 10
Empty rows or columns 11
Importing normalised data 11
Importing CANOCO solution file 12
Viewing / editing data 12
Viewing a dataset 12
Viewing the list of variables 13
Missing values 14
Editing data 14
Spreadsheet properties 14
Renaming a dataset 14
Deleting a dataset 15
Exporting a dataset 15
Saving your work 15
Opening a C2 file 15
Closing a workspace 15

Data manipulation 16

Data manipulation 16
Insert rows / columns 16
Delete rows / columns 16
Sorting by rows or columns 19
Join (merge) datasets 19
Transform data 20
Transpose data 21
Duplicate dataset 21
Classifying data 21
Make classes 21
Edit classes 22
Link dictionary 23
Harmonise 24

Descriptive statistics 25

Descriptive statistics 25

Data analysis 27

Transfer functions 27
Viewing model results 28
Deleting observations or variables from a model 30
Transforming species or environmental data and changing model options 31
Exporting model results 32
Converting model results into a dataset 32
Re-running a model 33
Deleting a model 33
Model diagnostic plots 33
Multivariate analyses 34

Graphical tools 37

Graphics overview 37
Creating a new Graphbook 37
Creating an x-y plot 37
Modifying an x-y plot 39
Series 39
Changing line and symbol styles 40
Symbols 41
Axes 42
Layout 43
Error bars 44
Fonts 44
Plotting multiple datasets on the same graph 45
Zooming in on data in an x-y plot 45
Identifying and labelling data points in an x-y plot 45
Interactive data tips 46
Interactive data labels 46
Labelling from a list 47
Hiding data points in a plot 47
Adding a lowess smooth 47
Adding additional x-y plots to a graph page 48
Deleting an x-y plot from a graph page 49
Undoing / redoing changes to x-y plot 49
Inserting a new graph page into a graph book 49
Deleting a graph page 50
Zooming the graph page 50
Copying a graph page 50
Export a graph page as a Windows metafile 50
Printing a graph page 50
Refreshing the view 51
Opening the graph in a new window 51
Brushing multiple graphs 51

Stratigraphic diagrams 53

Creating a stratigraphic diagram 53
Modifying the appearance of individual graphs 54
X-axis scaling 55
Style 56
Labels 57
Error bars 57
Extras 58
Copying graph formats 58
Changing the y-axis style and position 59
Setting the y-axis variable 59
Adding a secondary y-axis 60
Adding variables from a new dataset to the diagram 61
Removing variables from the diagram 61
Changing the order of variables in the diagram 61

Overlaying variables in a graph 63
Adding a lowess smooth 63
Adding zones 64
Adding groups 66

Global settings and defaults **67**

Global settings 67
User interface 67
Global settings 68
Model settings 68
Graph settings 69
Symbols 69
Fonts 70

Acknowledgements **71**

Acknowledgements 71

Appendix **72**

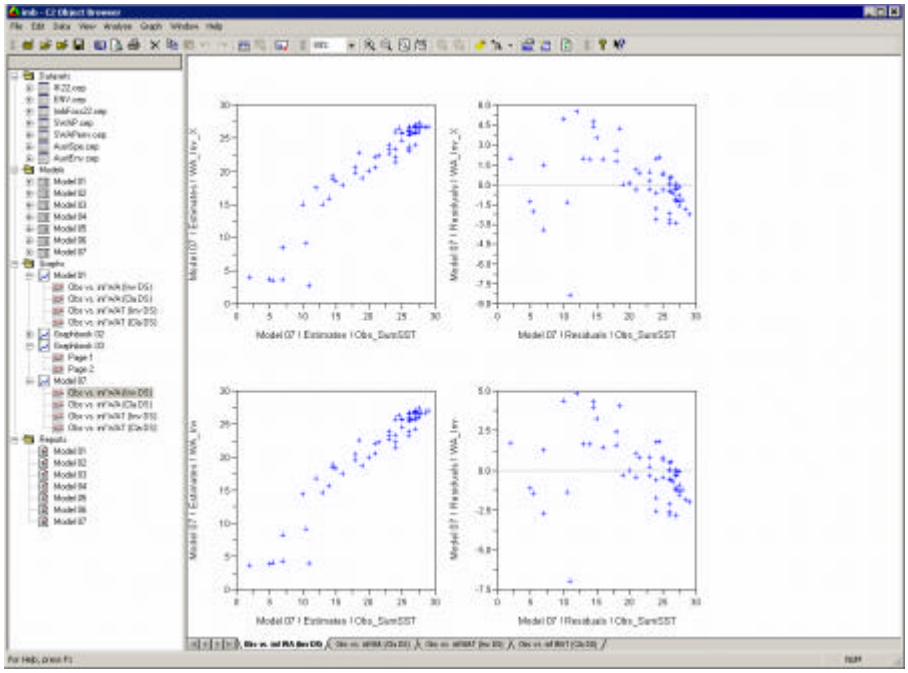
Importing data from Excel 72

Introduction

Welcome to C2

Version 1.5 (2007)

Welcome to C2, a Microsoft Windows program for the analysis and visualisation of ecological and palaeoenvironmental data.



C2 has the following features:

Data manipulation:

Import / export data to / from Excel, Access, Lotus, dBase, Paradox, Tilia, Cornell condensed and full formats, and comma- or tab-delimited text files. Allows basic editing, cut / copy / paste, sorting, transformation of data, and merging of multiple datasets.

Data analysis:

Fits palaeoecological transfer functions, or inference models, to modern training sets of species and environmental data. The following models are currently implemented:

Weighted averaging (WA),
Partial least squares (PLS),
Weighted averaging partial least squares (WAPLS),

Imbrie & Kipp factor analysis (a form of principal component analysis regression),
Correspondence analysis regression (CAR),
Modern analogue technique (k-nearest neighbour) (MAT),
Maximum likelihood regression and calibration (ML).

Options to delete variables and observations from the training set, to transform the species or environmental data, and to perform cross-validation using leave-one-out, n-fold leave out, or bootstrapping.

Performs environmental reconstructions using a secondary, or fossil dataset, with the ability to derive sample-specific reconstruction errors under bootstrapping.

C2 can also generate summary descriptive statistics for a dataset and can fit principal component and correspondence analysis models.

Visualisation:

X-Y scatter and line plots with options to control symbol and line styles, axis scaling.
Scale symbol size according to a third variable.
Add x- and y-error bars as a fixed value or third variable.
Overlay multiple data series with different numbers of observations on same graph.
Interactively zoom data ranges and label data points.
Plot stratigraphic diagrams in a variety of styles (bar, line, silhouette, symbol).
Combine multiple datasets (or multi-proxy data) with different numbers of observations in the same diagram.
Print, print preview and copy graphics to clipboard.

Program limits

The unlicensed free version of C2 is limited to datasets with less than 75 samples (observation, rows). Please contact Steve Juggins to obtain a license and unlock code that removes this limitation.

Contact

Please send comments / bug reports to Stephen.Juggins@ncl.ac.uk

Steve Juggins
School of Geography, Politics and Sociology
Newcastle University
Newcastle upon Tyne
NE1 7RU, UK
http://www.staff.ncl.ac.uk/stephen.juggins

September 2007

University of Newcastle © 2003-2007
Steve Juggins © 1991-2007

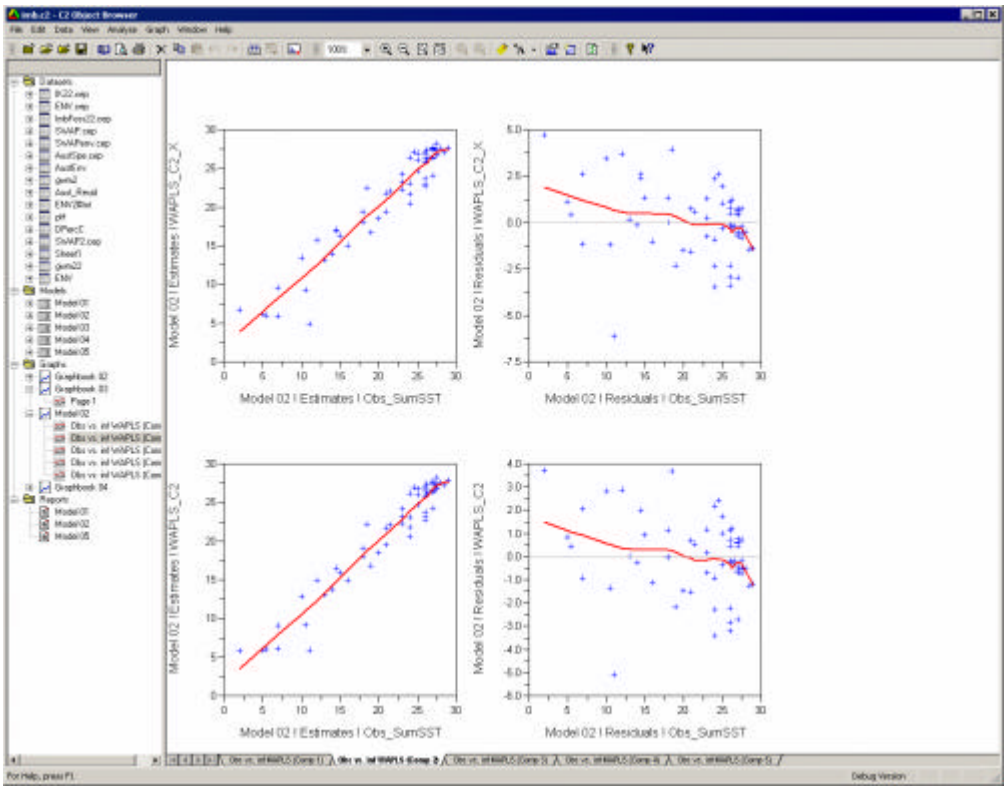
Suggested citation

Juggins. S. 2007 C2 Version 1.5 User guide. Software for ecological and palaeoecological data analysis and visualisation. Newcastle University, Newcastle upon Tyne, UK. 73pp.

Getting started

The C2 workspace

A C2 file or document can be thought of as a workspace that may contain one or more different datasets, the results of one or more analyses (models), one or more graphs, and a summary text report, or log, of each analysis. The tree view in the left pane of the Object Browser window shows a hierarchical view of each of the above C2 objects.



Importing data

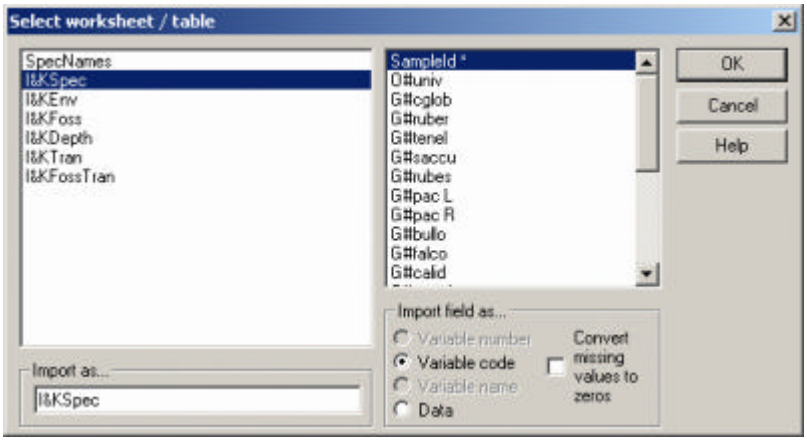
C2 is able to import data in Tilia, Cornell, Excel, dBase, Lotus 123, Paradox, Microsoft Access, and comma- or tab-delimited text (ASCII) file. See the Appendix for the rules for formatting data in spreadsheet or database formats before importing into C2.

To import a dataset into C2 select Open from the File menu or click on the File Open button on the main toolbar . The default file type is *.C2. Select the appropriate file using the drop-down list in the File Open dialog box.

Selecting the worksheet / table from an Excel spreadsheet or Database file

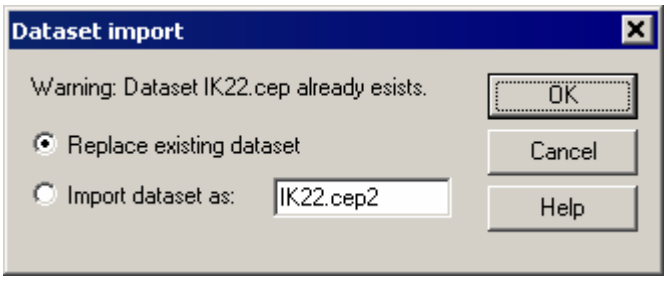
If you import a dBase or Paradox file you will get a popup "Select worksheet / table" dialog box asking you to select again the file from a list of similar files found in the directory. If you are importing data from Excel or Access you will see the same dialog box but with a list of worksheets (Excel) or tables and queries (Access) found in the selected spreadsheet or database. Select the appropriate file, worksheet or table and click OK.

You can optionally specify a name to be given to the dataset when imported. C2 will also list individual fields in the dataset. Character fields are indicated with an asterisk. The first character field will read as the sample or “case” code. The second character field (if there is one) will be read as the same name. Subsequent fields will be read as data. These defaults can be changed using the radio buttons in the right-hand pane. Missing values can be imported as missing (default) or recoded as zero.



Duplicate dataset names

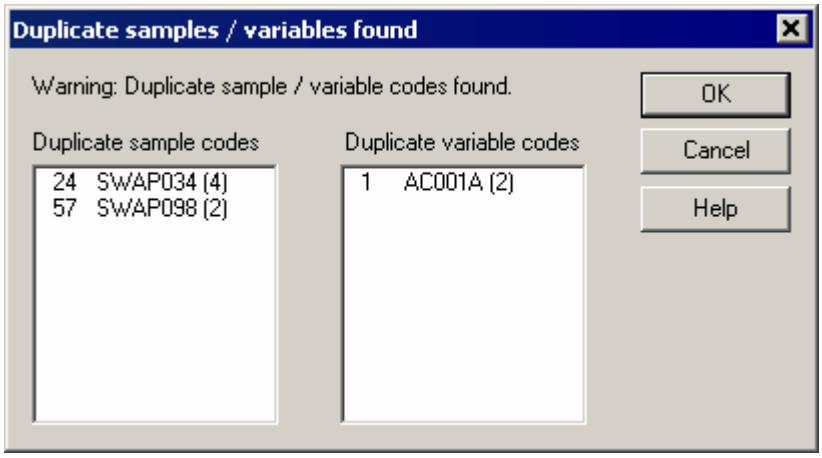
Each item of data in C2 is uniquely identified by its dataset (or model output) and row / column names. Dataset names must therefore be unique. If you try to import a file with the same name as one already loaded the following dialog box will popup and give you the option to either overwrite the existing dataset or to choose a new name.



Duplicate row or column labels

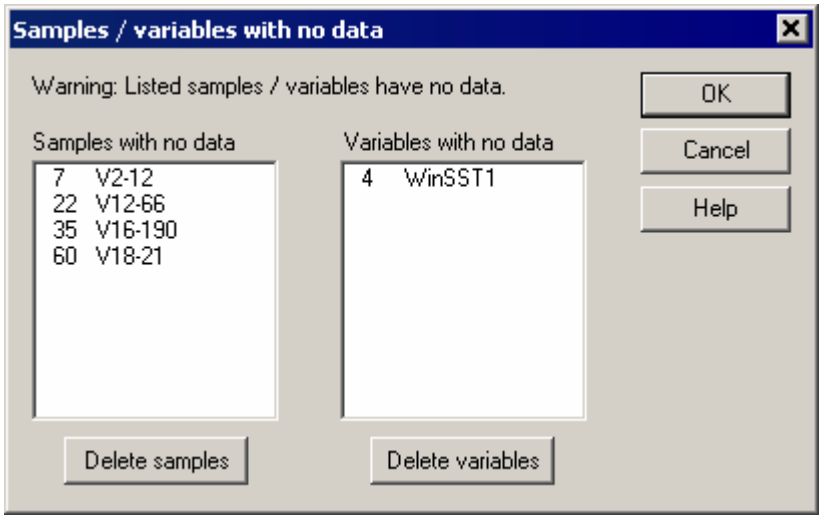
Each row and column of a dataset is identified by a row / column number, a unique short alphanumeric code, and a long name. C2 uses the short code (the species / sample code, or Id) to uniquely identify each row / column of a dataset.

C2 will issue a warning if duplicate row or column codes are found and display a list of the duplicate codes, together with the number of times each occurs (in brackets). You should edit the dataset to ensure that all codes are unique.



Empty rows or columns

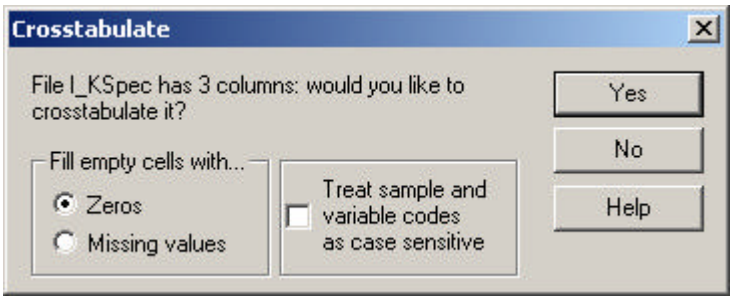
C2 checks the data on import for empty rows or columns. That is, rows or columns that are blank or contain all missing data. You are given the chance to delete the empty samples or variables. Empty columns may be the result of importing badly formatted data in Excel.



Importing normalised data

C2 also has support for reading / writing “normalised” data – that is, data in 3 column form where column 1 is the sample code, column 2 is the variable code and column 3 is the data value. This is a more efficient and effective way of storing and manipulating species abundance data in a database such as Access, Paradox or dBase.

When importing data, if the data consists of only 3 columns, and the first two contain alphanumeric data (corresponding to sample code and variable codes), C2 assumes this is in normalized form (ie. first column represents sample codes, second column variable codes, and third column data values). If C2 finds data like this it will display the following dialog box and ask if you would like to cross-tabulate (ie. convert to a rectangular matrix), using the values in the first two columns as the row and column labels for the new dataset respectively.



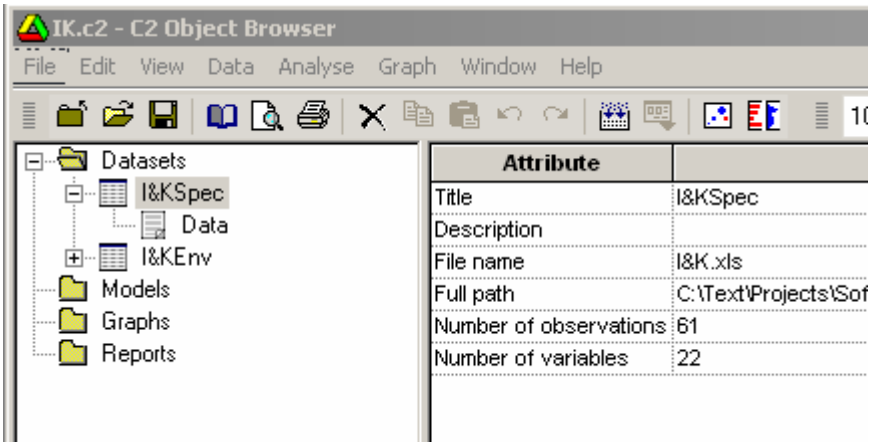
Click Yes to crosstabulate and convert the normalised data into a full table, or click No to import the data directly.

Importing CANOCO solution file

CANOCO solution files may be imported using the File Open dialog, enabling easy plotting of CANOCO output. When imported the CANOCO solution is imported as a model, with it's various components, not as a dataset.

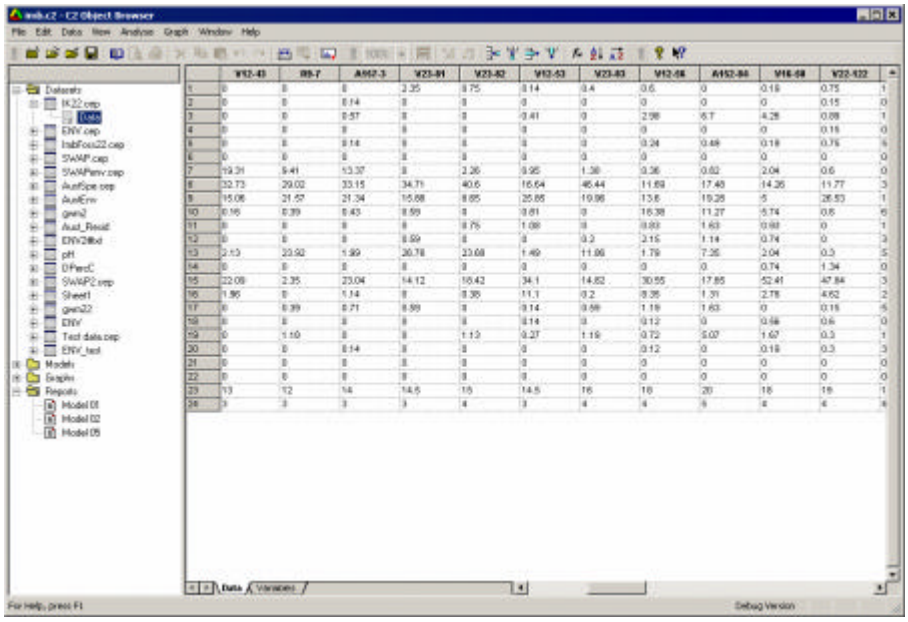
Viewing / editing data

Viewing a dataset



The tree view in the left hand pane of the C2 Object Browser gives a list of imported datasets in the "Datasets" folder. Selecting a dataset name show summary information for that dataset.

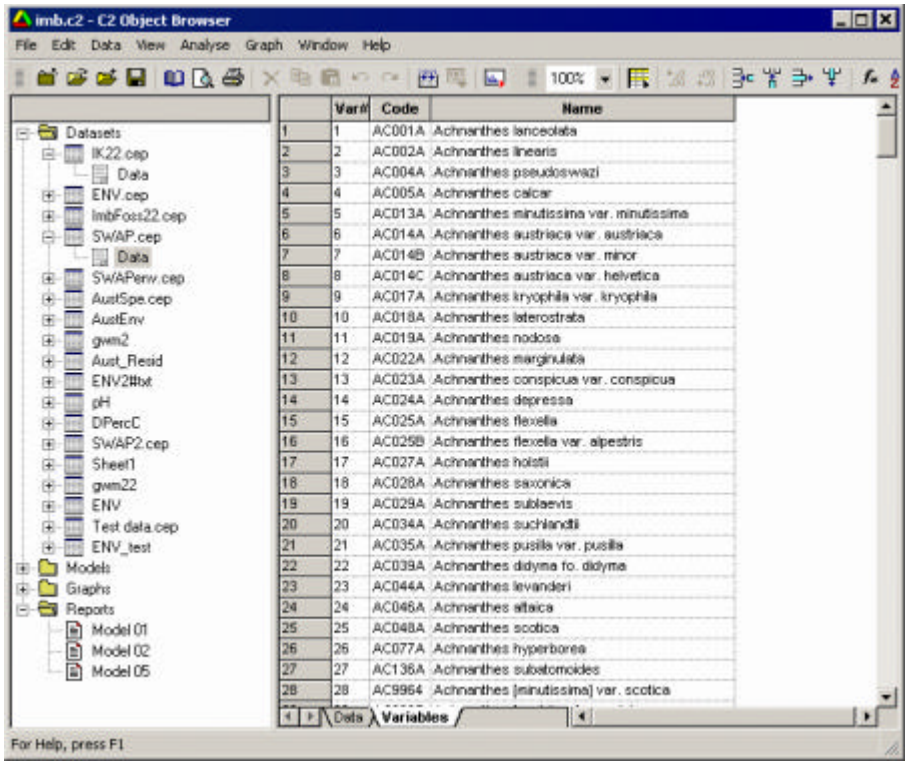
Click on the "Data" node to show a spreadsheet view of the dataset in the right-hand pane of the Object Browser:



The listing shows the sample number, a short unique sample code, a long sample name, and the data for each variable. Datasets are added to the workspace in the order they are imported. It is not possible to change the order.

Viewing the list of variables

Select the "Variables" tab at the bottom of the spreadsheet to show a list of variables for the dataset, listed by short code name and full long name.



Currently only Tilia files support full names for variables: data imported from other types of file will have the full name initially set to the same text as the code name. Long names can be subsequently linked in from a dictionary (see below).

Missing values

C2 understands missing values. These should be coded as null in Tilia, and -99.9 in other files before importing. Missing values are shown as (…) in the data windows.

C2 can handle missing values in the graphic routines (it ignores them) but missing values are not allowed in most of the numerical analyses.

Editing data

When initially opened in the viewer a dataset is in read-only mode. To edit the dataset click on the Edit button on the data toolbar, select Edit from the data menu, or right click the mouse over the spreadsheet and select Edit from the popup menu.

The spreadsheet supports full editing including Excel-like auto-fill, and cut, copy and pasting of cells, ranges, rows, columns or complete tables to Excel or other programs. The spreadsheet also supports full undo / redo, although undo information is lost when the user moves to another view – that is, either selects the variable view or selects another item on the left-hand tree-view.

The spreadsheet does not currently support formulae – any text pasted into cells will be converted to missing values.

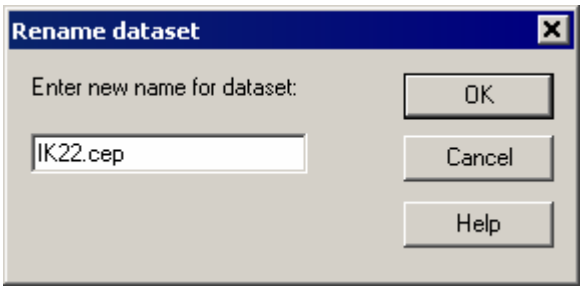
Spreadsheet properties

The zoom level of the spreadsheet can be changed using the drop-down zoom box on the data tool bar, by selecting a zoom-factor on the View menu, or by using the Ctrl-scroll-wheel function on a wheel mouse.

The number of decimal places displayed in the spreadsheet can be changed by clicking on the decimal increment / decrement buttons on the data tool bar. The default view displays numbers in a general view showing a maximum of 6 significant figures. Click these buttons to change the view to a fixed number of decimal places. To return to general format keep clicking the decimal decrement button.

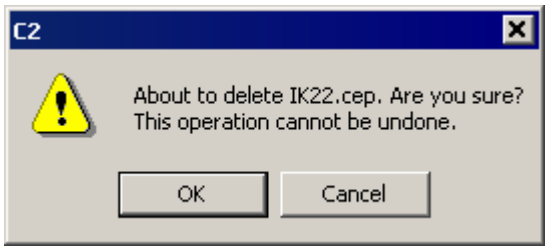
Renaming a dataset

Datasets can be renamed by right-clicking on the dataset name or the Data node in the tree-view, or by right-clicking on the spreadsheet and selecting Rename from the popup menu, or with the dataset selected, choosing Rename from the Data menu.



Deleting a dataset

Datasets may be removed from the workspace by selecting the dataset **name** in the Explorer Pane and selecting Delete from the Edit menu, clicking on the delete button, or pressing the delete key. Deleting a dataset from the workspace cannot be undone so you will be asked for confirmation.



Exporting a dataset

Datasets can be exported in the following formats: Access, Excel, Cornell condensed or full, Tilia, dBase, Lotus, Paradox or ASCII comma- or tab-delimited. Click on the dataset in the tree-view and select Export from the file menu, or right click on the name and select Export from the popup menu. When exporting to Access, if a database already exists with the name given in the File Export dialog the table will be added to the existing database. If the table name already exists in that database you will be asked to give a new name or overwrite the existing table. If the database (*.mdb file) doesn't already exist it will be created as an Access 2000 database and a table added.

To save data in normalised form (3-column form – see Data Import) there is now an additional checkbox on the File Export dialog that is enabled for spreadsheet, database, or comma / tab delimited text files. Zero data values can be included or excluded from the output.

Saving your work

The C2 workspace can be saved by selecting Save from the file menu or clicking on the Save button on the main toolbar. The single C2 file contains all the embedded datasets, models, graphs and reports in the Workspace.

Opening a C2 file

A saved C2 file can be loaded into the workspace using the Open command on the File menu or by clicking on the Open button on the main toolbar.

If there are existing unsaved objects in the workspace you will be given a chance to save them before opening a new C2 file.

Closing a workspace

An open workspace can be closed by selecting Close on the File menu or by clicking on the Close button on the main toolbar.

If there are existing unsaved objects in the workspace you will be given a chance to save them before closing the workspace.

Data manipulation

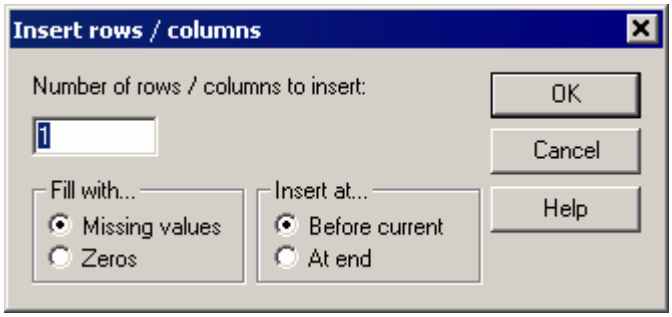
Data manipulation

This section describes C2's data editing and manipulation functions that can be useful in preparing data for subsequent numerical or graphical analysis.

Insert rows / columns

Select and view the appropriate dataset and choose insert rows or columns from the Data menu or click on the insert rows / columns button on the data tool bar . The dataset does not need to be in Edit mode to insert or delete rows / columns.

You are given the option to insert one or more rows or columns before the row / column of the currently active cell, or at the end of the spreadsheet. If there is no currently active cell this option will be disabled.

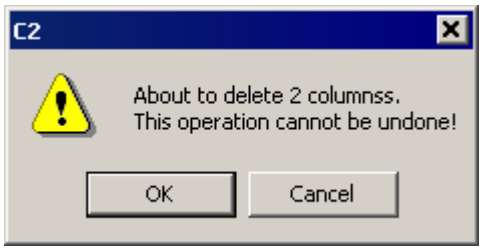


You can either fill the new rows / columns with zeros or missing values.

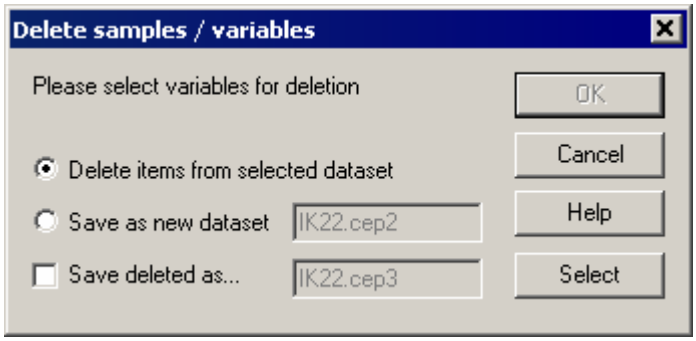
Delete rows / columns

Select and view the appropriate dataset and choose delete rows or columns from the Data menu or click on the delete rows / columns button on the data tool bar .

If there are currently selected rows or columns you will be asked to confirm their deletion:

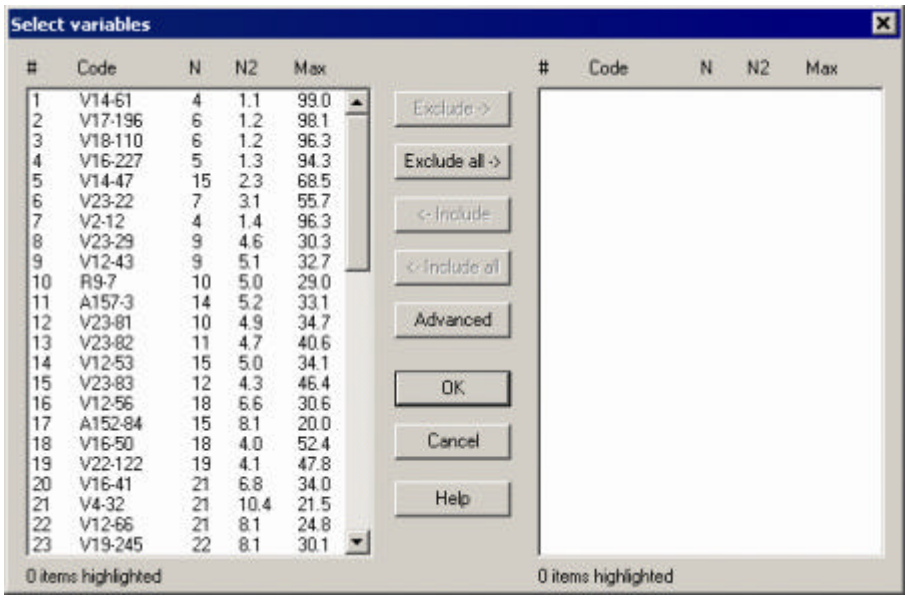


If there are no currently selected rows or columns you will be asked to select some in the Delete samples / variables dialog box:



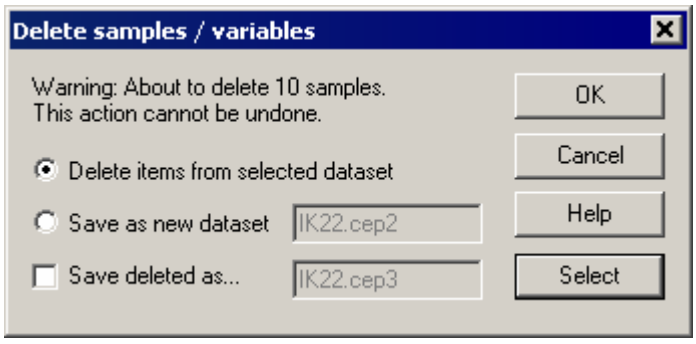
To select samples or variables for deletion click the "Select" button.

If you are deleting samples (rows) you then get a chance to select the appropriate samples from the Select samples dialog box. Move samples for deletion to the right-hand list using the exclude button.

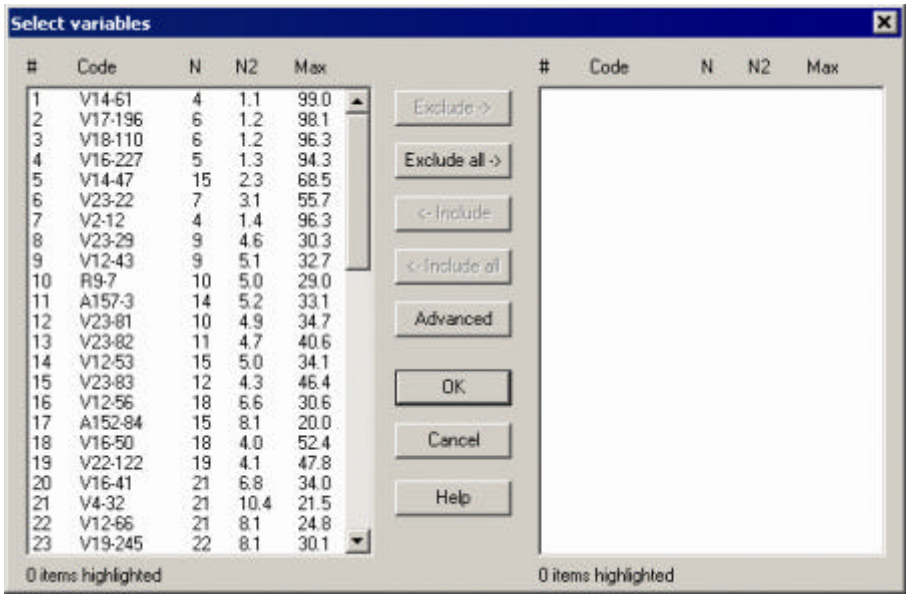


Click OK when done. The delete samples / variables dialog box now shows that some samples are selected and the OK button is enabled.

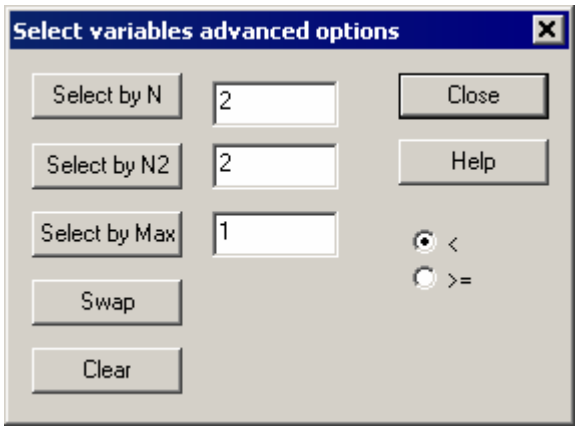
You can now choose to delete the samples from the original dataset or create a new dataset consisting of the original minus the selected samples. You also have the option to save the deleted samples as a new dataset. Thus the delete samples / variables function can also be used to split one dataset into two.



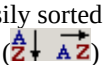
If you are deleting variables you are asked to select the variables for deletion from the Select variables dialog box. This is similar to above except it also shows N (number of occurrences), N2 (Hill's measure of effective number of occurrences), and maximum value for each variable.



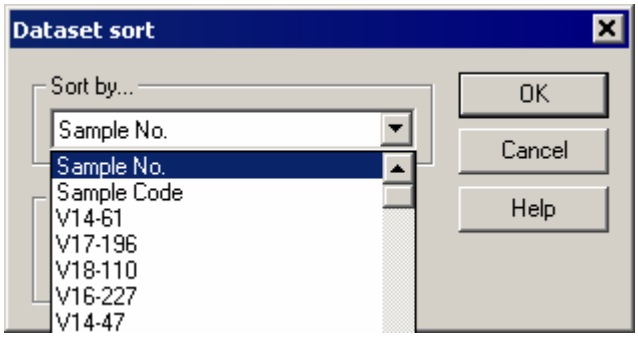
Clicking on the Advanced button invokes the Advanced options dialog box, allowing rapid selection using N, N2 or maximum abundance criteria. For example, clicking on the max button below would select all variables with a maximum value of less than 1.0. This function can be used, for example, to quickly delete rare species from a dataset.



Sorting by rows or columns

Datasets can be easily sorted by rows or columns by clicking on the sort buttons on the data toolbar () by opening the dataset and selecting Sort rows / columns from the Data menu.

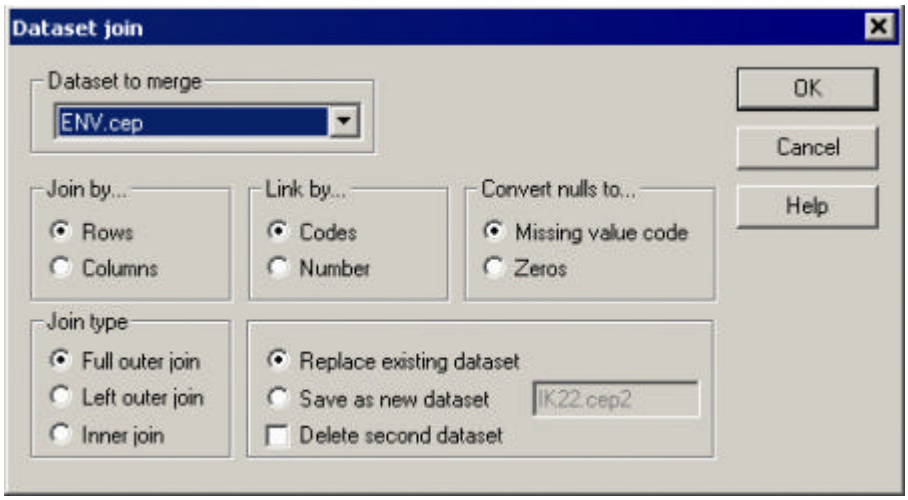
For samples, you can sort the dataset by sample number, sample code, or any variable in the dataset, in ascending or descending order.



For variables, you can sort by variable number, code name, or the data in any sample. You can also sort variables by their weighted average of their sample number. This is useful for species data that are in stratigraphic order – sorting them in this way orders the species in terms of their occurrence in the core, which provide a useful starting point for their order in a stratigraphic diagram.

Join (merge) datasets

The join (merge) function allows you to join two datasets together row-wise or column-wise, matching rows / columns by number or code name.



This function merges two datasets. Dataset 1 is the currently open or viewed dataset. Select dataset 2 in the top left combo box.

Joining by rows merges the two datasets by matching row numbers or codes. Joining by columns merges the datasets by matching column names or numbers.

There are three options for the type of join, which deal differently with non-matching rows / columns.

Let us say we are merging 2 datasets by rows. Full outer join will include all the rows from datasets 1 and 2, including rows in either dataset that are not matched in the other. The empty cells in mismatched rows will be filled with missing values or zeros, depending on what is selected in the convert nulls to… box.

Left outer join will contain all the rows of dataset 1 regardless of whether they are matched in dataset 2, and only those rows in dataset 2 that are matched in dataset 1. Inner join only includes rows that are matched in both datasets.

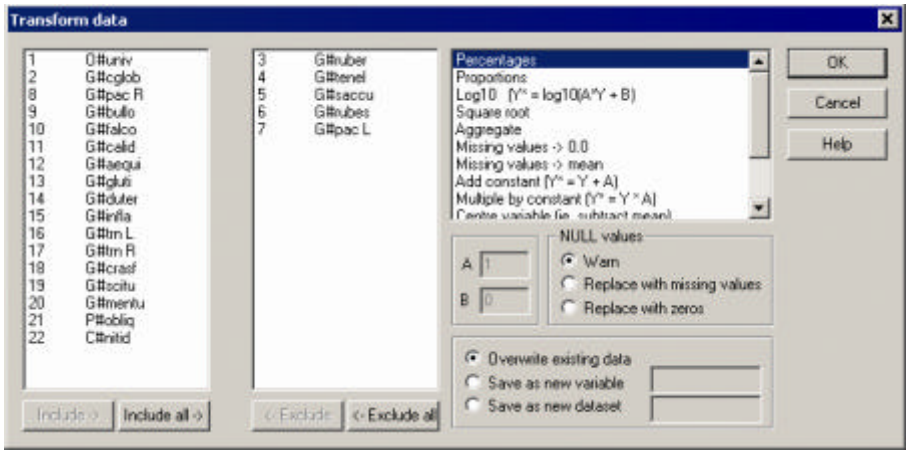
Link the rows or columns by sample code or number. By default, alphanumeric comparisons are not case sensitive. This behaviour can be changed in the global options.

Finally, give a destination for the merged dataset: either replace the original or save as a new dataset. You also have the option to delete dataset 2 after th emerge.

Note: by default code matching is case insensitive. Go to **Edit > Options > Global** tab on the main menu to make matching case sensitive.

Transform data

The transform function allows simple transformation of individual or groups of variables. Choose Transform from the Data menu or click the Transform button on the data tool bar .



Move the variables to be transformed to the right-hand list box using the include / exclude buttons. Then select a transformation. Current methods are:

1. Percentages – all selected variables are transformed to percentage based on the sum of those selected.
2. Proportions – as for percentages.
3. Log10 – all selected variables are transformed to log base 10, according to the formula log10(ax + b), where a and b are entered in the edit boxes.
4. Square root – all selected variables are converted to square roots.
5. Aggregate – a single new variable is created by summing the selected variables.
6. Missing values -> 0.0. Missing values are converted to zero.
7. Missing -> mean. Missing values are converted to the mean of the variable.
8. Add constant – the number entered in the "A" edit box is added to all values.
9. Multiply by constant – all values are multiplied by the number entered in the "A" edit box.
10. Centre variable – variables are centred ie. their mean is subtracted.
11. Centre and standardise. Variables are centred and divided by their standard deviation.
12. Absolute value – values are converted to their absolute value.

13.Change sign.

New variables created during the transformation can be:

1 copied over the original,
2. saved as new variables in the original dataset, or
3. Saved in a new dataset.

With (2), if only one variable is selected, the text entered into the adjacent text box will be used as the new name. If more than one variable is used, the test will be used as a suffix to the original names.

Any missing values generated during the transformation (e.g. by log of zero), can be converted to missing values, replaced with zeros, or prevented by a warning .

Transpose data

Selecting transpose from the data menu simply transposes a dataset and its row column names and numbers. This function can sometimes be useful to get around the 255 column limit in data imported from Excel. Note that C2 has no such limit.

Duplicate dataset

Selecting Duplicate dataset from the Data menu does just that. This option can be useful to keep an original versions of a dataset in the Workspace, prior to transformation or other editing of the copy.

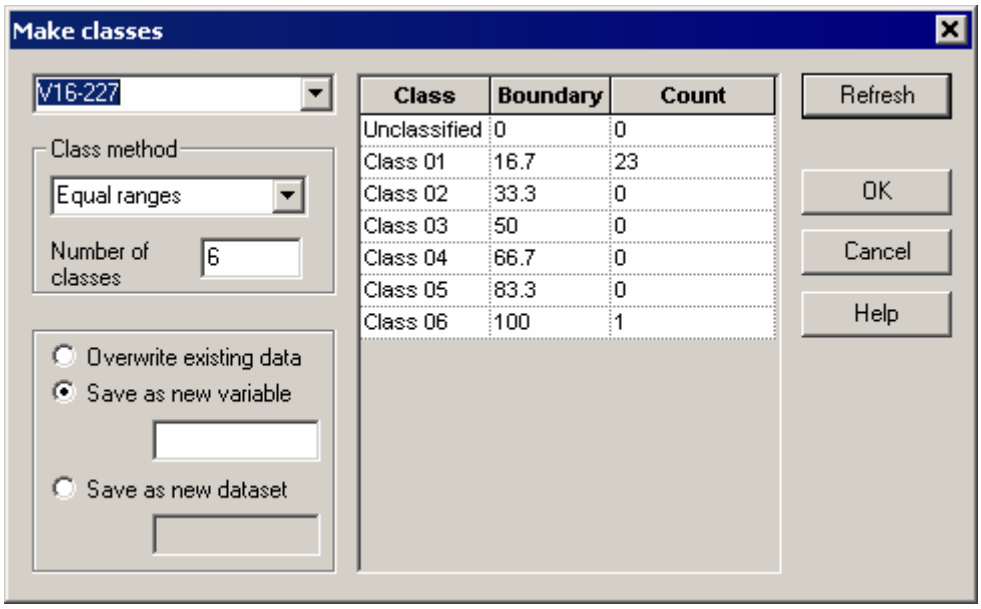
Classifying data

There are two functions to help you make classes or groups from the data. These functions are currently manual and should not be confused with numerical classification techniques like cluster analysis. Their main use at the moment is to generate class or grouping variables for use in visualisations.

There are two class functions. The first Make classes, helps define groupings of samples or variables from the raw data. The second, Edit classes, is a tool to edit an existing class variable.

Make classes

Open a dataset in spreadsheet view and select Classes -> Make classes from the Data menu.



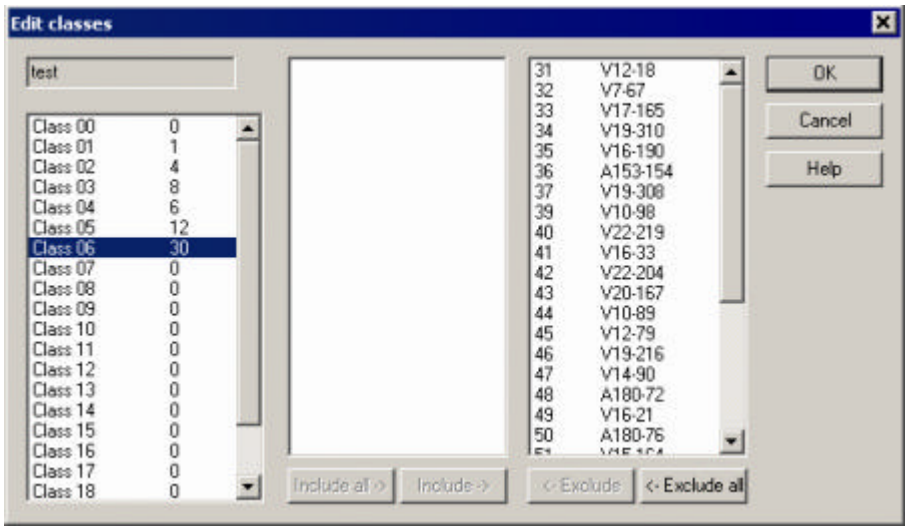
The Make classes dialog box accepts a single variable in the top right combo box and seeks to partition the individual data values into a number of groups. Each class can have up to 20 groups although in practise we usually use a much smaller number. Class method gives a number of choices for splitting the range of the selected variable. The default is Even range - nicely rounded intervals are formed based on the minimum and maximum values of the variable and the chosen number of classes. Change the variable, number of classes and click Refresh to update the grid showing the class boundaries and numbers of items in each group. Any unclassified items (ie. with missing values, or falling outside the class boundaries, are placed in class 0).

Other classification methods are Equal count and percentage (boundaries at 0, 5, 10, 20, 50, 100 – useful for percentage species data), and 10 user defined ranges. These can be set and saved in the C2 configuration database for future use.

You can save an integer variable that lists the class membership of each item, either overwriting the original variable, as a new variable in the existing dataset, or to a new dataset.

Edit classes

To use the Edit classes function first select a cell in the spreadsheet of the class variable you want to edit. This must be an integer variable – you will get an error if it is not an integer class variable. Then right click the mouse and select Classes -> Edit classes from the popup menu, or selected Classes -> Edit classes from the Data menu.



The Edit classes dialog box shows, on the left, a list of classes and the number of items in each class. Use the list boxes on the left to remove items from a class and to place them first into class 0 (unclassified) and then into a new class.

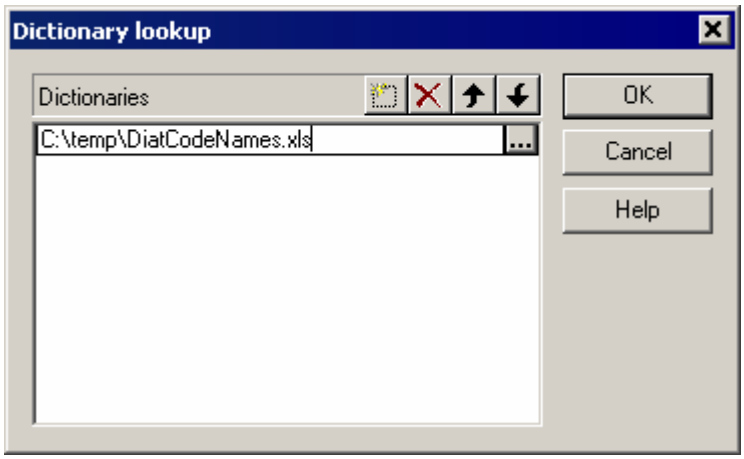
Link dictionary

The Link dictionary function allows you to link a dictionary of species full names into a dataset.

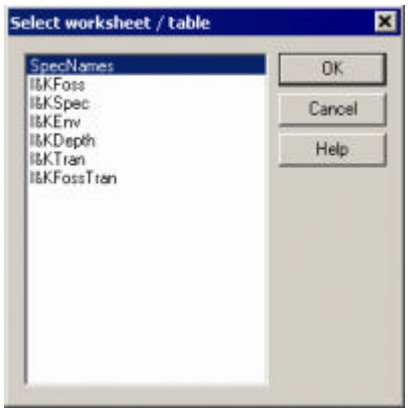
To link a dictionary, open the dataset in spreadsheet view and select Link dictionary from the Data menu. The dataset can be in either data or variable view, but it is more useful to have it in variable view.

Up to 10 dictionaries can be defined and saved in the C2 global settings database.

Click the new file icon to add a new dictionary to the list and click the small button on the right hand side of the list to browse for the file. A dictionary can be in Excel, Paradox, dBase, or Lotus format. It should be formatted so that the first field or column in the spreadsheet / database contains the variable short code, and the second field contains the full name. The content of other fields is not important. Matches between the dictionary and dataset codes are not case sensitive by default, although this behaviour can be changed in the global settings.



Highlight the dictionary name and click OK and C2 will read the names from the dictionary and match them, using the variable codes, to the codes in the dataset. This operation may take a few seconds if you have many names in the dictionary. If you have more than one worksheet in an Excel file (or table in an Access database) you will be asked to select the appropriate sheet from a popup list:

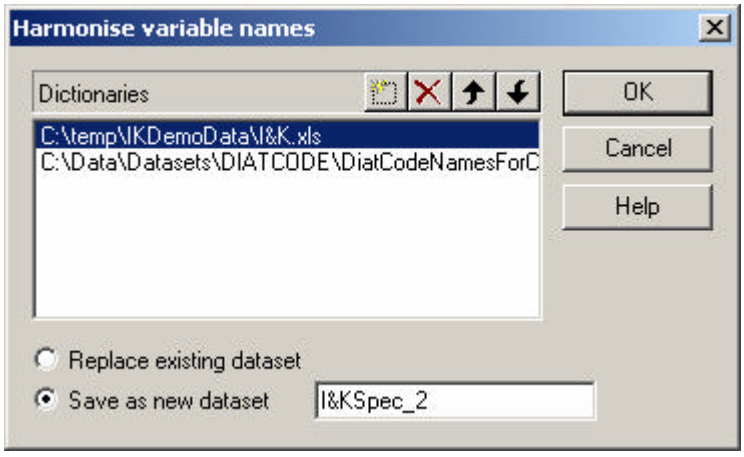


Select the sheet containing the dictionary and click OK.

When finished C2 will report how many names it found and matched.

Harmonise

Harmonising a dataset is a way of renaming and merging variables (taxa) to match the names in another file. This is useful to harmonise a biological dataset (e.g. fossil material) to match the taxonomic criteria used in a surface dataset. Harmonise takes an Excel file of names and records / merges and uses this to modify a dataset. The Excel file should contain alist of at least two columns: the first column is the old variable code, the second the new code. An optional 3rd column may give the long name for any new variable codes. C2 will then rename variables in a dataset from the codes in column 1 to those in column 2 of the dictionary. After renaming it will merge (sum) the data for any duplicate codes.

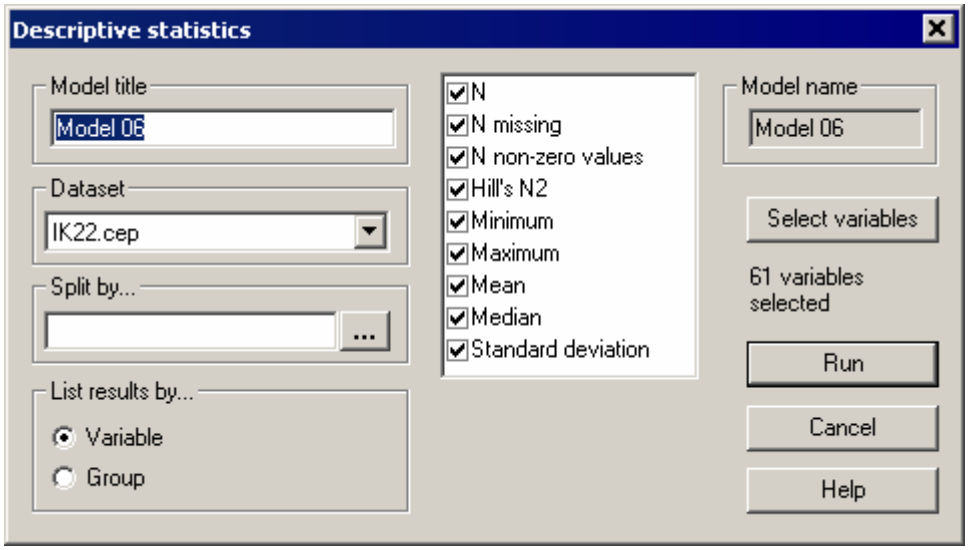


Descriptive statistics

Descriptive statistics

C2 has a single function to calculate summary descriptive statistics for a dataset. Such statistics are a useful starting point in analysing a new dataset.

To calculate descriptive statistics for a dataset choose Descriptive statistics from the Analyse menu.



Output from this function is treated as a model result and will be added to the Model folder on the tree view. Models in C2 have two names, an internal name (which is simply Model 01, 02 etc.) and a user given title. The internal name is used to identify the model within the program. The user title can be more recognisable description of the model. The internal name is generated by the program and is shown on the top right of the dialog box. The user title is initially set to the same text as the internal model name but can be changed before the model is run.

To run the model first select the appropriate dataset in the dataset combo box. All summary statistics are calculated for all variables by default. Click Select variable to remove some from the analysis, and deselect any summary statistics that you don't want. Summaries can be calculated for the dataset as a whole, or for groups of samples by specifying a "Split by" variable. The split by variable must be an integer with group numbers indicating group membership for each sample in the dataset. It must have the same number of rows as the dataset.

If you are splitting the data by group you can either have the output listed by variable or by group. For the former the output lists the summaries for all the

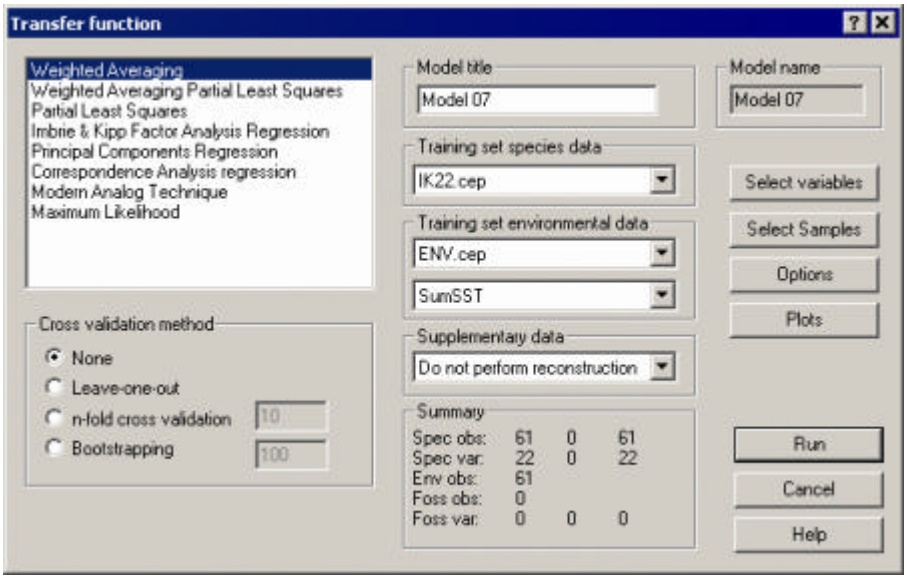
variables for group 1, then group 2 etc. For the latter, the list contains the summaries for variable 1 for the different groups, then variable 2 etc. The latter listing makes it easier to compare differences between groups for each variable.

The results of the model are added to the tree-view under the model number.

Data analysis

Transfer functions

Choose Transfer function from the Analyse menu or click on the Run transfer function button on the main toolbar .



Select the desired transfer function type, choose a cross-validation method, and select the datasets to be used for the training set species data, the training set environmental data, and the environmental variable to be predicted. You may also provide an optional fossil dataset under "Supplementary data" if you want to perform a reconstruction.

C2 will automatically generate a name for your transfer function in the sequence Model01, Model02 etc. This name is used within the program to refer to the results of a particular model run. You may also specify a separate model title to provide a more useful memory-jogger name for a particular model.

The matrix of numbers under Summary list the total numbers of observations and variables in the species, environment and fossil data. For the number of observations and variables in the species data and the number of variables in the fossil data the Summary also lists the number of explicit deletions, implicit deletions, and number of active observations or variables (ie. the number after any deletions). See below under "Deleting observations or variables" for a definition of these terms.

Click Run to run the model. A dialog box shows the progress of the model calculations. When complete the model results are added to tree in the Explorer pane under Models and a summary is added under Reports.

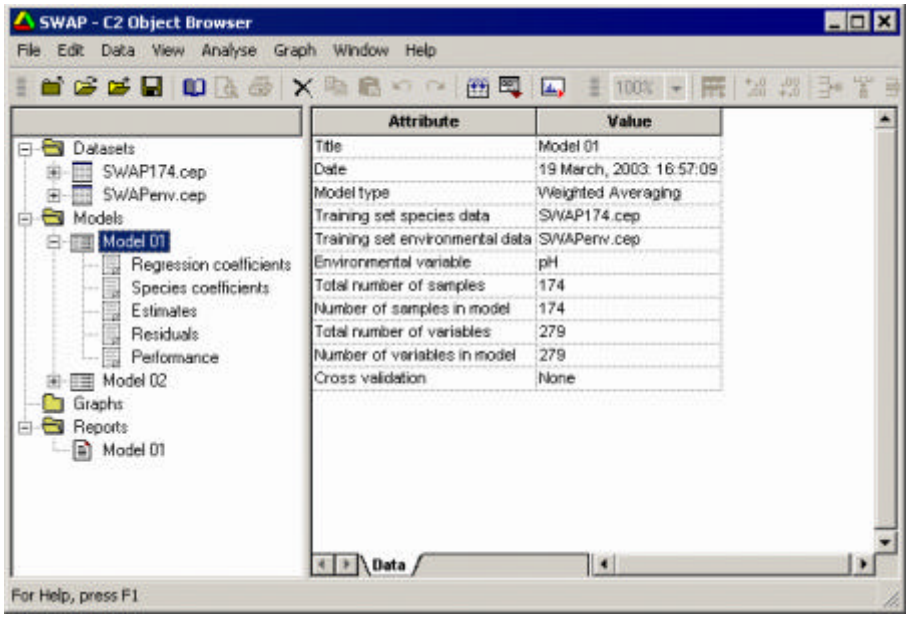
Note: any variables that occur in the fossil data but are missing from the training set are excluded from the reconstruction for all methods except Modern Analogue Technique. For this method, all taxa in the fossil dataset (even if they

have been deleted from the training set) are still used in the reconstructions. For this method it is therefore important that the fossil data only contains species that should be included in the reconstruction.

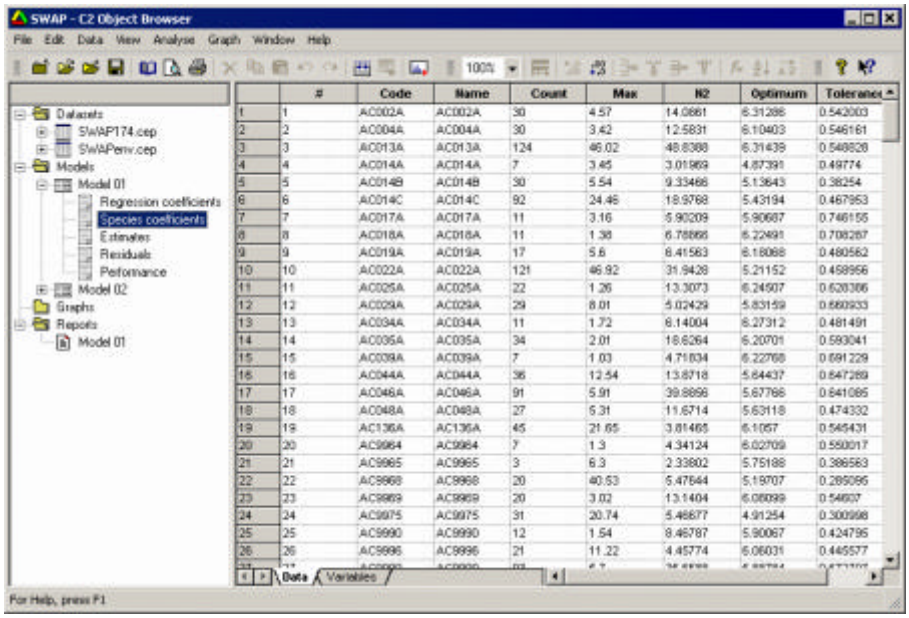
Note also: modern and fossil datasets are matched according to taxon codes. This matching is case insensitive by default. Go to **Edit > Options > Global** tab on the main menu to make matching case sensitive.

Viewing model results

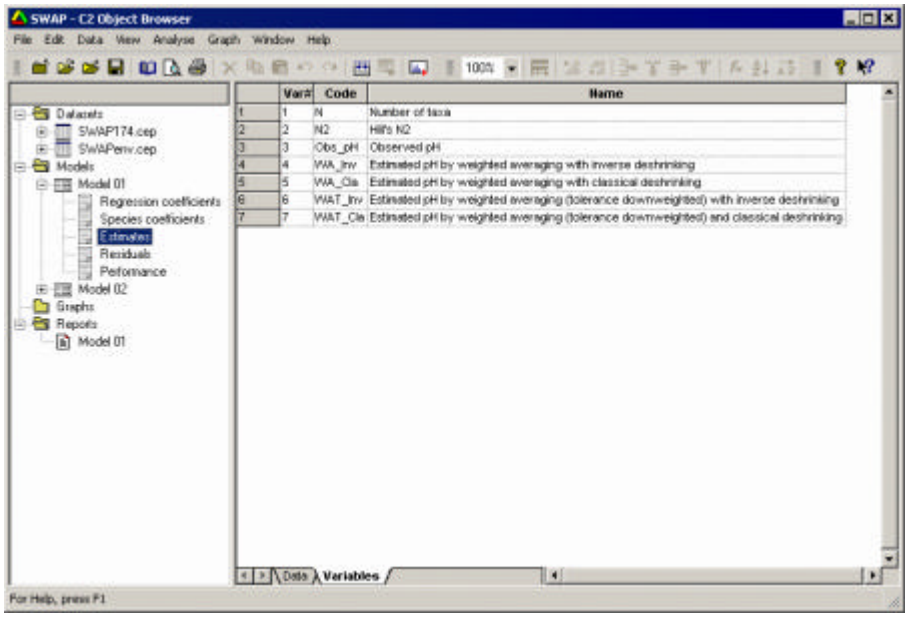
The various components of the fitted model are stored as separate datasets under the model name in the tree-view. Selecting the node of a named model displays a summary of the fitted model in the Data Pane:



Selecting a model component displays the results for that component in a spreadsheet:



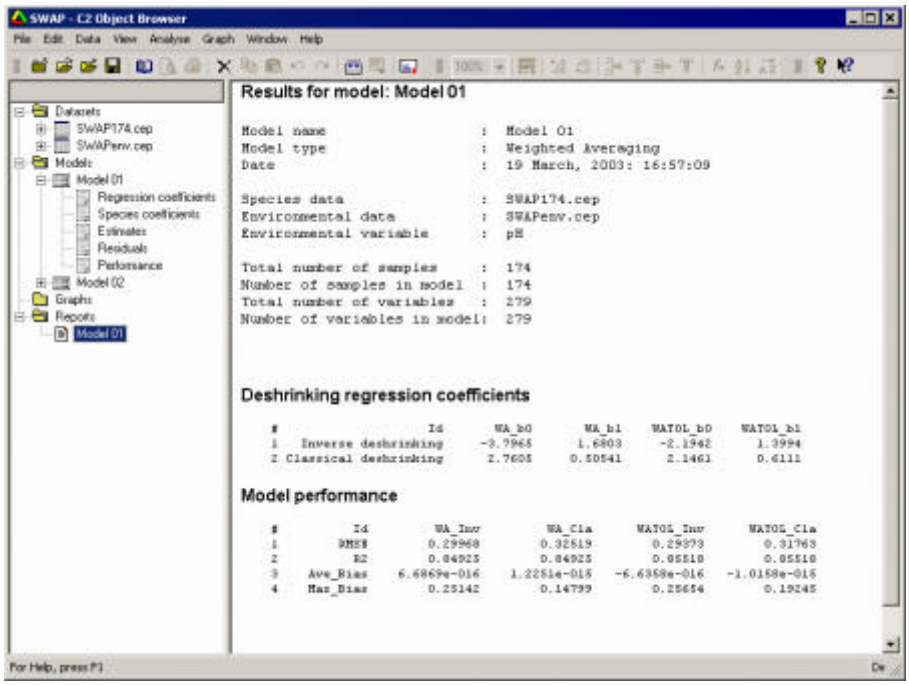
Selecting the Variables tab at the bottom of the spreadsheet displays a list of variables for that model component, together with the full explanatory name:



The list of components is slightly different for each model. For the training set, all models list species coefficients, sample estimates, sample residuals, and model performance. If bootstrap cross-validation is performed all models also list sample specific standard errors.

If a fossil dataset is also specified all models also list the reconstructions, species diagnostics, and reconstruction diagnostics. If bootstrap cross-validation is performed all models also list sample specific standard errors for the fossil samples.

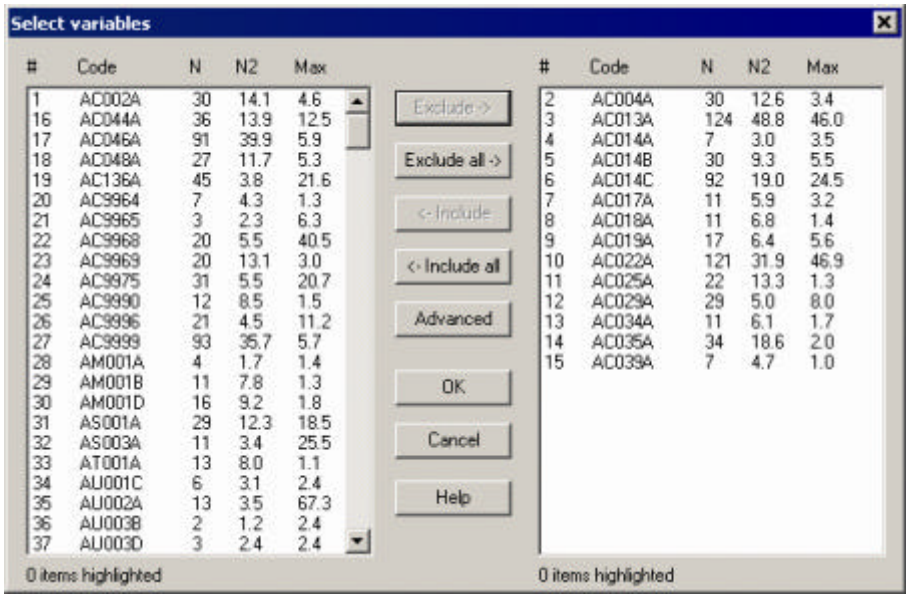
A summary of the fitted model is also listed under the Reports section of the left-hand tree view. This displays the model types, datasets etc., the model performance, and a list of variables excluded from the transfer function or reconstruction.



Deleting observations or variables from a model

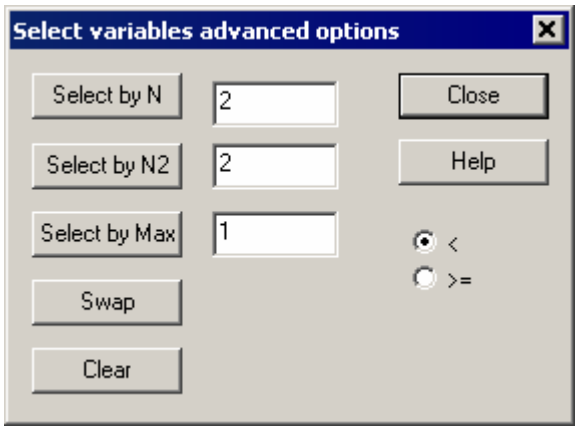
By default all observations and variables in the training set are used to generate a transfer function.

To delete variables click the Select variables button on the Transfer function dialog box. The list box displays a list of variables included in the model, together with the number of occurrences (N), the effective number of occurrences (N2) and the maximum abundance (Max) for each variable.

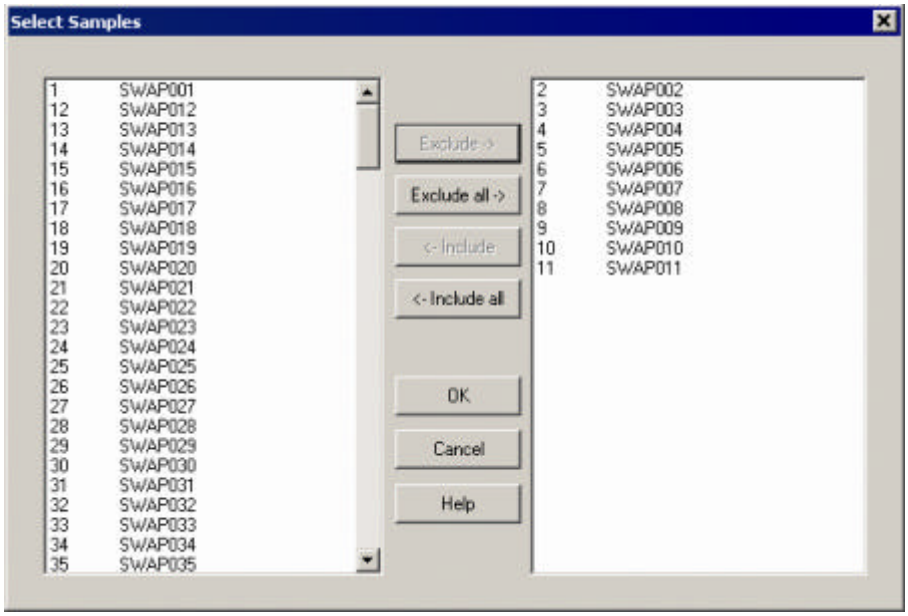


Exclude the variables to be omitted from the model and click OK.

The Advanced button allows selection of variables by N, N2, or Max.



The Select samples option on the Transfer function dialog box similarly allows the exclusion of observations:

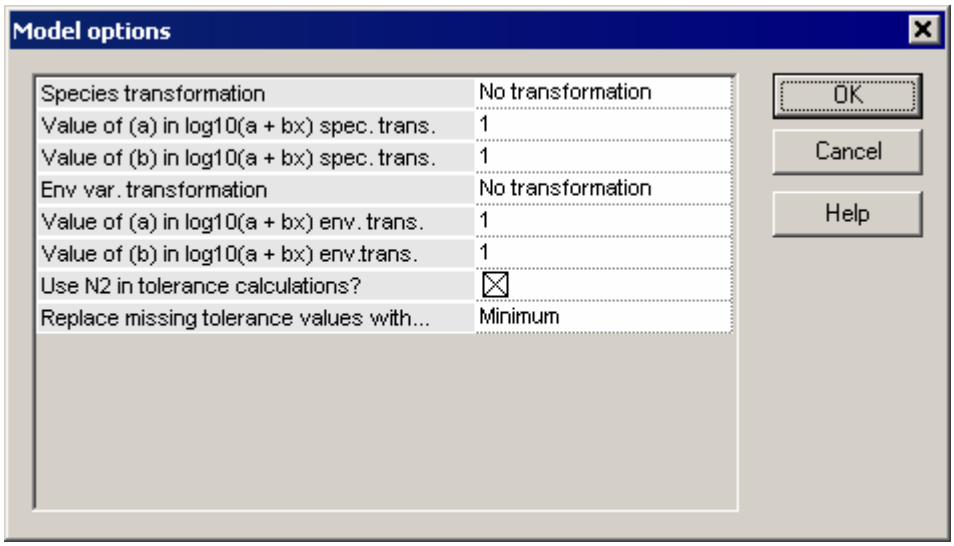


Note: excluding variables or observations does not delete them from the original dataset – it simply omits them from this model run.

The number of explicit and implicit deleted observations and variables is updated in the "Summary" list in the transfer function dialog box. Explicit deletions are those variables or observations excluded by the user. Implicit deletions are those variables or observations that would have no data once explicit deletions have been made. For example, if a user excludes 10 taxa from the training set, but retains all observations, and in doing so is left with an observation that now contains no taxa, this observation has no data and becomes an implicit deletion.

Transforming species or environmental data and changing model options

Species and / or environmental data can be transformed and options for each model (transfer function type) can be changed by selecting Options in the Transfer function dialog box:



Species data may be transformed to percentages, proportions, square roots or log10. Environmental variables may be transformed to square roots or log10. Invalid transformations (e.g. square root of a negative number) will result in an error.

Note: log10 transformations are only applied to non-zero values in the dataset. Original zero values remain zero.

Any species transformations are also applied to fossil data if a dataset is given for the supplementary data

The exact options displayed will vary according to the type of model. Most are self-explanatory.

Exporting model results

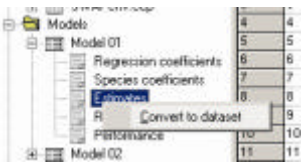
Model results can be exported to an Excel file by either highlighting the model title in the tree view and selecting Export from the file menu, or right click on the model in the tree view and select Export from the pop-up menu. Each model component is inserted into a separate worksheet of the Excel file, and a copy of the model summary that is written to the Report is inserted into the first worksheet.

If you only want to export one table of results from the model you can either copy and paste it into Excel, for example, or you can convert the model component to a dataset and export the dataset.

Converting model results into a dataset

Model results cannot be edited. If you want to modify a results table, or export it to a format other than Excel you first have to convert it to a dataset.

To do this right-click on the component of interest and select Convert to dataset from the pop-up menu.

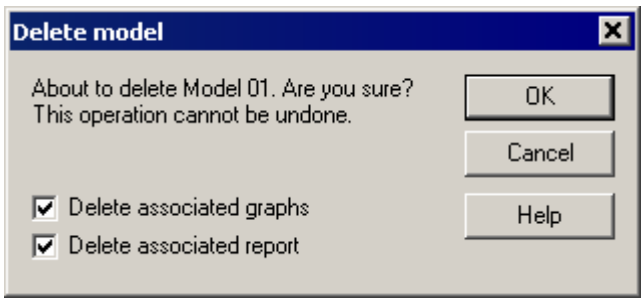


Re-running a model

Developing a transfer function often involves an iterative sequence of model fitting, model diagnosis, identification of outliers, and re-fitting. To aid this process it is possible to open the Transfer function dialog box with the settings (cross-validation options, transformations, model options and variable or observation deletions) of a previous model run. To do this highlight the model in the Explorer tree view and choose re-run model from the Analyse menu, click on the Re-run model button on the main toolbar, or right click on the model in the Explorer view and select Re-run model from the pop-up menu

Deleting a model

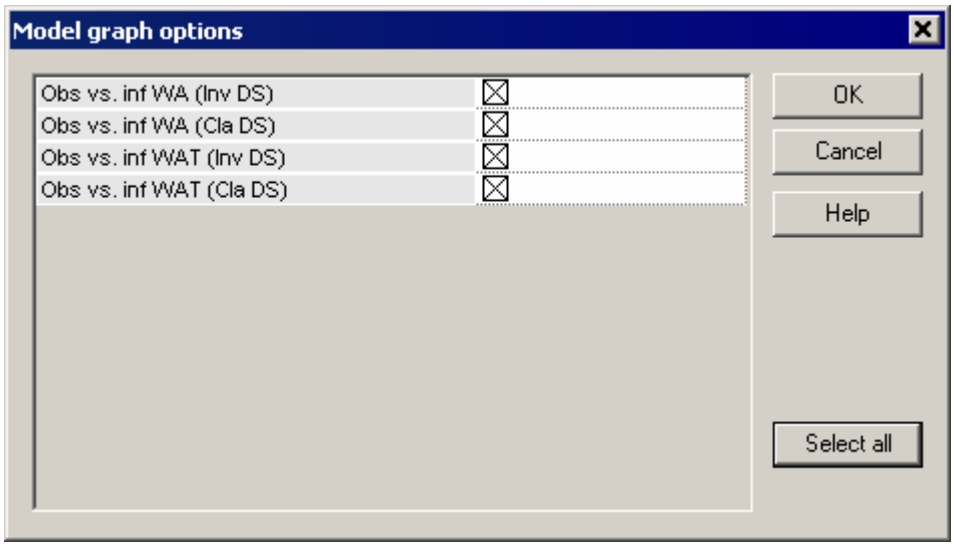
To delete a model highlight it in the Explorer tree view and choose Delete from the Edit menu, click on the Delete button on the main toolbar, or right click on the model in the Explorer view and select Delete from the pop-up menu.



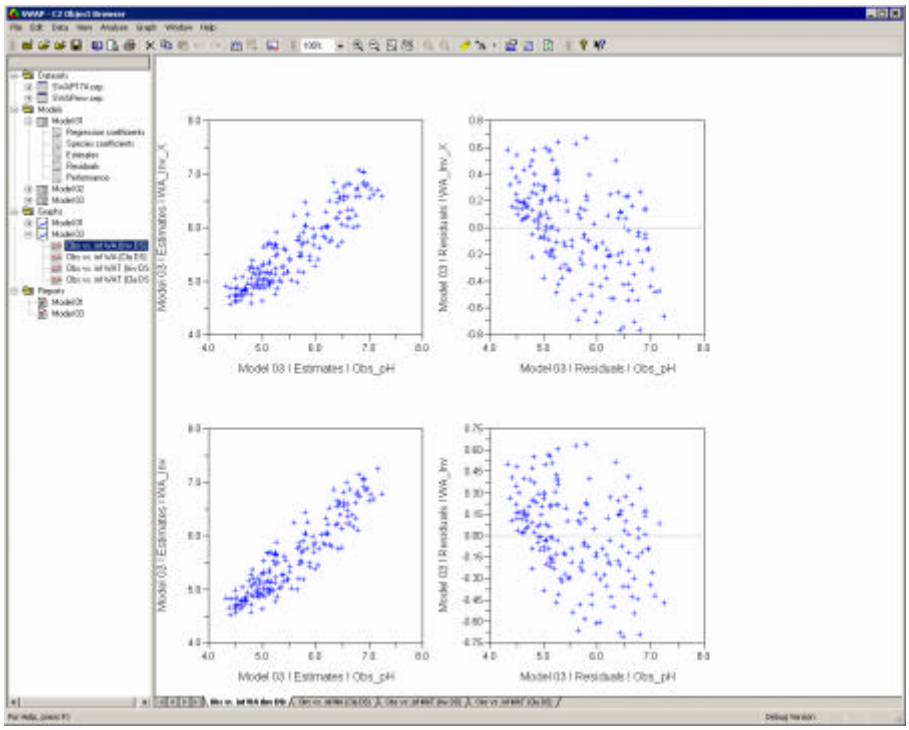
When you delete a model you are asked for confirmation and also get the chance to delete any auto-generated graphs (see below) or Reports that are associated with that model.

Model diagnostic plots

Diagnostic plots can be generated automatically for the training set analysis pressing the Plots button on the transfer function dialog box.



Select the desired plots on the Model graph options dialog box and click OK. When the model is run the requested plots will be generated and added to Graphs folder in the tree view:

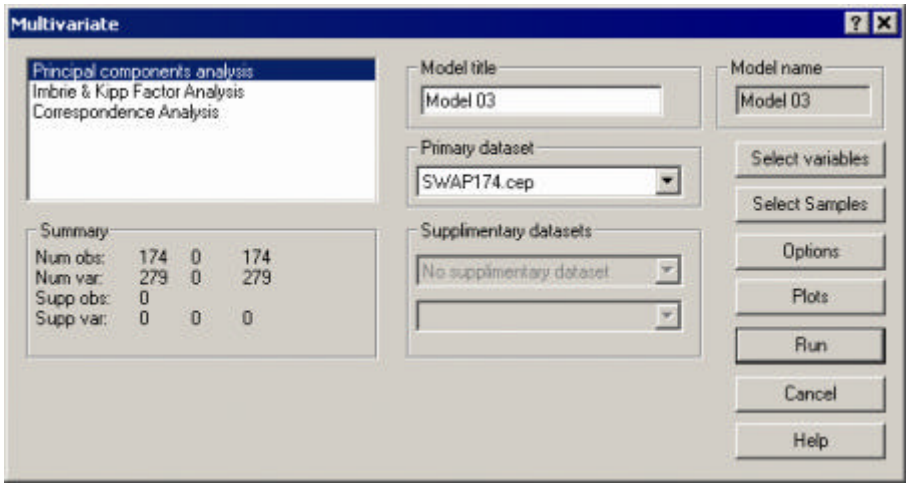


Diagnostic plots may also be created after a model has run by right-clicking on the model name in the tree-view and selecting Create graphs from the pop-up menu.

Multivariate analyses

C2 currently has three types of ordination methods currently grouped under the Multivariate heading . These are Principal components analysis, Imbrie and Kipp Factor analysis, and Correspondence analysis. Imbrie and Kipp Factor analysis is really a PCA with standardisation by samples and varimax rotation, but is included separately here for ease of use.

Select Multivariate from the Analyse menu.



The Multivariate dialog box is very similar in style and operation to the Transfer function dialog box. Choose the appropriate method, any options graphs, and press Run.

The resulting model output is added to the tree view and can be explored, listed and exported as normal.

Graphical tools

Graphics overview

C2 currently offers two types of graphics: x-y scatter / line plots or stratigraphic plots. Graphics are organised in graph books. A graph book contains one or more pages, and each page may contain one or more graphs. Also, each graph may contain one or more data series. Source data for graphs may be combined from different datasets in the workspace, making it easy to plot raw data alongside model results or to merge plots from different datasets into a single stratigraphic diagram.

This section describes x-y scatter and line plots. Stratigraphic diagrams are described in the next section.

Creating a new Graphbook

To create a new Graphbook containing a blank graphpage to the workspace by selecting "New graph book" from the Graphs menu or clicking the New graph book button on the main toolbar.

A new graph book will be added to the Explorer tree view:

Graph books are automatically named Graphbook 01, 02 etc. Graph pages are automatically named Page 1, Page 2 etc.

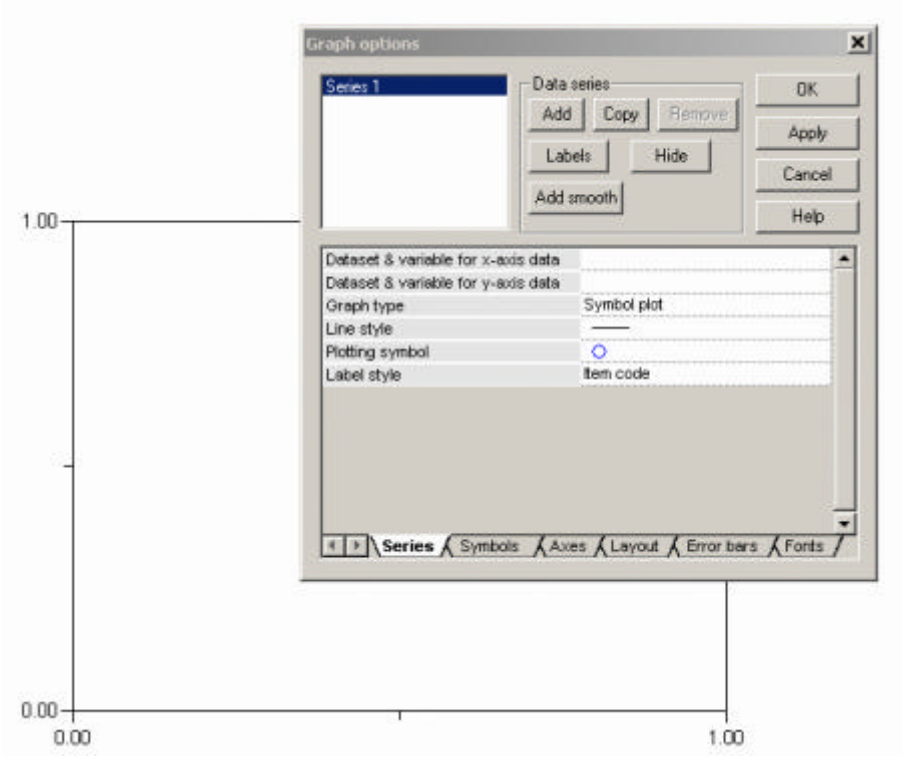
Creating an x-y plot

To create a new Graphbook containing an X-Y plot select **New graph book > XY Scatterplot** from the Graph menu or clicking the New graph book with scatterplot button on the main toolbar .

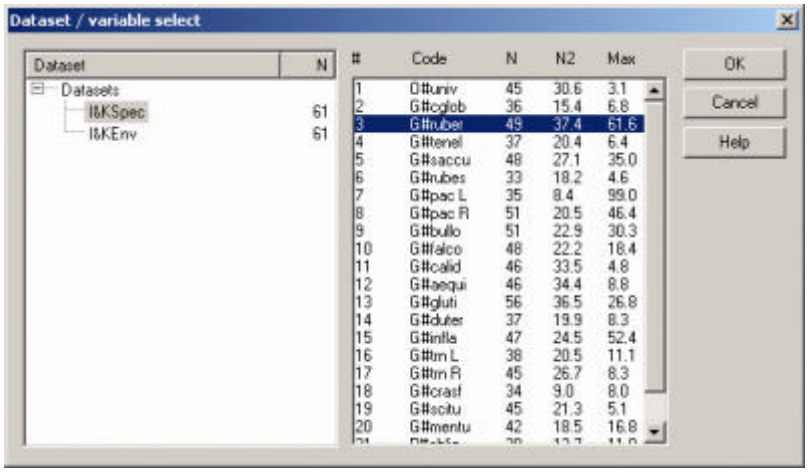
A new graph book will be added to the Explorer tree view:

Graph books are automatically named Graphbook 01, 02 etc. Graph pages are automatically named Page 1, Page 2 etc. The can be renamed using function ley F2 and given a description.

This will create an empty plot and display the Graph options dialog box:



To specify a variable for the x- or y-axis place the cursor in the field labelled Dataset & variable for x-axis (or y-axis) data. Click the Field Edit icon to display the "Dataset / variable select" dialog box:

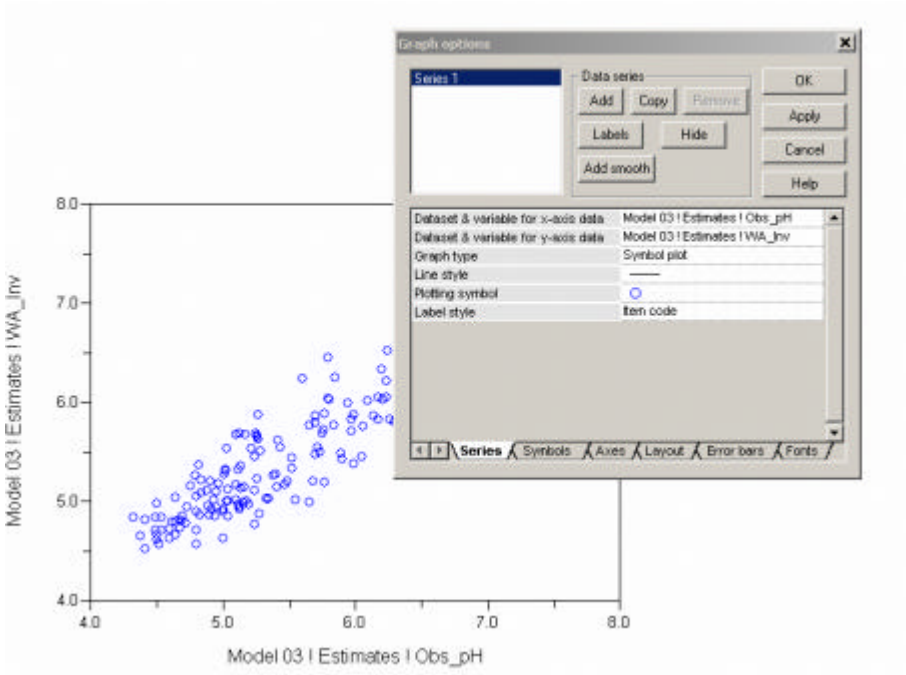


Navigate through datasets or model results and select a variable.

Click Apply or OK on the Graph options dialog box to display the graph:

Both x- and y-variables must have the same number of observations.

If either the x- or y- variable is left blank C2 will generate an index plot – that is, the sample number, or index, will be used for that variable in the plot.



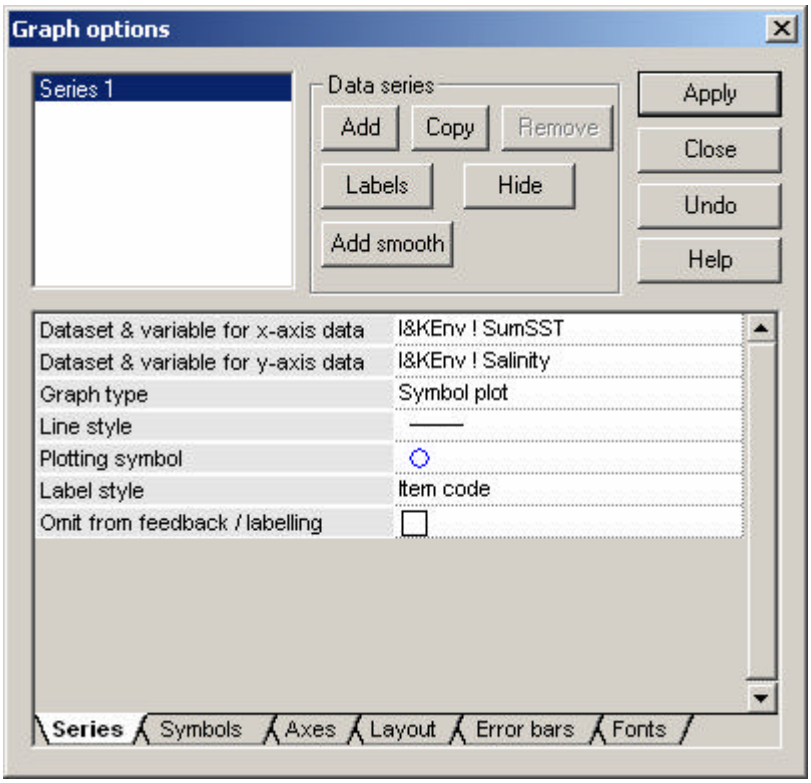
Modifying an x-y plot

Various properties of the x-y plot can be modified using the Graph options dialog box. To change a property of a graph click Edit graph button on the Graph toolbar , right-click the mouse over the graph and choose Edit from the popup menu, or double-click the mouse over the graph.

Different graph properties of are grouped into different pages in the Graph options dialog box. These are:

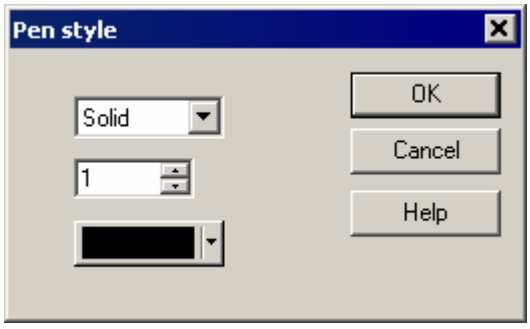
Series

The Series tab contains options for defining the data series used in the plot. That is, setting the x- and y-axis variables, the graph-type (line, symbol or line + symbol), the line and symbol styles, and the label style.

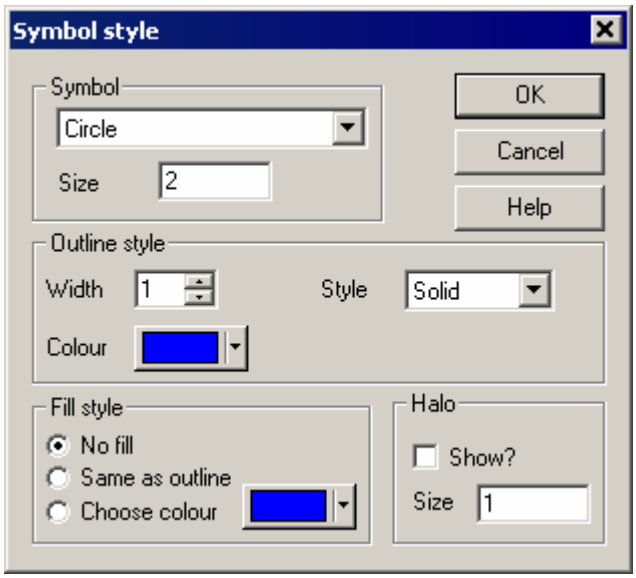


Changing line and symbol styles

Click the Field Edit icon in the Line style field to display the Line style dialog box:



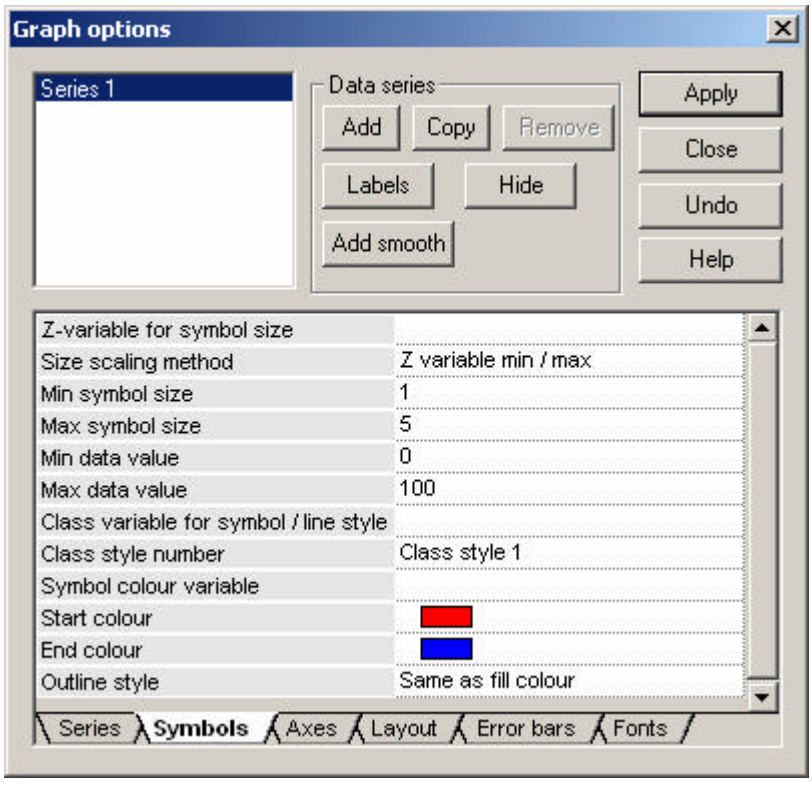
Click the Field Edit icon in the Symbol style field to display the Symbol style dialog box:



Halo draws a thin white line around the symbol. This is useful when there are multiple overlapping symbols with a solid fill style.

Symbols

The Symbols tab contains options for setting the symbol size and style used in bubble or coded scatter plots.



Symbols are, by default, plotted using the size defined by the symbol chosen on the Series tab (default = 2.5mm). The options here allow you to scale the size of a symbol according to the value of a 3rd variable (Z-variable).

There are two ways to scale the points:
Z variable min / max maps the symbol size linearly from the minimum to

maximum values of the Z variable to the values in Min symbol size to Max symbol size (in mm).

Constrained min / max maps the symbol size linearly from the Min data value to Max data value to the values in Min symbol size to Max symbol size (in mm).

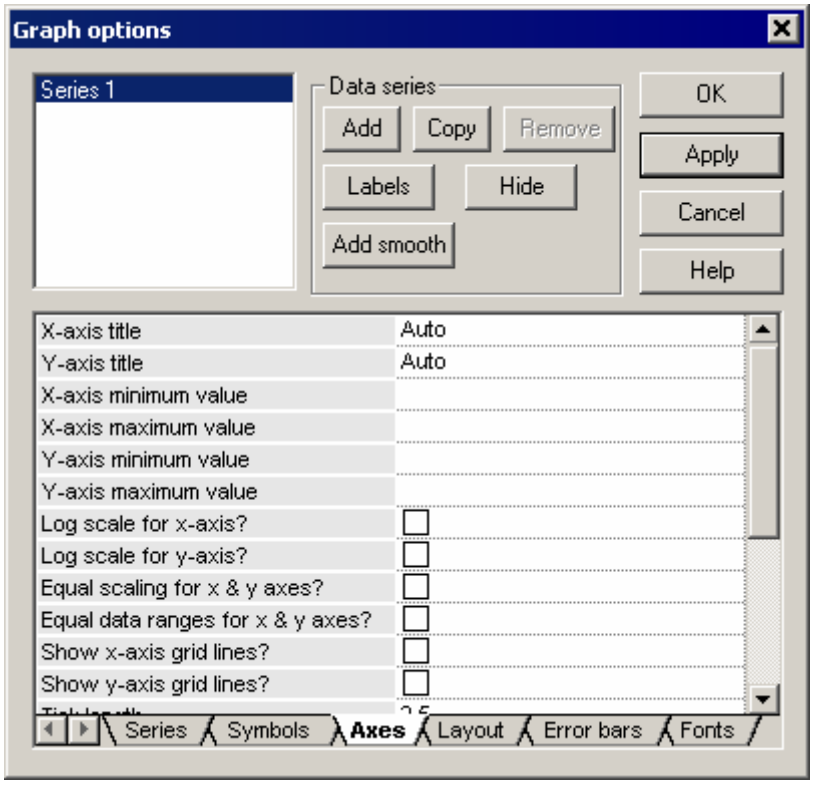
Z variable min / max scaling is useful for individual plots whereas Constrained min / max is useful for retaining a consistent scaling across multiple plots.

The class variable allows you to specify an integer variable that defines the class or group membership for each data point. Different groups of data are then plotted using different symbols.

The symbols used for plotting are defined in the global options. Up to 10 different symbol class styles can be defined. For example, class style 1 may define circles of different colour for colour plots, and Class style 2 may define different symbols of a single colour for B&W plots. The class style number field allows you to choose which symbol style to use for the current data series.

Variable for symbol / bar colour and start / end colours changes the colour of bars / symbols according to the value of a third quantitative variable. Given a variable name and start / end colour each point is coloured according to its position in the data range.

Axes

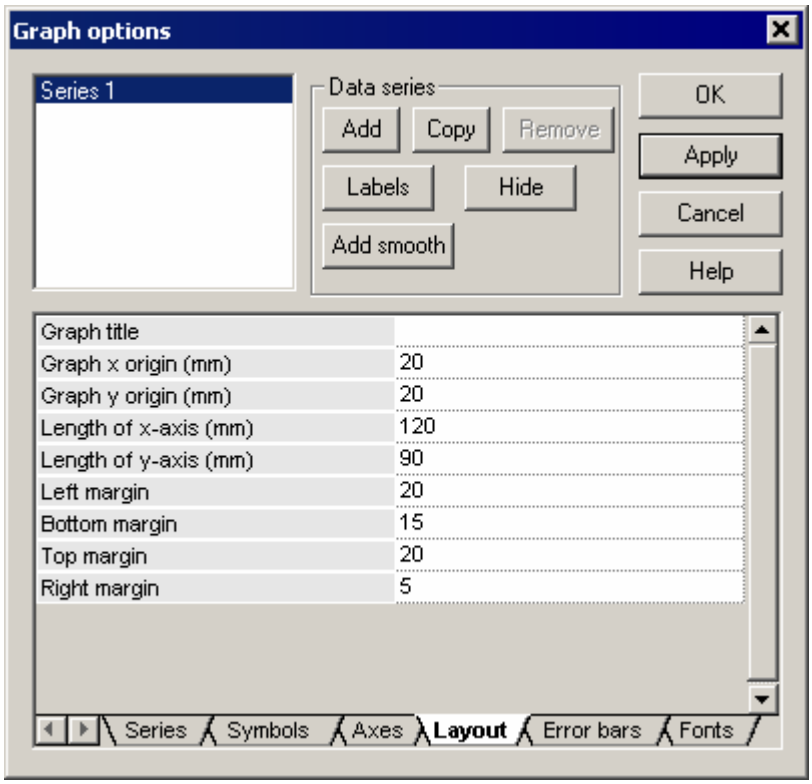


This tab allows you to:

Change the text used to label the axes,
Override the automatic calculation of axes limits,
Plot x- or y-axes on a log10 scale,
Scale the x- and y-axes equally (that is, one data unit on x = one data unit on y),
Force the x- and y-axes to have the same data limits (if the Equal scaling for x- and y-axis is also selected the length of the y-axis will be adjusted accordingly) ,
Display x- or y- grid lines,
Change the length of the axes tick marks,

Draw a horizontal line through the x- or y-axis origin (zero data value),
Reverse the scale of the x- or y-axis.

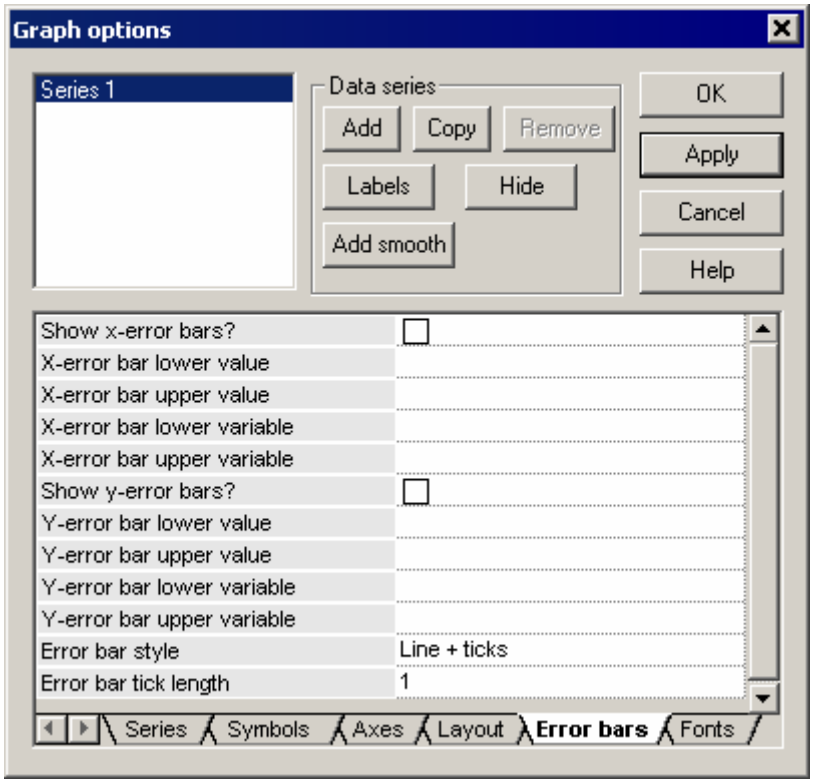
Layout



This tab controls the layout and positioning of the graph on the page.

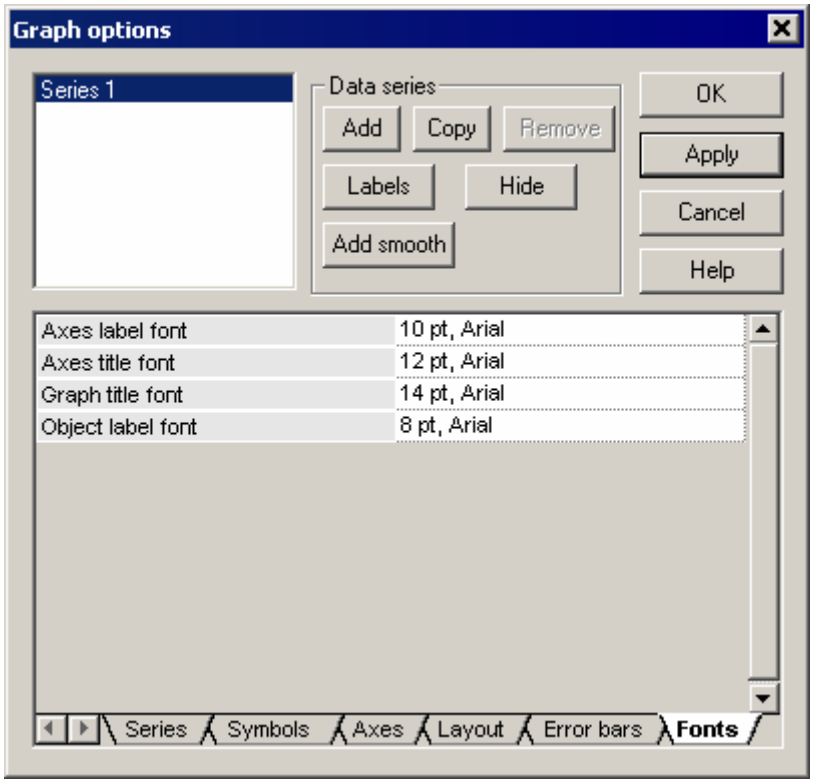
The graph origin is the location of the bottom left-hand corner of the x- and y-axes. The margins are the areas outside the axes containing ticks, and axes labels.

Error bars



X- and y- error bars can be added to a plot by either specifying a fixed value to be added / subtracted from each data point, or by specifying a variable whose value will be added / subtracted from each data point.

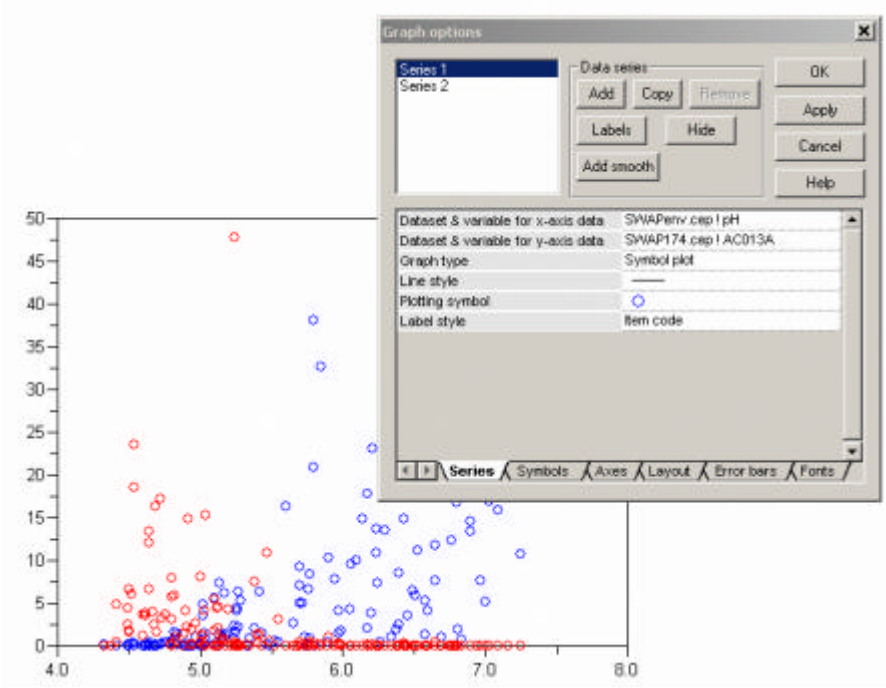
Fonts



Change the font face, style and size of the graph labels.

Plotting multiple datasets on the same graph

An individual graph can contain multiple datasets, each plotted in a different style. In C2, each plotted dataset is known as a **series.** To add additional data series simply click Add or Copy in the Graph options dialog box. Add will add a new blank series with default options. Copy will add a copy of the highlighted series. The figure below shows an x-y plot with two data series plotted with different symbols.



A data series can be deleted by highlighting it and clicking remove. Note: you cannot delete the first data series – each graph must contain at least one data series.

Zooming in on data in an x-y plot

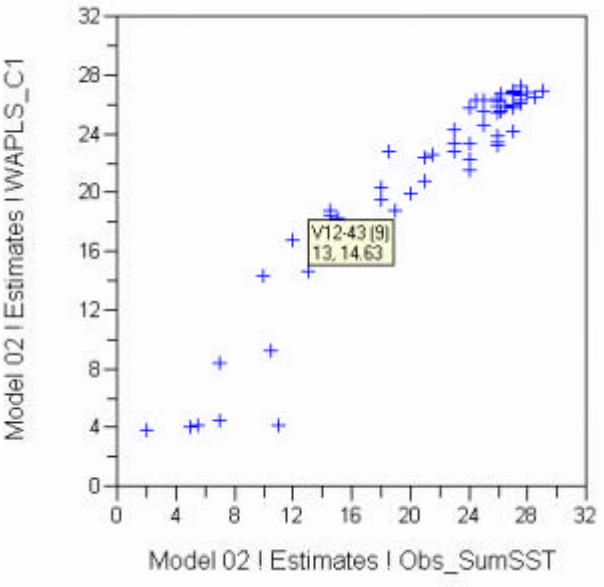
The data range in an x-y plot can be zoomed simply by dragging the mouse to select the zoomed range. Zooming is incremental and can be undone (ie. zoomed out) by clicking on the (incremental) Zoom out or Zoom out to original range buttons on the Graph toolbar.

Identifying and labelling data points in an x-y plot

C2 supports two types of interactive labelling, data tips and data labels. Both types of labelling carry a computational overhead and will slow down the redrawing of the graph on screen. For pages with a few graphs consisting of a few hundred data points this will not be noticeable but for multiple graphs with many thousands of points the delay may be significant. For this reason labelling is disabled for the page when it is created.

Interactive data tips

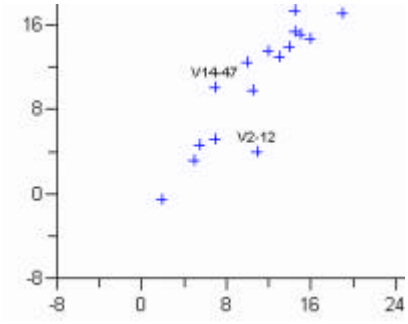
To identify a data point make sure the **Enable graph tips** button on the Graph toolbar is selected. Use the drop-down selector to show graph tips or on the program status bar. Hover the mouse over a data point and the sample code and x- and y- data values will appear on the graph or in the status bar:



Data tips are available for both x-y plots and stratigraphic diagrams.

Interactive data labels

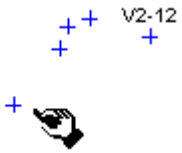
To permanently label a data point make sure the Enable graph labels button on the graph toolbar is selected. There are two labelling styles. The default is fixed position labels. Clicking on a point will simply place a label above the data point:



Click the point a second time to remove the label.

The second method is moveable labels. Choose this by clicking the down-arrow on the Enable graph bales button:

The Enable graph labels button with change to blue text. Now, when a data point is left-mouse clicked the cursor changes to a hand + pen:

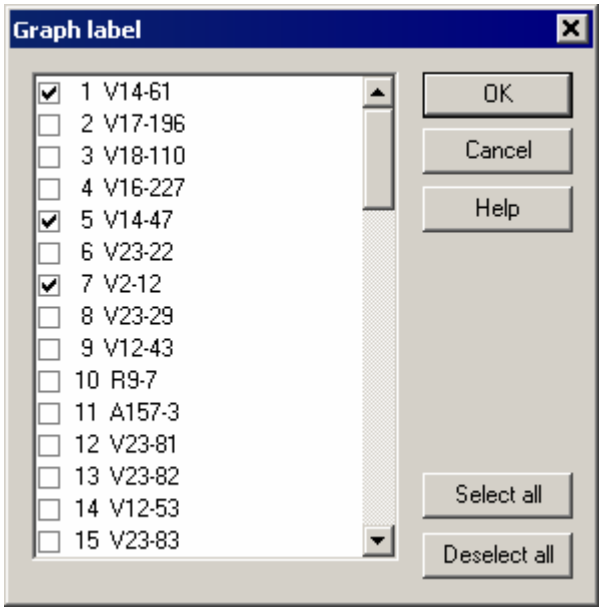


Move the mouse to where you want the label placed a left-click again. Click the data point again to remove the label.

Labels are drawn with the font specified in the Object label font field of the Fonts tab in the Graph option dialog box.

Labelling from a list

To find specific data points in a graph it is sometimes useful to select the point from a list. Click on the Label button on the Graph options dialog box and select the data points to be labelled from the check list:



If a symbol is specified for a point the label is drawn horizontally centred above the symbol. If the symbol type is set to No symbol the label is drawn horizontally and vertically centred over the data point.

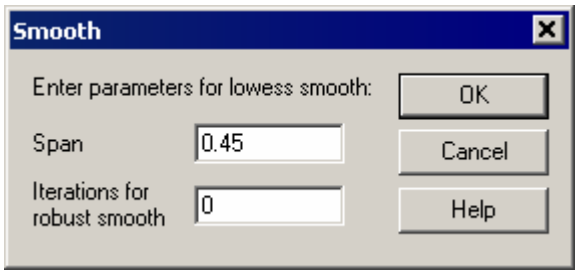
Hiding data points in a plot

To hide specific data points in a graph click on the the Hide button on the Graph options dialog box and select the data points to be hidden from the list.

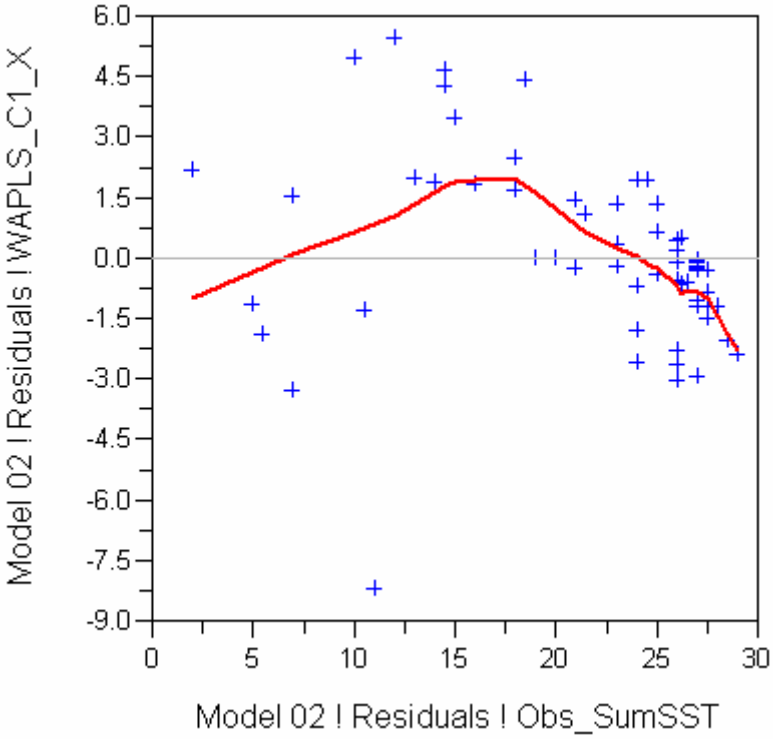
Adding a lowess smooth

A locally weighted scatterplot smooth (lowess) for a data series can added to the plot to identify or highlight the relationship between the x- and y-variables.

Click the Add smooth button on the Graph options dialog box:



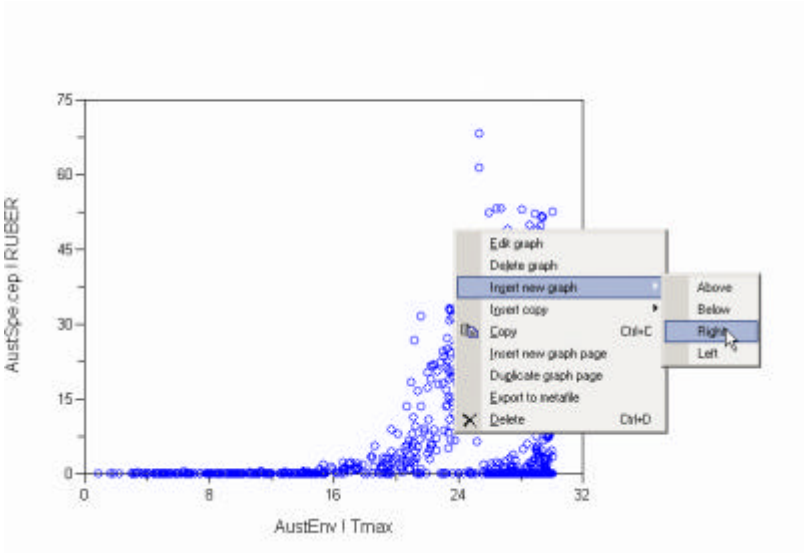
The two available options for the smooth are the span (the proportion of data points used in the local regression at each data point), and the number of iterations used in computing robust estimates of the local regressions.



The smooth is generated as an additional data series and can be deleted or modified using the functions in the Graph options dialog box.

Adding additional x-y plots to a graph page

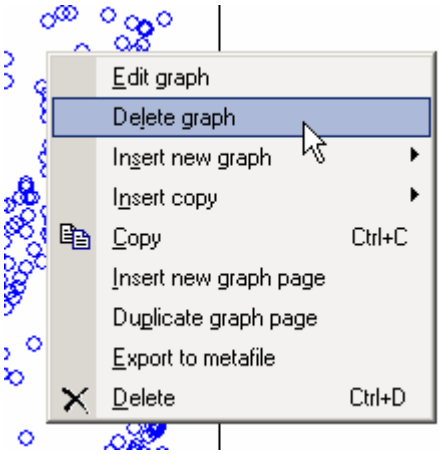
A graph page can contain multiple x-y plots. To add additional plots right click the mouse over an existing plot. The following popup menu will appear.



Choose Insert new graph to insert a new, empty plot to the right, left, above or below the existing plot. Choose Insert Copy to insert a copy of the existing graph.

Deleting an x-y plot from a graph page

To delete an x-y plot right click the mouse over the plot and select Delete graph from the popup menu.



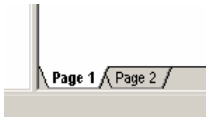
Undoing / redoing changes to x-y plot

Most modifications to a plot can be undone or redone by selecting undo / redo from the Edit menu or clicking on the undo / redo buttons on the main toolbar

Inserting a new graph page into a graph book

To insert a new, blank graph page into an existing graph book simply select **Insert new graph page** from the Graph menu or right click the mouse over an existing graph page and select **Insert new graph page** from the popup menu.

New graph pages are labelled Page 2, Page 3 etc., and can be selected by clicking on the tab at the bottom of the graph window:

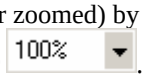


If you want to use an existing graph as a template or starting point for a new plot you can **Insert a duplicate**. This will create a copy of the selected graph page that can be further modified.

Deleting a graph page

To delete a graph page simply select the page and choose Delete from the Edit menu or right click the mouse over the page and select Delete from the popup menu.

Zooming the graph page

The plots on a graph page can be scaled (or zoomed) by choosing a scaling factor in the zoom box on the main toolbar 

. The selected zoom factor applies to the screen display and printed version of the plot. Fit window will scale the plot to fit the existing size of the graph window or printed page.

The plot may also be zoomed in or out using the zoom buttons , zoomed to full size (100%) by clicking the button, or zoomed to fit the current window by clicking the button.

Copying a graph page

A graph page can be copied to the clipboard as an enhance Windows metafile by selecting Copy from the Edit menu or clicking the Copy button on the main toolbar. The graph can then be pasted into Word, Excel or your favourite graphics program for further editing.

Note: you may need to experiment with Paste and Paste special and pasting as a Picture (Metatfile) or Enhanced Windows Metafile to achieve best results.

Export a graph page as a Windows metafile

A graph page can be saved to disk as an Enhanced Windows Metafile by selecting Export from the file menu or right–clicking over the graph and selecting Export to metafile from the pop-up menu. The metafile can then be imported into Word, Excel or your favourite graphics program for further editing.

Printing a graph page

A graph page can be printed or print-previewed in the normal way.

Note that only horizontal tiling is supported – any part of a graph extending above the top margin of the page will be truncated.

Graphs that extend beyond the right-hand page margin will be printed on multiple pages, with an overlap specified in the Global options.

Using the Fit to window zooming is useful to scale a large plot (or a graphpage containing multiple plots) to a single printed page.

Refreshing the view

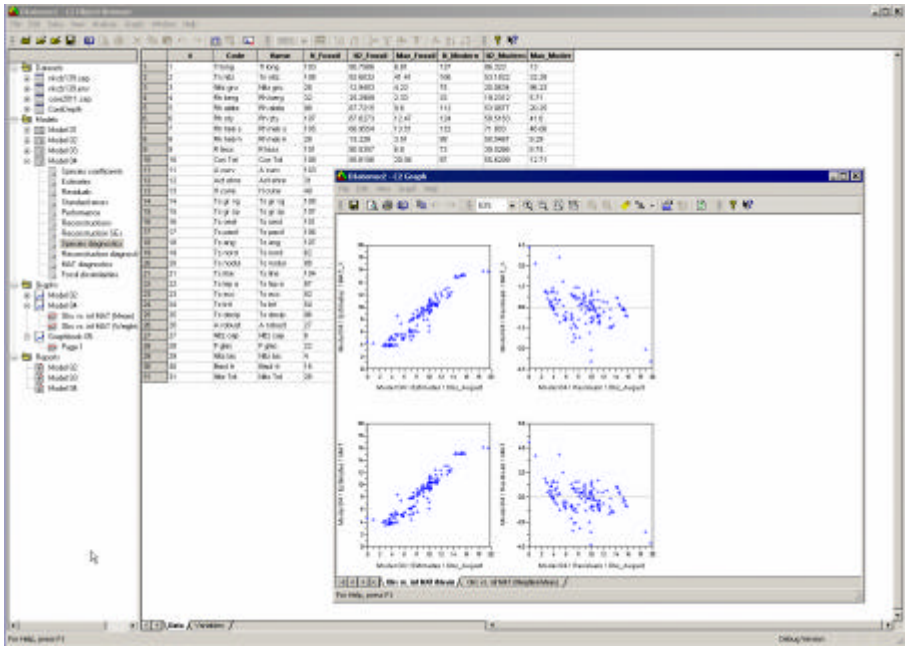
Sometimes changes are made to the diagram or underlying data and the graph is not automatically redrawn. To manually redraw the graph press F5 or click the Refresh button on the graph toolbar.

Opening the graph in a new window

It is sometime useful to open the graph in a separate window – e.g. when working with the graph and inspecting / editing the underlying data.

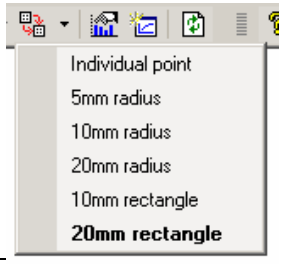
To open the graph in a separate window either click the button on the graph tool bar or right-click the graph book or graph page in the tree-view and select Open graph in a new window from the pop-up menu.

The new graph window has a restricted menu and tool bar limited to graph-editing functions. The window can be closed in the normal way, or will be automatically closed when you exit from C2.



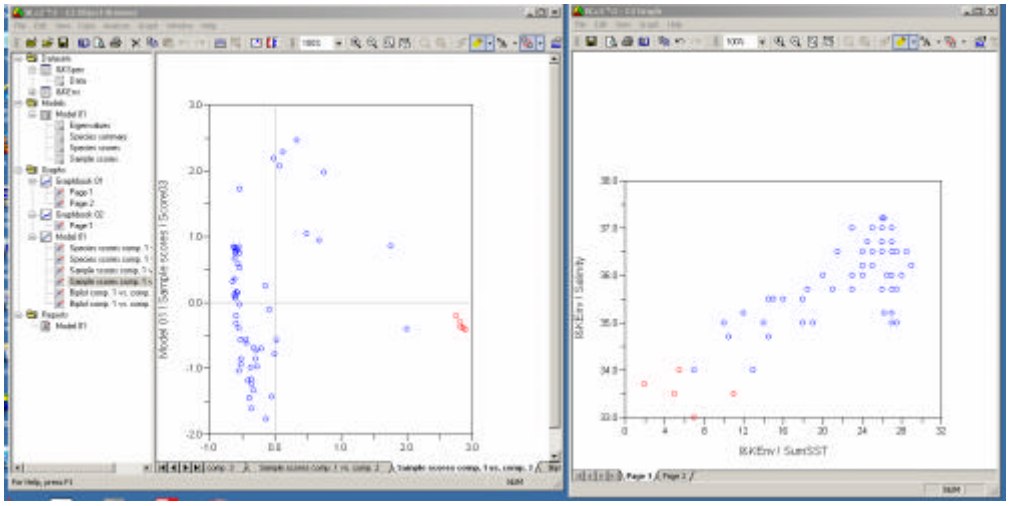
Brushing multiple graphs

Brushing allows you to identify points in multiple scatter or stratigraphic plots. Turn this function on / off using the “brush” button and choose the brushing radius from the drop down menu. Now just move the mouse over a graph to brush data points. Selected



data will plot red (default colour – can be changed in Edit menu -> options). Corresponding points in any other visible graph will also be highlighted (even if open in a different window). C2 links points in different graphs by row number. If the data in two graphs are unrelated they will still get highlighted. This can cause some problems – for example, if you have two series (e.g. species scores and samples scores from a CA overlain), C2 will highlight points from both series – which is not very useful. There is a new option on the first tab of the scatter plot options dialog to omit a series from triggering a hit – this should provide a work around for most situations).

Used sensibly brushing can be very helpful for diagnosing models. For example, in a transfer function, to diagnose outliers, make a symbol or bar "stratigraphic" plot of dominant taxa in the training set and use e.g. pH as the chronology. Open this is a separate window and selected a residual plot of the TF in the main window. Turn brushing on and brush the outliers – the corresponding samples in the stratigraphic plot will be highlighted, identifying any "unusual" assemblages associated with the outliers.



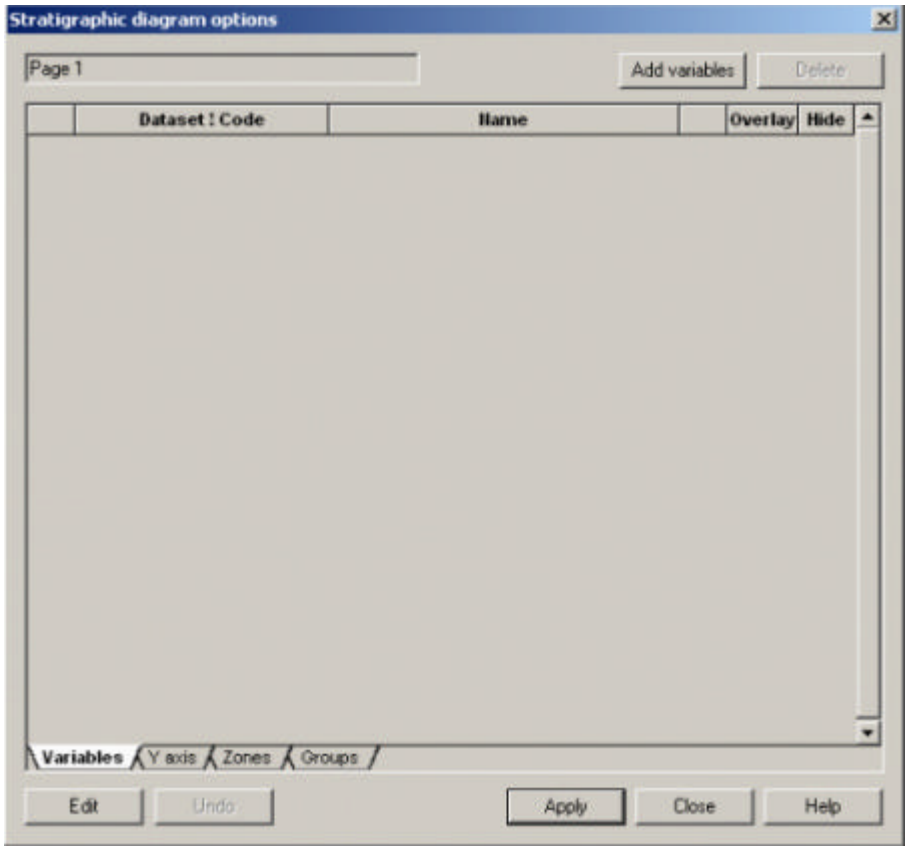
Stratigraphic diagrams

Creating a stratigraphic diagram

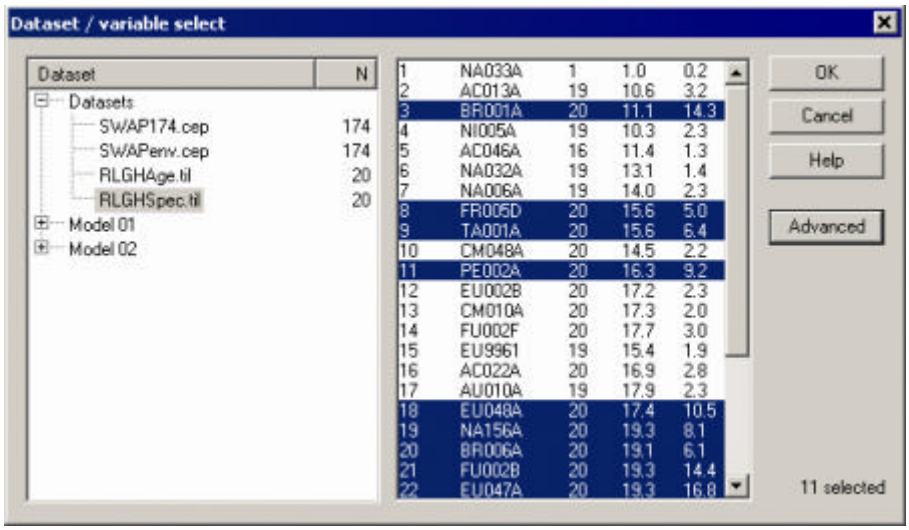
Please read the previous section on creating a graph before attempting to create a stratigraphic diagram.

To create a graph book containing a stratigraphic diagram simply click the **New graph book with stratigraphic diagram** or select **Insert new graph book > Stratigraphic diagram** from the Graph menu.

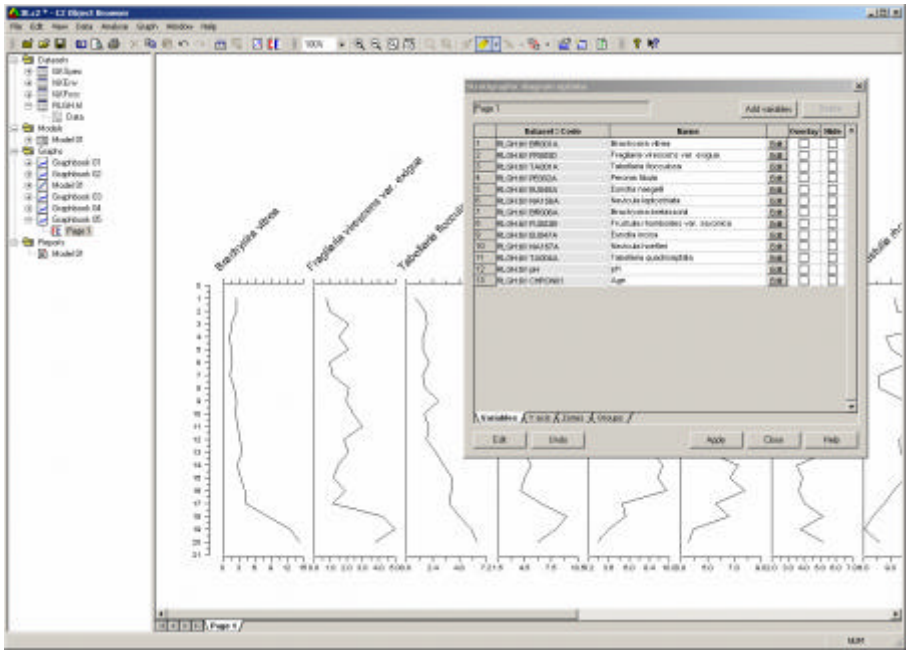
This will create an empty plot and display the Stratigraphic diagram options dialog box:



The next step is to add variables to the diagram. Click the Add variables button and select a dataset and list of variables from the Dataset / select variables dialog box:



You can select multiple variables by control-clicking (ie. holding down the Ctrl key while clicking). The Advanced button also allows selection of multiple variables using N, N2, and Max criteria. Click OK and the selected variables will be added to the Stratigraphic diagram options dialog box. Click Apply to update the diagram:



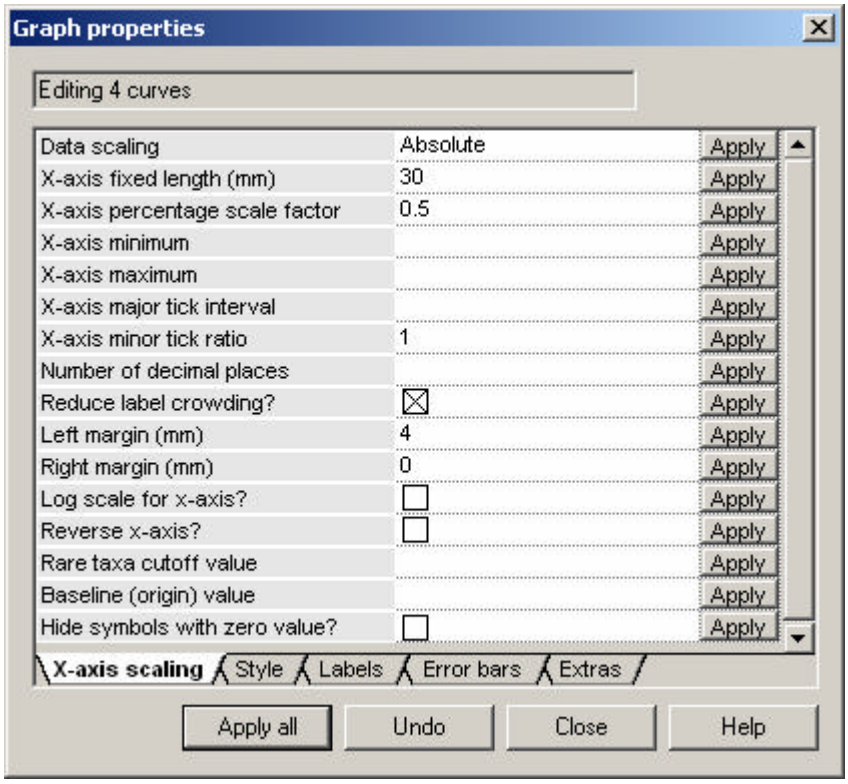
Modifying the appearance of individual graphs

The properties of a single graph can be modified by clicking the edit button next to the variable name. To change the properties of more than one curve first select the rows to be modified in the list of variables (by clicking on the row number) and click the Edit button at the bottom left of dialog box. Either action will display the Graph properties dialog box.

Alternatively you can double-click the left mouse button over an individual graph to invoke the Graph properties dialog box for that graph.

The various modifiable properties are listed under four tabs. Clicking Apply changes the setting for that property, clicking Apply all changes the settings for all the properties on that page.

X-axis scaling



Data scaling: Absolute sets the length of the x-axis of the selected graphs to be the value specified in the x-axis fixed length field (in mm). Percentage scales the length of the x-axis using the value in the x-axis percentage scale factor field multiplied by the length of the x-axis in data units. For example, if the graph plots data that ranges from 0 to 50 percent, and the scale factor is 0.5, the x-axis will be set to 50 * 0.5 = 25mm.

X-axis minimum, maximum, tick interval, tick ratio, and number of decimal places override the automatic calculation of these values. Leave blank for auto calculation.

Reduce label crowding avoids overlapping labels.

The left and right margins set the margins which are used to calculate the spacing between graphs.

Log scale for x-axis sets the x-axis to be plotted using a log10 scale.

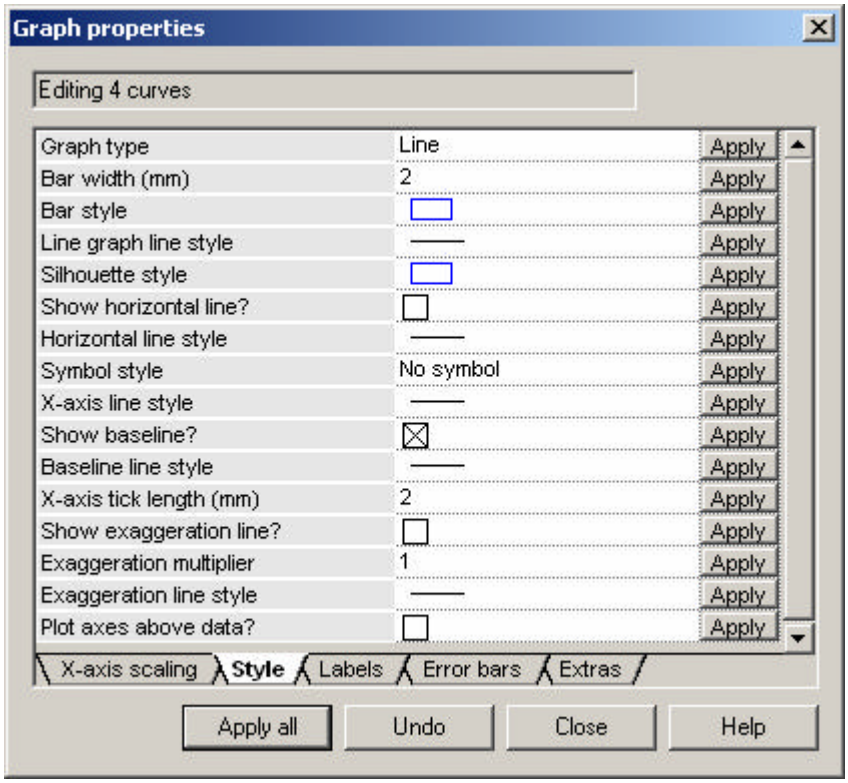
Reverse x-axis does just that.

Rare taxa cut-off value displays a symbol for values lower than the cut-off, rather than a bar.

Baseline (origin) value. If this field is blank the baseline is drawn as the left-hand y-axis of the graph. If a value is entered that lies between the minimum and maximum of the x-axis it is used to position the baseline. Bars, silhouettes and horizontal lines are drawn as deviations from this baseline.

Hide symbols with zero values does just that.

Style

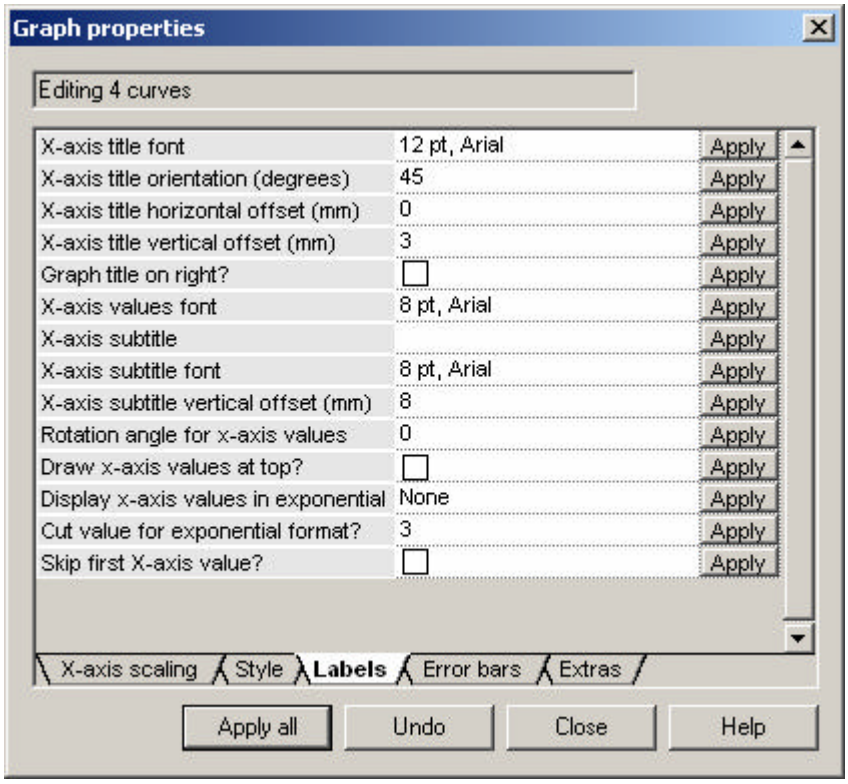


Graph type: Choose from Bar, Line, Silhouette, Symbol and Numeric.

Most of the options on this page are self explanatory. If a symbol is set in the Symbol style field it will be overlain on silhouette and line graphs.

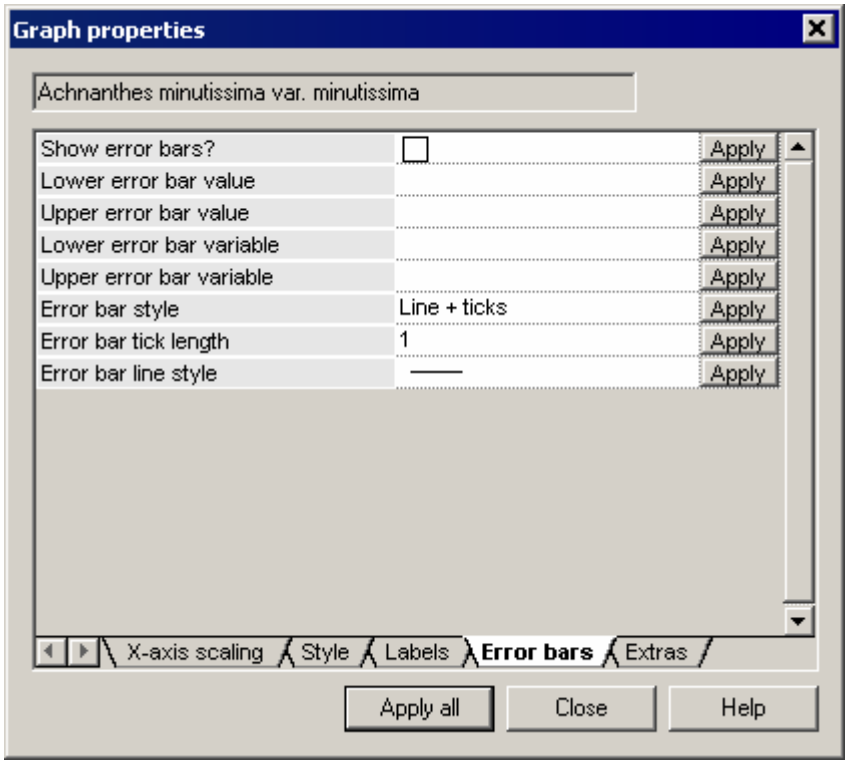
Show horizontal line draws a horizontal line from the y-axis to the data point in silhouette, line and symbol graphs.

Labels



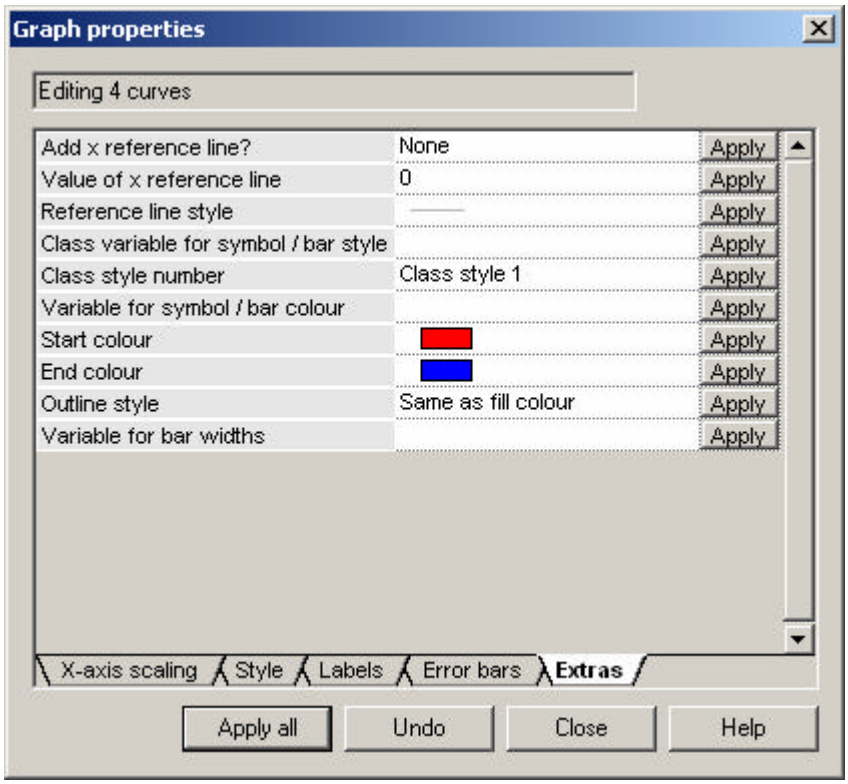
Specify an x-axis sub-title drawn under each plot and change the font face, size and styles used to label the graphs. X-axis title offset gives the label offset from the top of the y-axis.

Error bars



Supply either a fixed number or a variable for the upper / lower error bars to be added / subtracted from each data point.

Extras



Add a reference line (vertical line) to the plot at the specified x-value.

Class variable / class style changes the colour of bars or colour / shape of symbols in a curve according to a 3rd, class (integer) variable.

Variable for symbol / bar colour and start / end colours changes the colour of bars / symbols according to the value of a third quantitative variable. Given a variable name and start / end colour each point is coloured according to its position in the data range.

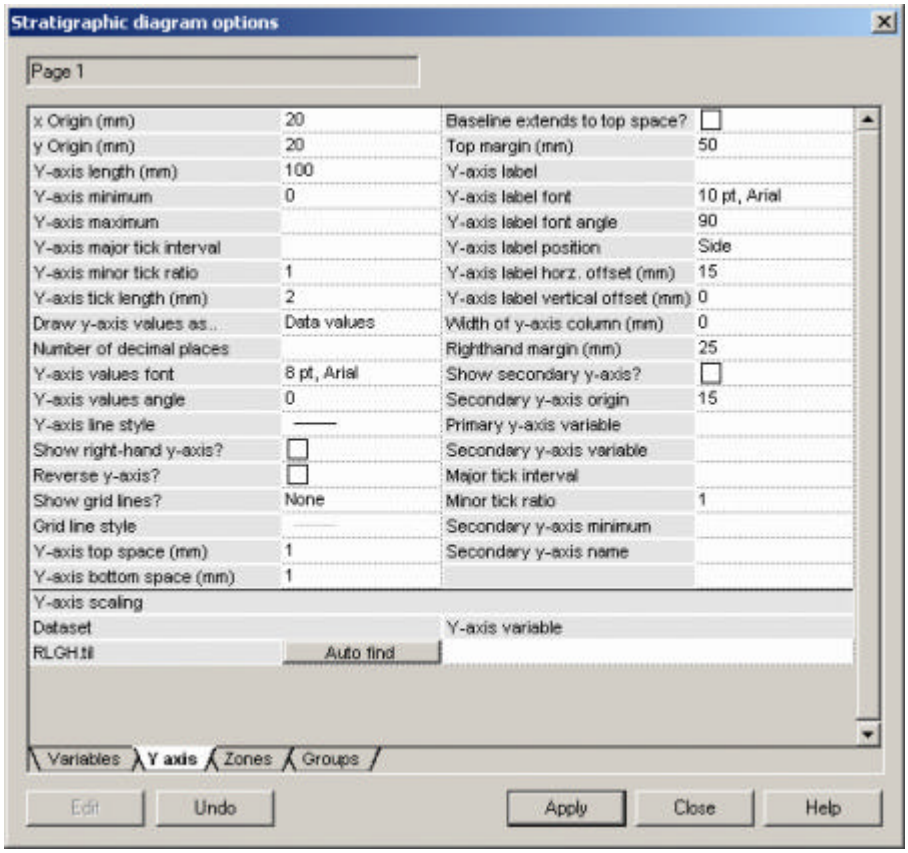
Variable for bar widths changes the width (thickness) of bars in stratigraphic plots. Each sample can be sized to represent the depth or time it spans. To use you need to insert into the dataset a variable that contains the width or thickness of each bar, in the same units as you will use for the y-axis.

Copying graph formats

Graphs for each variable can be individually edited via the Graph properties dialog box but in some cases it is easier to copy the formats of an adjacent graph. This function works in a similar way to the format painter in Word and Excel. Select a graph whose format you want to copy and click the new Copy Format button . Then just click other graphs to copy the format. Click outside the graph area to end. Double-click the Copy Format button to select which attributes of the curve are copied (graph style, axis scaling etc.). The default is to copy all attributes.

Changing the y-axis style and position

Various properties of the y-axis can be changed by selecting the y-axis tab in the Stratigraphic diagram options dialog box:



X and y origin sets the position of the bottom left corner of the y-axis relative to the bottom left corner of the window or printed page. Y-offset is the horizontal distance of the y-axis label from the y-axis.

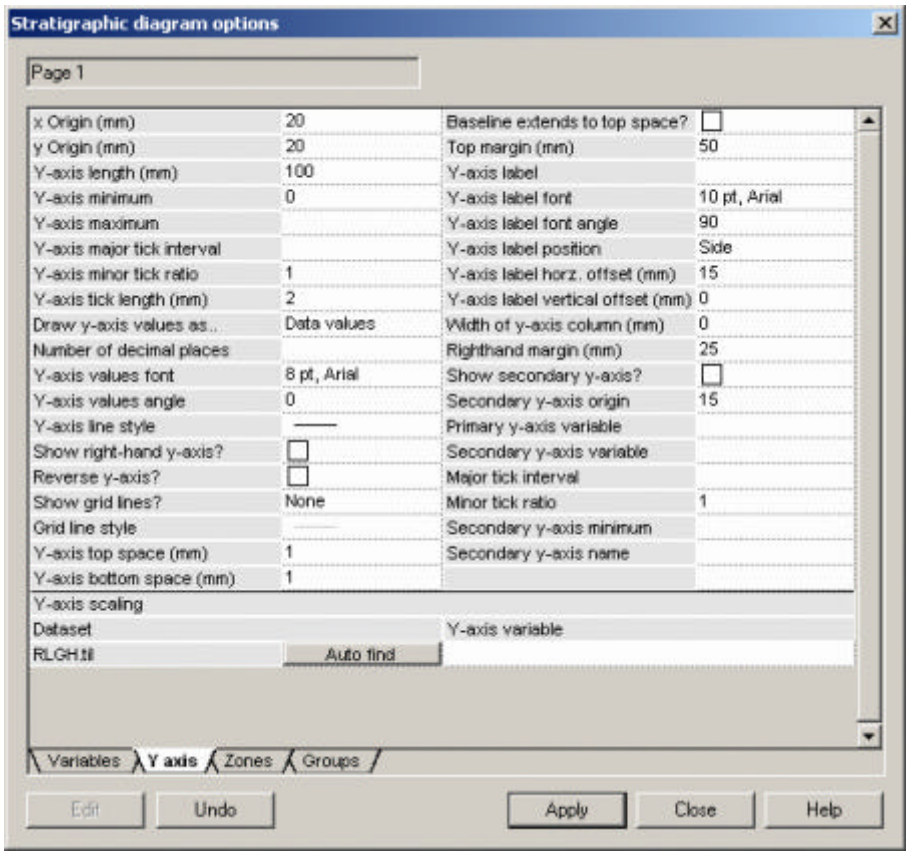
The y-axis can now be drawn as a column. There is an option on the y-axis properties to set the width of the column. Zero draws a normal y-axis, any other number draws a column of that width. You will have to find the best width to fit the labels by trial and error but 12mm is a good starting point.

Y-axis labels can be plotted as values (default) or sample codes / names using a new option on the y-axis properties. Note: this only works when data points are plotted against sample numbers on the y-axis (ie. the default). It will not label items correctly if a chronology variable is given for the y-axis.

The rest of the options are self explanatory.

Setting the y-axis variable

By default, each variable in the diagram is plotted against sample number on the y-axis. To change the y-axis variable select the y-axis tab on the Stratigraphic diagram options dialog box:

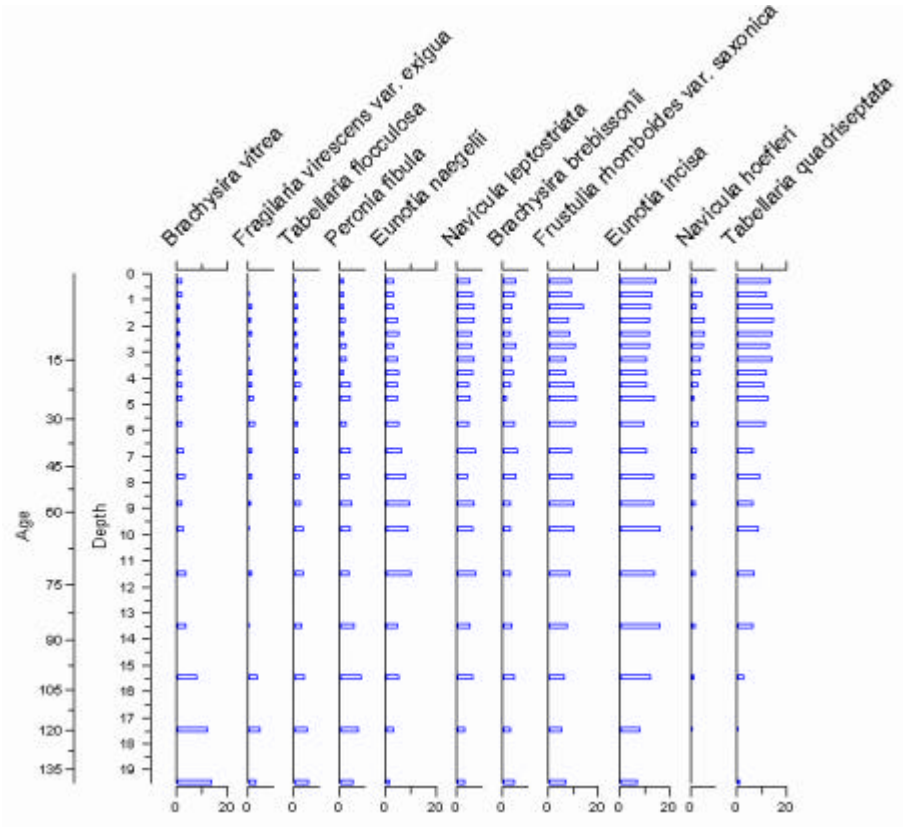


An entry appears at the bottom under Y-axis scaling for each dataset that is included in the diagram. Click in the right-hand field and select a dataset / variable to be used as the y-axis for all graphs based on the listed dataset. Alternatively click the Auto find button to give a list of any variables starting with Age.., Chron.. or Depth, that matches the dataset.

This process must be repeated for each listed dataset. Different y-axis variables can then be specified for different datasets.

Adding a secondary y-axis

To add a secondary axis select the y-axis tab in the Stratigraphic diagram options dialog box and select the variables used for the primary and second axes.



The above diagram has a depth primary axis and a secondary axis showing samples ages. Try adjusting the x-origin and label offsets to produce an optimal layout.

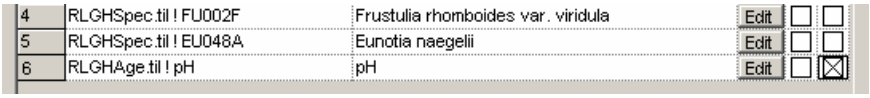
Adding variables from a new dataset to the diagram

Simply click the Add variables button and select the variables to be added. Remember to set a variable for the y-axis for the new dataset.

Removing variables from the diagram

Variables can be deleted from the diagram by selecting the row for that variable (by clicking on the row number) and pressing the Delete button.

Alternatively variables can be temporarily hidden by checking the Hide checkbox in the variables list:

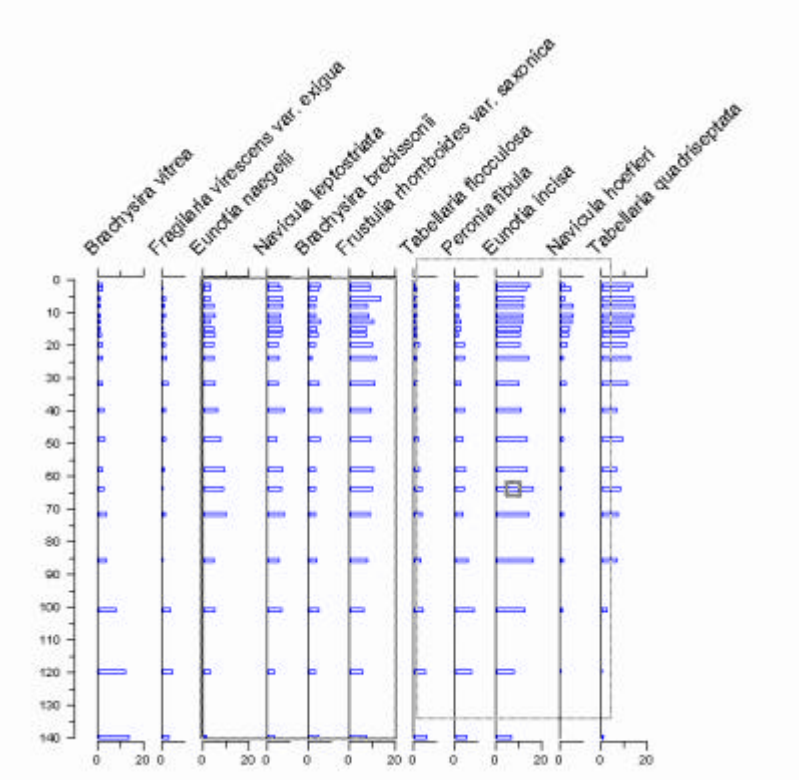


Changing the order of variables in the diagram

Variables are plotted in the order they appear in the variables list. To change the order simply select the rows to be changed and drag them to their new position:

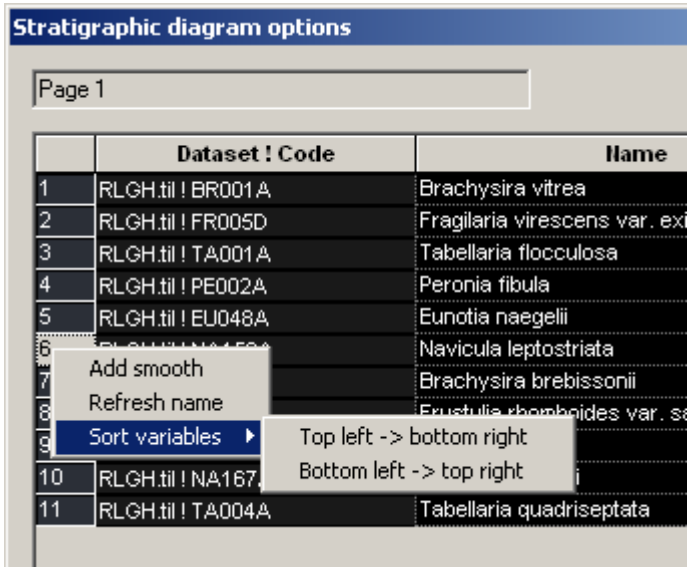
	Dataset ! Code	Name			
1	RLGHSpec.til ! NA032A	Navicula cocconeiformis	Edit	☐	☐
2	RLGHSpec.til ! FU002F	Frustulia rhomboides var. viridula	Edit	☐	☐
3	RLGHSpec.til ! EU048A	Eunotia naegelii	Edit	☐	☐
4	RLGHAge.til ! pH	pH	Edit	☐	☒
5	RLGHSpec.til ! TA001A	Tabellaria flocculosa	Edit	☐	☐
6	RLGHSpec.til ! EU002B	Eunotia pectinalis var. minor	Edit	☐	☐

It is also possible to drag and drop individual graphs to change their order in a stratigraphic diagram. Select one or more graphs by clicking (shift click for range) or dragging a box to intersect the desired graphs. Selected graphs will now have a shaded border. Now click and drag the selected graphs. The cursor will change shape according to drop zone. Drag to the top of the window to delete from diagram. Drag to bottom to hide. Drag over another graph to move selected graph(s) in front of selected. Drag and drop over upper 1/3 of another graph to overlay.



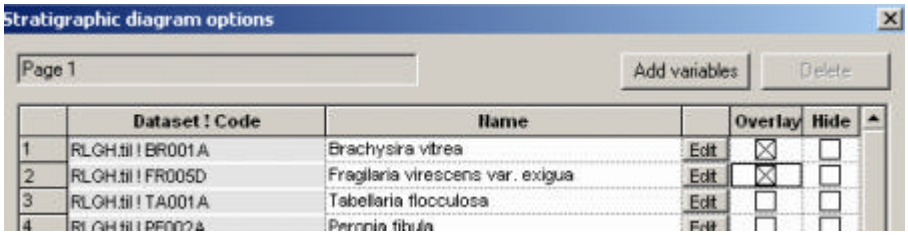
You can also select a range of graphs in a stratigraphic diagram and right click the mouse to edit the selected graphs.

It is also possible to automatically sort variables in a stratigraphic plot according to their "optima" on the y-axis. This provides a rough rearrangement of taxa from those more abundant at the top of the core on the left to those most abundant at the base of the core on the right. Simply select the variables to be sorted, right click over the names and choose the sort direction.



Overlaying variables in a graph

To overlay a curve on the previous graph simply check the Overlay button:

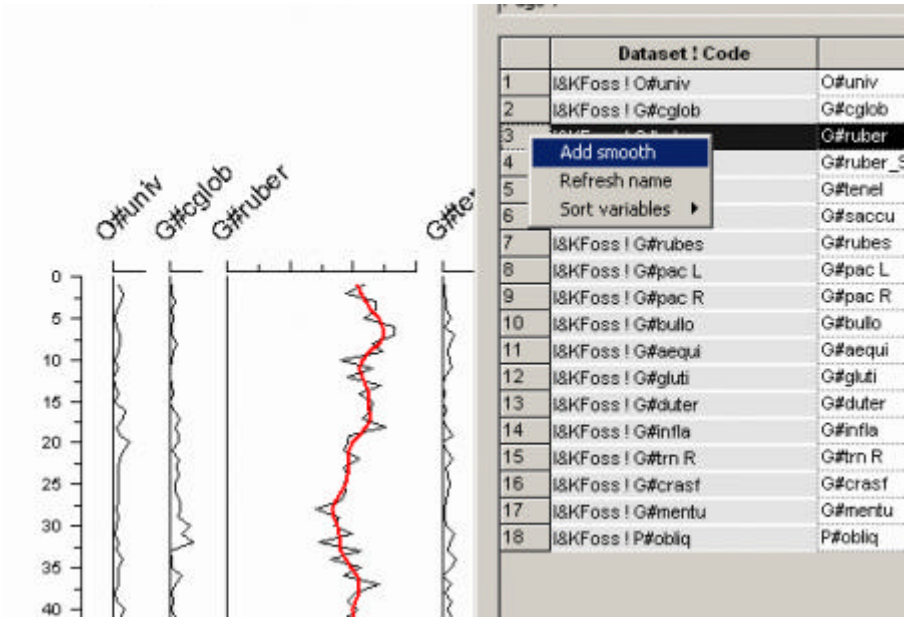


Note: The x-axis label is only shown for the first variable in the plot.

Adding a lowess smooth

A locally weighted scatterplot smooth (lowess) for a data series can be added to an individual plot to highlight temporal rends.

Select the variable(s) to be smoothed in the Stratigraphic options dialog, right click over the name and select Add amooth:

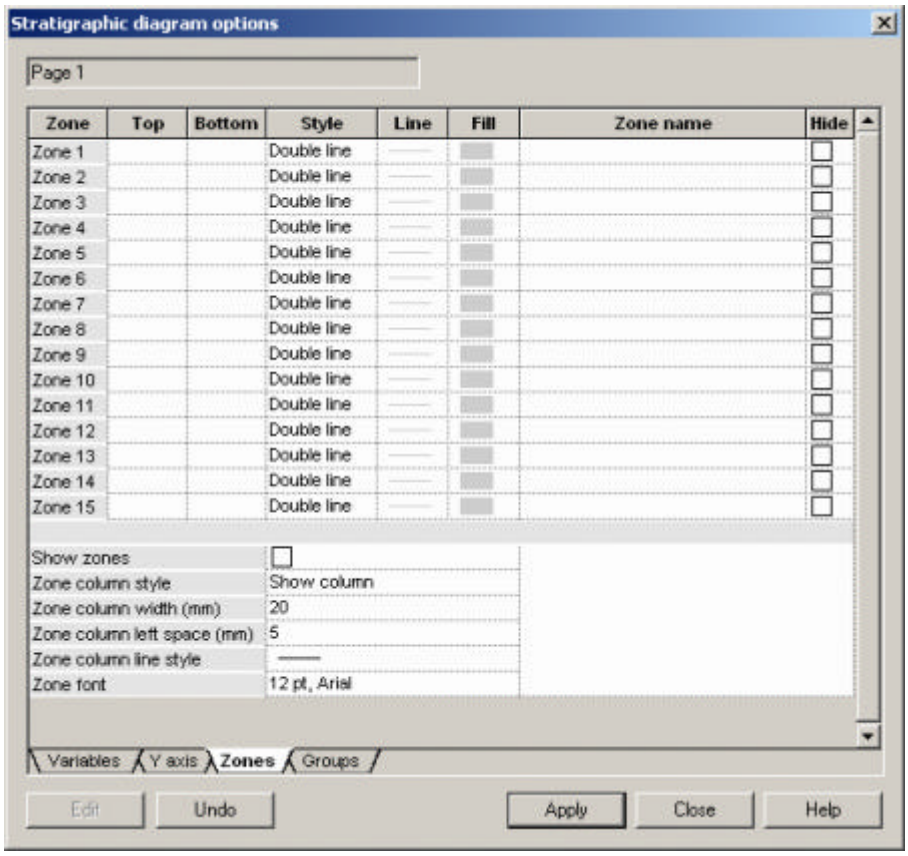


Smoothes are added to each selected variable as overlain curves. You can remove the overlay, delete, or otherwise modify the curve as with any other data series.

Note: The x and y data for the smooth are stored with the graph – the y is simply a copy of the y-axis variable currently selected for that graph. If you change the y-axis scaling you must also re-create the smooth. This is easily done by selecting the smooth in the Stratigraphic options dialog box and clicking Add smooth. If the selected variable is a smooth it will be re-created using the currently selected y-axis variable.

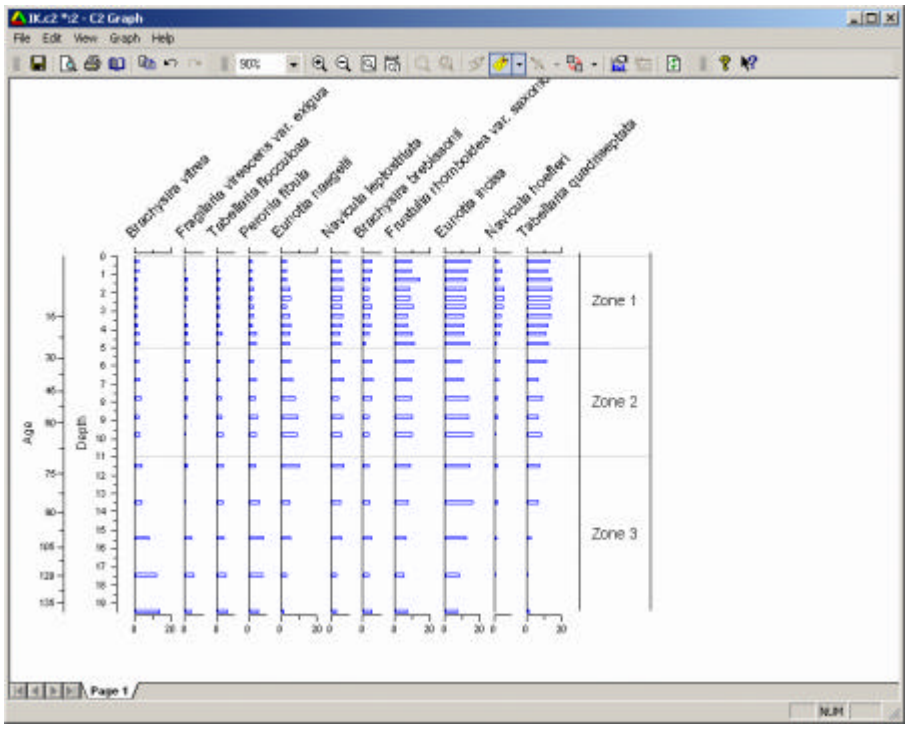
Adding zones

Select the Zones tab on the Stratigraphic diagram options dialog box.



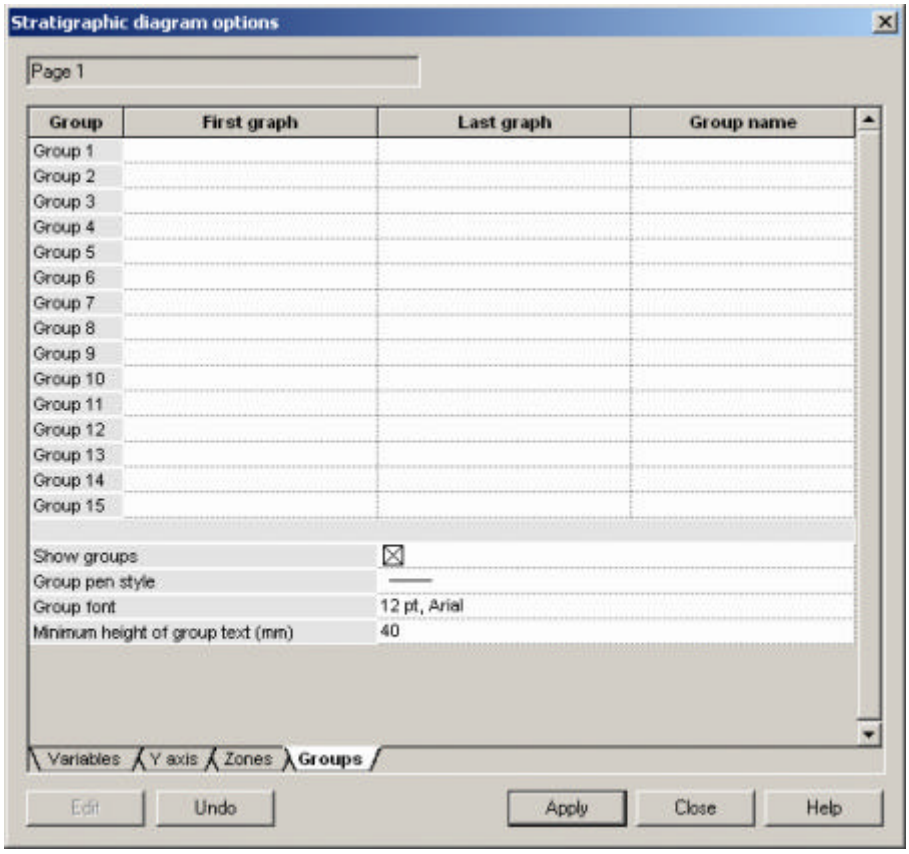
The box has seven columns of data for each zone. The first two are depths (ages etc.) for the zone top / bottom. The third column sets the zone style – single line drawn at the “top” value, lines at top and bottom values, or a solid fill rectangle. The forth and fifth set the line and fill style for the zone. The sixth column gives the zone name, and the seventh allows you to hide a zone. The zone name is plotted in a column on the right of the diagram if the Zone column style is “Show column").

Other options change the zone column width, line style and font.



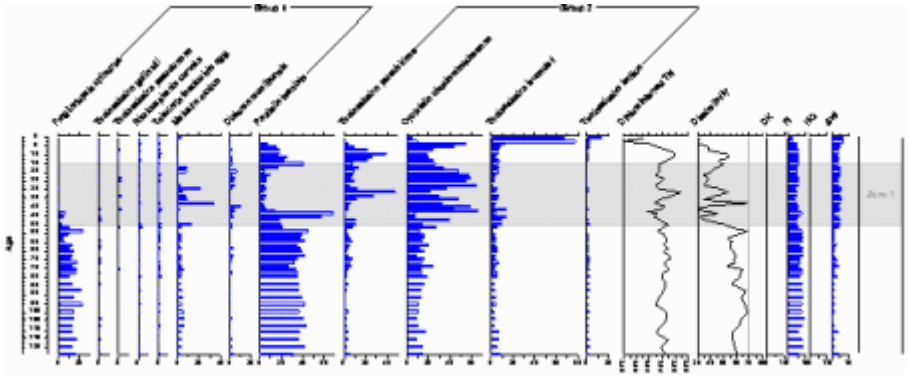
Adding groups

Select the Groups tab on the Stratigraphic diagram options dialog box.



For each simply group specify the start and end variables (graphs) and group name.

The plot below shows 2 groups, one solid fill zone and a column-style y-axis.



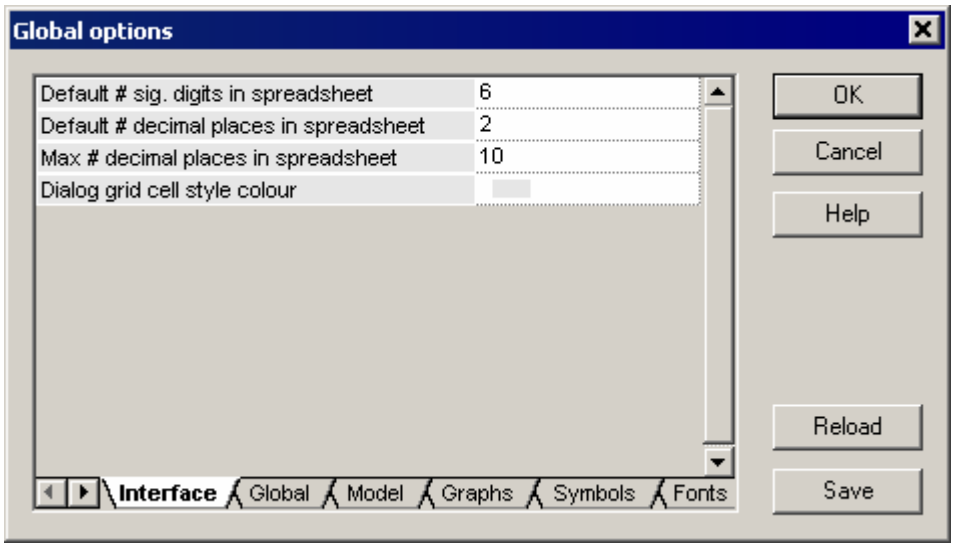
Global settings and defaults

Global settings

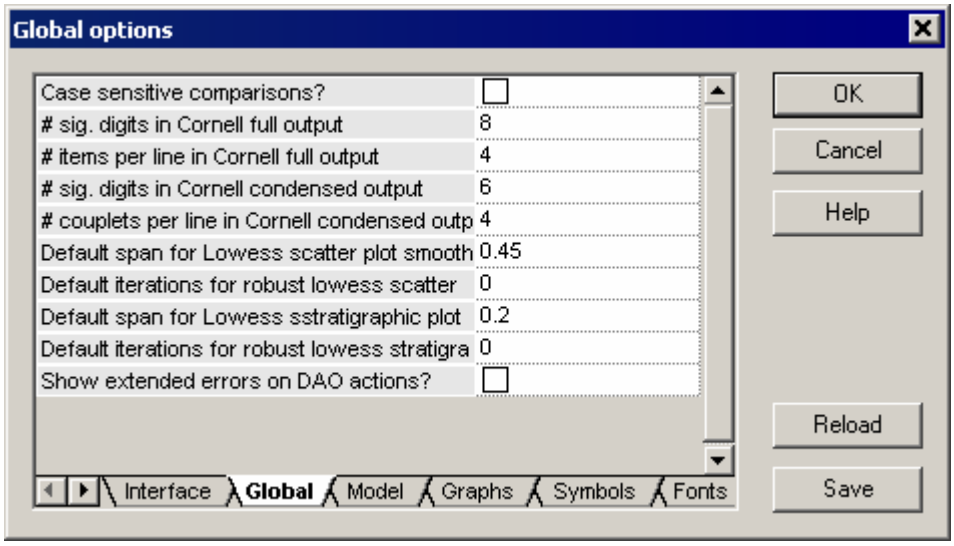
The default values of many of the parameters used to control the operation of C2 can be modified by the user by selecting Options from the Edit menu.

The six tabs in the Global options dialog box change the settings for:

User interface



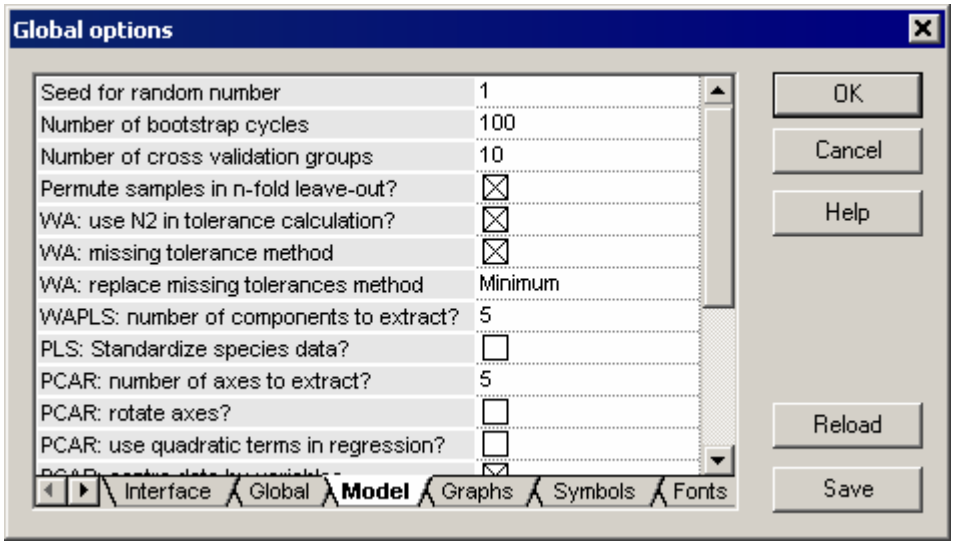
Global settings



If case sensitive comparisons is checked then comparisons between sample or variable codes when matching datasets or looking for duplicates is case sensitive.

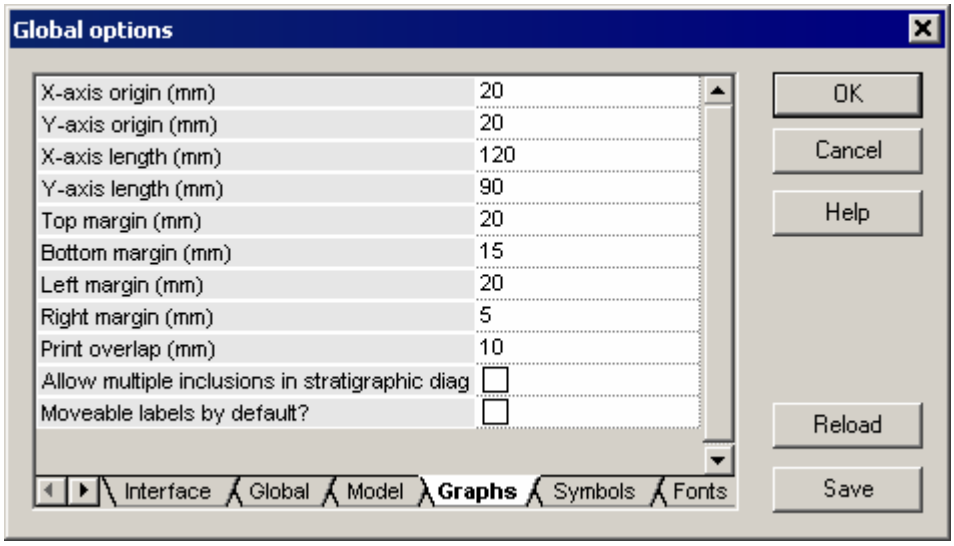
The Show extended errors… function reports extended errors if problems arise importing Excel and other spreadsheet / database files.

Model settings



Sets the defaults used in the various transfer function models.

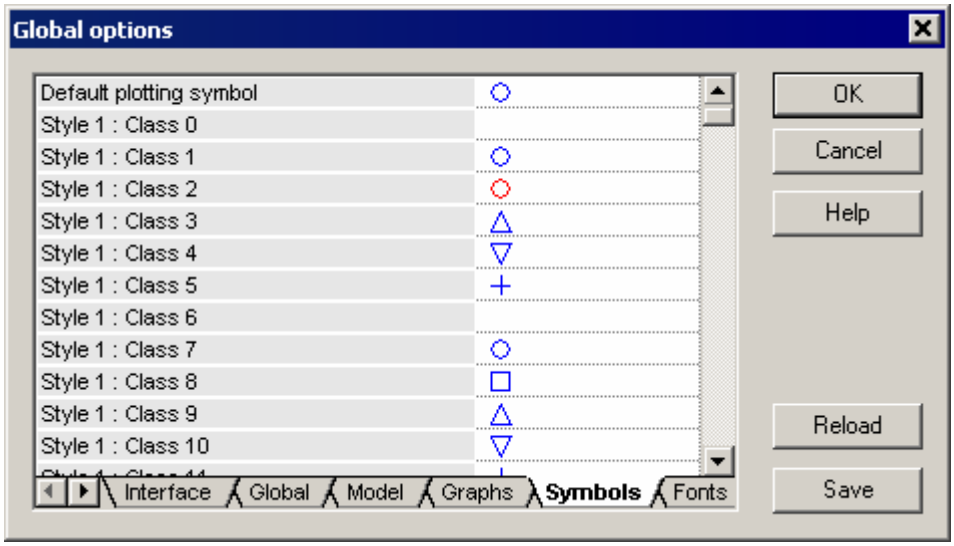
Graph settings



Sets the defaults for new graphs.

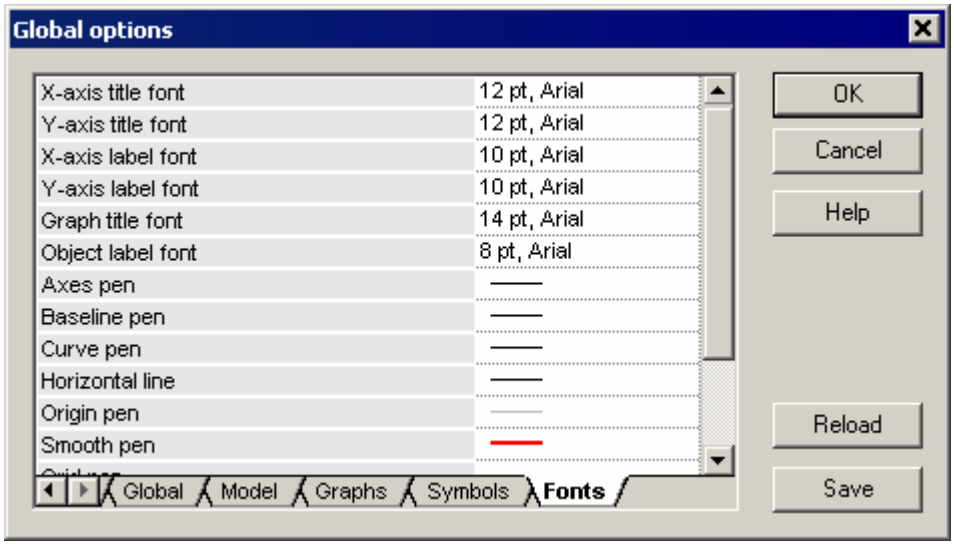
Allow multiple inclusions … allows you to select and include a variable in a stratigraphic diagram more than once.

Symbols



Sets the style of the default symbol and the symbols used in the 10 built-in class styles, with each class style having 21 distinct symbols.

Fonts



Sets the font and line styles.

Acknowledgements

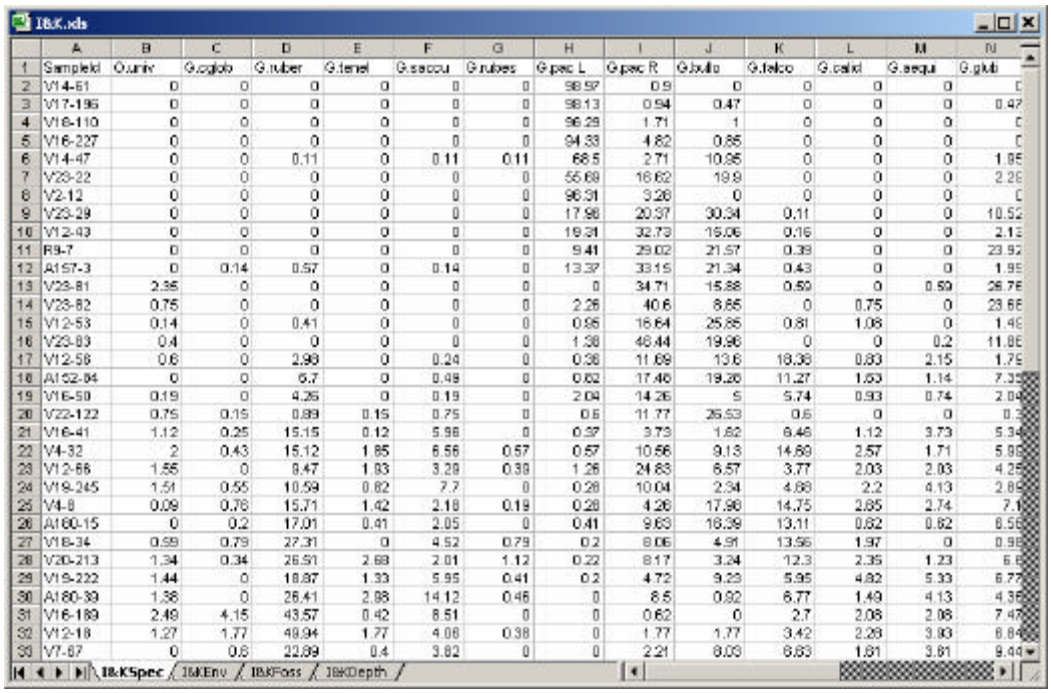
Acknowledgements

C2 brings together the data analysis functions of several of my earlier DOS programs, namely CALIBRATE, WAPLS, and MAT, and integrates them with a range of graphical visualisation tools into a single program. Some earlier code has been reused but C2 represents a re-write of many of the graphical, user-interface and numerical routines. I am very grateful to John Birks for providing the impetus and encouragement to tackle this project, and to the Norwegian Research Council (NORPAST project) for financial support.

Appendix

Importing data from Excel

C2 can import data from Excel and Lotus spreadsheets, dBase, Paradox and Access databases and ASCII comma- or tab-delimited files. To import correctly the data must be formatted according to certain rules. The example below shows data formatted in an Excel spreadsheet but the same rules apply to all the above formats.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	SampleId	O.univ	G.oglob	G.ruber	G.tenel	G.saccu	G.rubes	G.pac L	G.pac R	G.bulla	G.falco	G.calid	G.aequ	G.glob
2	VI4-61	0	0	0	0	0	0	98.97	0.9	0	0	0	0	0
3	VI7-195	0	0	0	0	0	0	98.13	0.94	0.47	0	0	0	0.47
4	VI6-110	0	0	0	0	0	0	96.29	1.71	1	0	0	0	0
5	VI6-227	0	0	0	0	0	0	94.33	4.82	0.85	0	0	0	0
6	VI4-47	0	0	0.11	0	0.11	0.11	68.5	2.71	10.95	0	0	0	1.95
7	V23-22	0	0	0	0	0	0	55.69	16.62	19.9	0	0	0	2.29
8	V2-12	0	0	0	0	0	0	96.91	3.28	0	0	0	0	0
9	V23-29	0	0	0	0	0	0	17.96	20.37	30.34	0.11	0	0	10.52
10	VI2-93	0	0	0	0	0	0	19.31	32.73	16.06	0.16	0	0	2.13
11	R9-7	0	0	0	0	0	0	9.41	29.02	21.57	0.39	0	0	23.92
12	A157-3	0	0.14	0.57	0	0.14	0	13.37	33.15	21.34	0.43	0	0	1.95
13	V23-81	2.35	0	0	0	0	0	0	34.71	15.88	0.59	0	0.59	26.76
14	V23-82	0.75	0	0	0	0	0	2.26	40.6	8.65	0	0.75	0	23.68
15	VI2-53	0.14	0	0.41	0	0	0	0.96	16.64	25.85	0.81	1.08	0	1.49
16	V23-83	0.4	0	0	0	0	0	1.38	46.44	19.96	0	0	0.2	11.86
17	VI2-56	0.6	0	2.98	0	0.24	0	0.36	11.69	13.6	18.38	0.83	2.15	1.79
18	A152-84	0	0	6.7	0	0.49	0	0.62	17.46	19.26	11.27	1.53	1.14	7.35
19	VI6-50	0.19	0	4.26	0	0.19	0	2.04	14.26	5	5.74	0.93	0.74	2.04
20	V22-122	0.75	0.15	0.89	0.15	0.75	0	0.6	11.77	25.53	0.6	0	0	0.3
21	VI6-41	1.12	0.25	15.15	0.12	5.96	0	0.37	3.73	1.62	6.46	1.12	3.73	5.34
22	V4-32	2	0.43	15.12	1.85	6.56	0.57	0.57	10.56	9.13	14.69	2.57	1.71	5.85
23	VI2-66	1.55	0	8.47	1.83	3.29	0.39	1.26	24.83	6.57	3.77	2.03	2.03	4.25
24	VI9-245	1.51	0.55	10.59	0.82	7.7	0	0.28	10.04	2.34	4.68	2.2	4.13	2.85
25	V4-8	0.09	0.78	15.71	1.42	2.18	0.19	0.28	4.26	17.96	14.75	2.85	2.74	7.1
26	A160-15	0	0.2	17.01	0.41	2.05	0	0.41	9.63	16.39	13.11	0.82	0.82	8.56
27	VI8-34	0.59	0.79	27.31	0	4.52	0.79	0.2	8.06	4.91	13.56	1.97	0	0.98
28	V20-213	1.34	0.34	26.51	2.68	2.01	1.12	0.22	8.17	3.24	12.3	2.35	1.23	6.8
29	VI9-222	1.44	0	18.87	1.33	5.95	0.41	0.2	4.72	9.23	5.95	4.82	5.33	6.77
30	A160-39	1.38	0	26.41	2.88	14.12	0.46	0	8.5	0.92	6.77	1.49	4.13	4.36
31	VI6-189	2.49	4.15	43.57	0.42	8.51	0	0	0.62	0	2.7	2.08	2.08	7.47
32	VI2-18	1.27	1.77	49.94	1.77	4.06	0.38	0	1.77	1.77	3.42	2.28	3.93	6.84
33	V7-87	0	0.6	22.89	0.4	3.82	0	0	2.21	8.03	6.83	1.81	3.81	9.44

1. If the first (leftmost) column contains numbers this column will be interpreted as sample numbers, and the second column will be interpreted as sample codes.
2. If the first column contains alphanumeric codes this column will be interpreted as sample codes and the sample number will be set to the row number.
3. If the column after the sample codes (ie. second or third column depending on 1 or 2 above) contains alphanumeric data this column will be interpreted as sample names. If the column after the sample codes contains numbers only it will be interpreted as data.
4. All other columns will be interpreted as data. Only numeric data is allowed – text data will generate an error. Empty cells are not allowed – either fill them with zero (if species data) or –99.9 to indicate missing values.
5. The first row should contain text not numeric labels for the variable codes.

6. Spaces and the following characters (+,.-*) should be avoided in the header row variable codes. These characters will be replaces by "#" on import.