C² A Microsoft Windows program for developing and applying palaeoecological transfer functions and for visualising multi-proxy stratigraphic datasets

Overview

The C² program consists of a data manager, data modelling tools, graphical / visualisation tools and a report manager, all integrated in an Explorer-like interface. The program can import datasets in Cornell, Tilia or Excel format and different datasets can have different numbers of observations or variables. Each dataset can have up to 20,000 observations (rows) and 20,000 variables (columns). Each observation or variable is identified by a unique 8-character code and an optional name of up to 60 characters. C2 contains a number of data editing and manipulation features, including a spreadsheet data browser / editor, (Figure 1), common data transformations, and merging of multiple datasets by row or column.

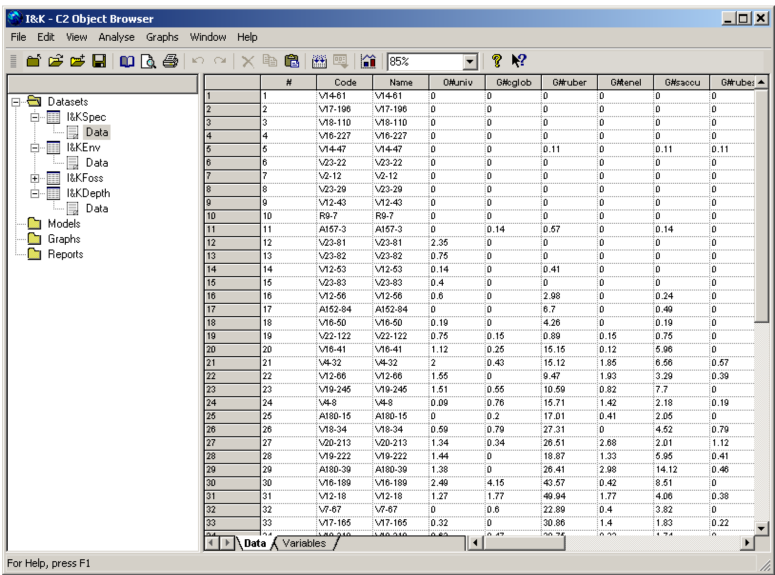

Figure 1: The C2 interface, with the Object Explorer on the left and Object Viewer on the right.

Transfer functions

Data modelling tools currently consist of numerical methods for deriving transfer functions (inference models) from a modern training set of microfossil, geochemical or other response variables, and an associated dataset of environmental, or predictor variables. The following methods are currently implemented: weighted averaging (WA), partial least squares (PLS), weighted averaging partial least squares (WAPLS), Imbrie and Kipp factor analysis (IKFA), correspondence analysis regression (CAR), modern analogue technique (MAT), and maximum likelihood regression and calibration. All methods allow arbitrary sample and variable deletion, percentage, $\log_{10}$ and square-root transformation of the species data, and $\log_{10}$ and square-root transformation of the environmental data. All methods also allow cross validation by leave-one-out, n-fold leave out, and bootstrapping. A range of model diagnostics, including sample specific errors, are calculated for the training set and fossil samples.

Each component of a model or reconstruction is saved and can be browsed, spreadsheet style, in the data viewer. Multiple models can be fitted in a single program run, allowing easy comparison of models generated using different numerical techniques, data transformations, sample deletions etc. Model results can be exported in Excel format, and existing models can easily be re-run, for example, to apply them to different fossil datasets (Figure 2).

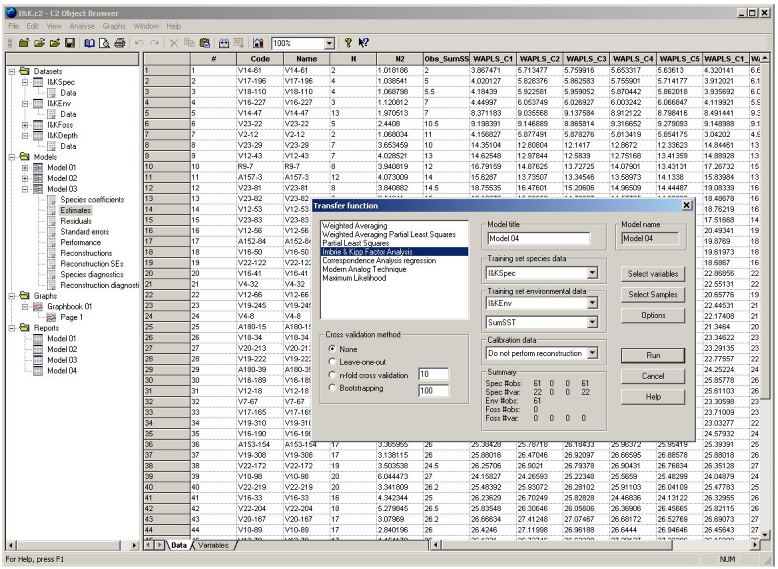

Figure 2: A list of fitted models on the left being browsed in the viewer, overlain by the transfer function dialog box.

A summary of each fitted model is also recorded in the Report window, which acts as a log file for the current session (Figure 3).

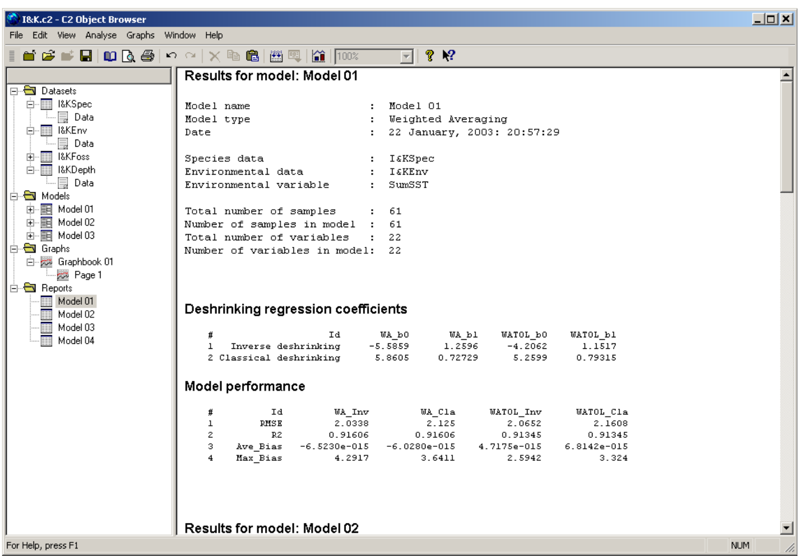

Figure 3: Report viewer showing model summary.

Graphics

C^2 currently implements two types of graphs: X-Y scatter, line or bubble plots, and stratigraphic diagrams. All graph types can mix variables from different datasets or model results, the only requirement is that the variables for the x- and y-axes have the same number of observations. X-Y scatter graphs are useful for exploring species / environment relationships, diagnosing models, and plotting model results. Each plot can contain multiple data series, making it easy to compare, for example, reconstructions derived using different numerical methods, and each page may contain multiple plots. A range of plotting symbols and line styles is available, and symbol size may be scaled according to a third variable. The data ranges of each plot can be independently and interactively zoomed, data points can be identified interactively, and graphs can be scaled to fit the current window size or printed page. Graphs can be printed, print-previewed or be copied to the clipboard as a Windows enhanced metafile.

The example in Figure 4 shows a diagram with four plots: a scatter plot of the relative abundance of the foraminifera *G. ruber* vs. summer SST (bottom left), a plot of summer SST vs. salinity, with symbols scaled according to the relative abundance of the foraminifera *G. sacculifer* (bottom right), a plot of observed vs. inferred summer SST for a weighted-averaging transfer function, with axes equally scaled and an outlier interactively identified (top left), and a plot of reconstructed summer SST derived from fossil foraminifera data, for WAPLS components 1 and 2.

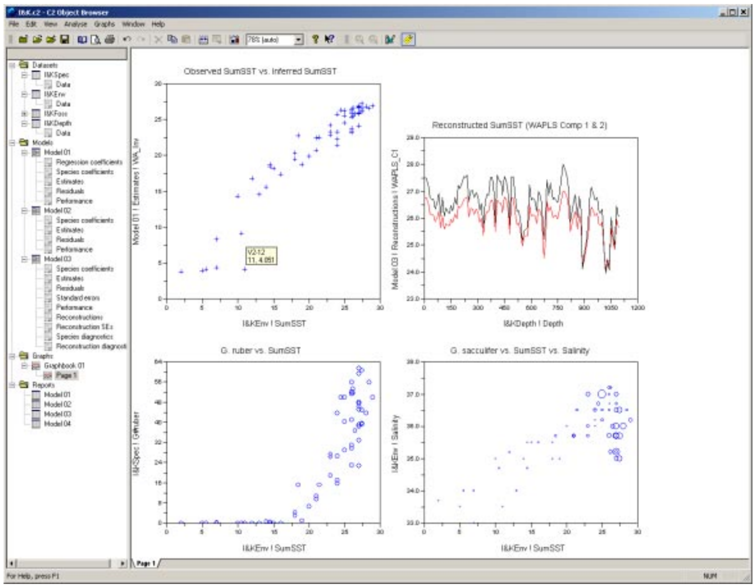

Figure 4: Example x-y scatter, line and bubble plots.

Figure 5 shows how a file of coastline coordinates can be imported and used to plot simple maps to explore spatial distributions in species occurrence and to identify trends in model residuals.

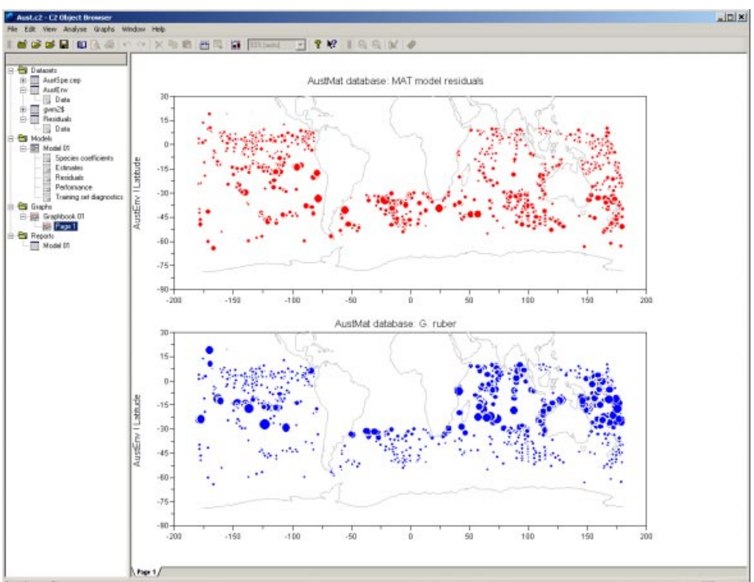

Figure 5: Bubble plots showing the distribution of G. ruber in the AustMat core-top database (bottom), and the spatial distribution of residuals for a MAT model fitted to the same database (top: notice the latitudinal trend in residuals).

Stratigraphic diagrams

C^2 plots stratigraphic diagrams and allows data from different datasets or models, recorded at different sampling intervals, to be combined into the same diagram. Each individual plot in the diagram can contain one or more data series, and each may be separately customised as a bar, line, symbol or silhouette graph. Data can be plotted on a linear or $\log_{10}$ scale and the length of the x-axis can be auto-scaled for percentage data. Fixed or variable error bars can also be specified for each curve. The first example below (Figure 6) shows the diatom stratigraphy for core RF55E from Roskilde Fjord, Denmark, with reconstructed total nitrogen derived from a diatom-based WAPLS transfer function, and bootstrap estimated sample-specific error bars.

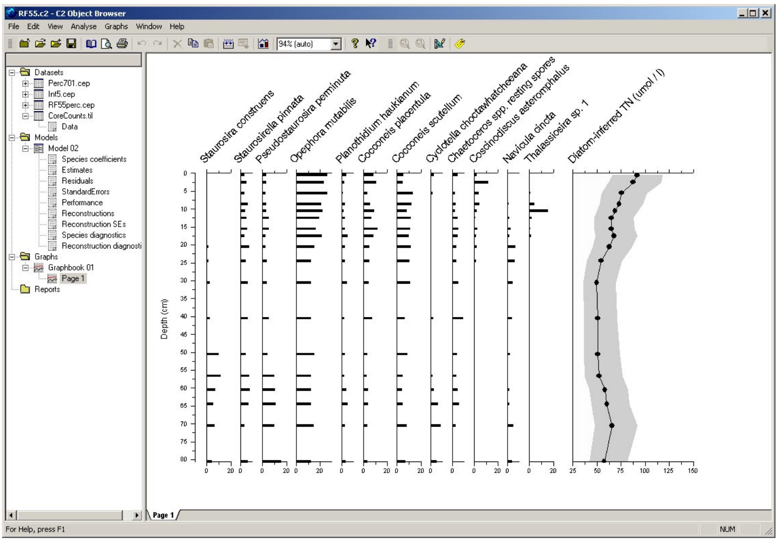

Figure 6: Diatom stratigraphy and reconstructed total nitrogen (TN), with bootstrap derived sample specific standard errors for core RF55E, Roskilde Fjord, Denmark.

Figure 7 shows multi-proxy pollen, diatom, geochemistry, magnetics and pigment data from Lago Albano (Italy) combined in a single diagram using a range of different graph styles for different data types. Each data type represents a different original dataset, and each contains a different numbers of observations recorded at different depth intervals in the core.

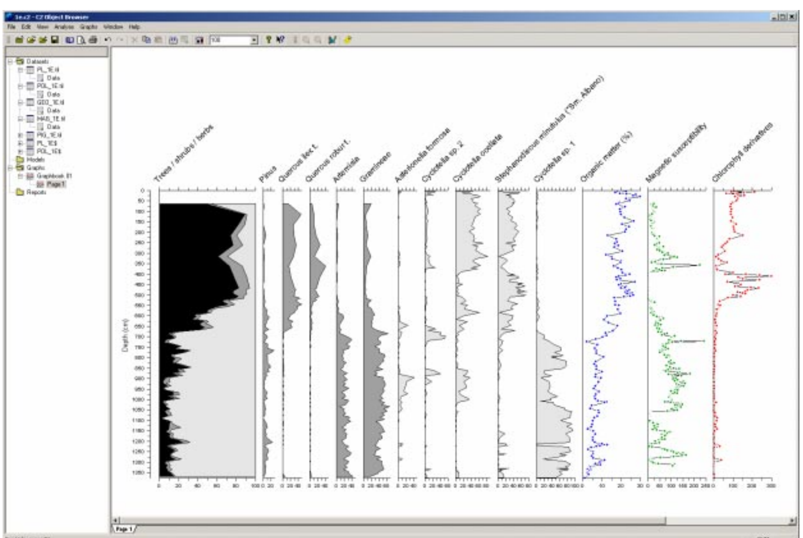

Figure 7: Multi-proxy stratigraphic data from Lago Albano, Italy, covering the last c. 30kyr.

C^2 comes with a standard Windows installation program that will install the program file, on-line help, a manual with tutorial, and example datasets.

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