C^2 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.

🖆 📽 🖼 🔛 🗟 🖑	$ \rangle$	< 🖻 🛍	🎬 🔍 🖌	1 85%	•	? №?					
		*	Code	Name	OMuniv	G#cglob	G#ruber	Gitternel	G#saccu	G#rube:	
Datasets	1	1	V14-61	V14-61	0	0	0	0	0	0	
E I&KSpec	2	2	V17-196	V17-196	0	0	0	D	0	0	
Data	3	3	V18-110	V18-110	0	0	0	0	D	0	
	4	4	V16-227	V16-227	0	0	0	0	0	0	
🖻 🥅 I&KEnv	5	5	V14-47	V14-47	0	0	0.11	D	0.11	0.11	
Data	6	6	V23-22	V23-22	0	0	0	0	0	0	
i I&KFoss	7	7	V2-12	V2-12	0	0	0	0	0	0	
🖃 🔲 I&KDepth	8	8	V23-29	V23-29	0	0	0	Û	Û	0	
Data	8	9	V12-43	V12-43	0	0	0	0	0	0	
Models	10	10	R9-7	R9-7	0	0	0	0	0	0	
- Graphs	11	11	A157-3	A157-3	0	0.14	0.57	D	0.14	0	
	12	12	V23-81	V23-81	2.35	0	0	0	0	0	
🛅 Reports	13	13	V23-82	V23-82	0.75	0	D	D	0	0	
	14	14	V12-53	V12-53	0.14	0	0.41	D	0	0	
	15	15	V23-83	V23-83	0.4	0	0	0	0	0	
	16	16	V12-56	V12-56	0.6	0	2.98	D	0.24	0	
	17	17	A152-84	A152-84	0	0	6.7	0	0.49	0	
	18	18	V16-50	V16-50	0.19	0	4.26	0	0.19	0	
	19	19	V22-122	V22-122	0.75	0.15	0.89	0.15	0.75	0	
	20	20	V16-41	V16-41	1.12	0.25	15.15	0.12	5.96	0	
	21	21	V4-32	V4-32	2	0.43	15.12	1.85	6.56	0.57	
	22	22	V12-66	V12-66	1.55	0	9.47	1.93	3.29	0.39	
	23	23	V19-245	V19-245	1.51	0.55	10.59	0.82	7.7	0	
	24	24	V4-8	V4-8	0.09	0.76	15.71	1.42	2.18	0.19	
	25	25	A180-15	A180-15	0	0.2	17.01	0.41	2.05	0	
	26	26	V18-34	V18-34	0.59	0.79	27.31	0	4.52	0.79	
	27	27	V20-213	V20-213	1.34	0.34	26.51	2.68	2.01	1.12	
	28	28	V19-222	V19-222	1.44	0	18.87	1.33	5.95	0.41	
	29	29	A180-39	A180-39	1.38	0	26.41	2.98	14.12	0.46	
	30	30	V16-189	V16-189	2.49	4.15	43.57	0.42	8.51	0	
	31	31	V12-18	V12-18	1.27	1.77	49.94	1.77	4.06	0.38	
	32	32	V7-67	V7-67	0	0.6	22.89	0.4	3.82	0	
	33	33	V17-165	V17-165	0.32	0	30.86	1.4	1.83	0.22	

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, log₁₀ and square-root transformation of the species data, and log₁₀ 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).

🖆 😂 🖬 🔛 🖪 😂 🗠	aX	b 🖪 😷	四 論 1	00%	- 8	N?								
		#	Code	Name	K	I H2	Obs_SumSS	WAPLS_C1	WAPLS_C2	WAPLS_C3	WAPLS_C4	WAPLS_C5	WAPLS_C	1_ WA
E Datasets	1	1	V14-61	V14-61	2	1.018186	2	3.867471	5.713477	5.759916	5.653317	5.63613	4.320141	6.E
E- I I&KSpec	2	2	V17-198	V17-196	4	1.038541	5		5.828376	5.862583	5.755901	5.714177	3.912021	6.1
Data	3	3	V18-110	V18-110	4	1.068798	5.5	4.18439	5.922581	5.959052	5.870442	5.862018	3.935692	6.0
E ISKENV	4	4	V16-227	V16-227	3	1.120812	7	4.44997	6.053749	6.026927	6.003242	6.066847	4.119921	5.5
Data	5	5	V14-47	V14-47	13	1.970513	7	8.371183	9.035568	9.137584	8.912122	8.798416	8.491441	9.3
⊕-Ⅲ I&KFoss	6	6	V23-22	V23-22	5	2.4408	10.5	9.198391	9.146889	8.865814	9.316652	9.279093	9.148988	9.1
E- ILKDepth	7	7	V2-12	V2-12	2	1.069034	11		5.877491	5.878276	5.813419	5.854175	3.04202	4.5
Data	8	8	V23-29	V23-29	7	3.653459	10		12.80804	12.1417	12.8672	12.33623	14.84461	13
Models	9	9	V12-43	V12-43	7	4.028521	13		12.97844	12.5839	12.75168	13.41359	14.88928	13
E Model 01	10	10	R9-7	R9-7	8	3.940819	12		14.87625	13.72725	14.07901	13.43131	17.26732	15
H Model 02	11	11	A157-3	A157-3	12	4.073009	14			13.34546		14.1338	15.83984	13
E-III Model 03	12	12	V23-81	V23-81	8	3.840882	14.5		16.47601	15.20606	14.96509	14.44487	19.08339	16
- Species coefficients	13	13	V23-82	V23-82		0.01011		10 10070					18.48678	16
Estimates	14	14	V12-53	V12-00	ansfer fu	nction						×	18.76219	16
- Residuals	15	15	V23-83	V23-83	Weighted	Australia		- Model titl			- Model nam		17.51668	14
Standard errors	16	16	V12-56	V12-56	Weighted	Averaging Partial	Least Squares	Model 0			Model 04		20.49341	19
Performance	17	17	A152-84	A152-84	Partial Lea	ast Squares		Iwoge n	4		IMODE U4		19.8769	18
- Reconstructions	18	18	V16-50			ipp Factor Analys		Testilian	set species dal	-			19.61973	18
- Reconstruction SEs	19	19	V22-122	V22-12;	Correspon	dence Analysis re nalog Technique	gression						18.6867	16
- Species diagnostics	20	20	V16-41	V16-41	Maximum			18KSpec		•	Select variables		22.86856	22
Reconstruction diagnosti	21	21	¥4-32	V4-32				Teriting		and state			22.55131	22
🔠 Graphs	22	22	V12-66	V12-66				Training set environmental data I&KEnv SumSST			Select Samples Options		20.65776	19
😑 🚾 Graphbook 01	23	23	V19-245	V19-245									22.44531	21
Page 1	24	24	V4-8	V4-8									22.17408	21
E Reports	25	25	A180-15	A180-11	C	idation method		Toomoor					21.3464	20
- Model 01	26	26	V18-34 V20-213	V18-34				Calibratio	n data				23.34622	23
- Model 02				V20-21:	None	6		Do not perform reconstruction			B	23.29135 22.77557	23	
- Model 03	28	28	V19-222	V19-22	C Leave-one-out							nuri		22
- Model 04	29	29	A180-39	A180-35	C nick	d cross validation	10	Summary			Can	on -	24 25224	24
	30	30	V16-189	V16-185	Spec #obs: 61 0 0 61					-	25.85778	26		
	31	31	V12-18	V12-18		suapping	100	Spec #va		0 22	He	in I -	25.61103	26
	32	32	V7-67	V7-67				Env #obs Foss #ob				<u> </u>	23.30598	23
	33	33	V17-165	V17-16				Foss #va		0 0			23.71009	23
	34	34	V19-310	V19-31(23.03277	22
	35	35	V16-190	V16-19(24.57932	24
	36	36	A153-154	A153-154	17	3.365955	26		25.78718	26.18433	25.96372	25.95419	25.39391	25
	37	37	V19-308	V19-308	17	3.138115	26		26.47046	26.92097	26.66595	26.88578	25.88018	26
	38	38	¥22-172	V22-172	19	3.503538	24.5		26.9021	26.79378	26.90431	26.76834	26.35128	27
	39	39	V10-98	V10-98	20	6.044473	27		24.26593	25.22348	25.5659	25.48299	24.04879	24
	40	40	V22-219	V22-219	20	3.341809	26.2		25.93072	26.28102	25.91103	26.04109	25.47783	25
	41	41	V16-33	V16-33	16	4.342344	25		26.70249	25.82828	24.46836	24.13122	26.32955	26
	42	42	V22-204	V22-204	18	5.279845	26.5		26.30646	26.05606	26.36906	26.45665	25.82115	26
	43	43	V20-167	V20-167	17	3.07969	26.2		27.41248	27.07467	26.68172	26.52769	26.69073	27
	44	44	V10-89	V10-89	17	2.840196	26	26.4246	27.11998	26.96188	26.6444	26.94646	26.45643	27
		Variables	7				1							

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).

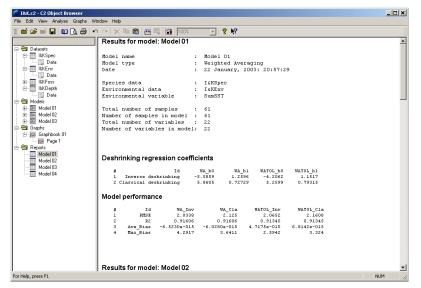


Figure 3: Report viewer showing model summary.

Graphics

C² 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 weightedaveraging 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.

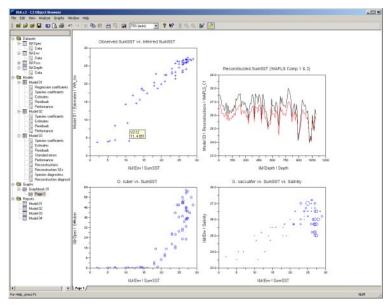


Figure 4: Example x-y scatter, line and bubble plots.

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.

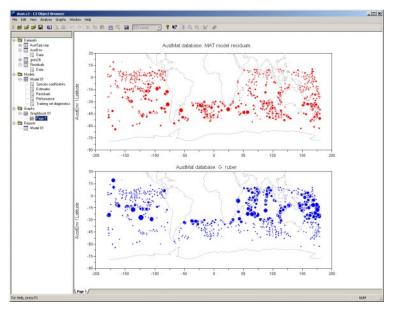


Figure 5: Bubble plots showing the distribution of G. ruber in the AustMat core-top database (bottom), and the spatial distribution of residuals for a MAT model fitted to the same database (top: notice the latitudinal trend in residuals).

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 log₁₀ 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.

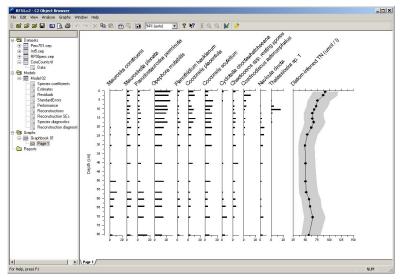


Figure 6: Diatom stratigraphy and reconstructed total nitrogen (TN), with bootstrap derived sample specific standard errors for core RF55E, Roskilde Fjord, Denmark.

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.

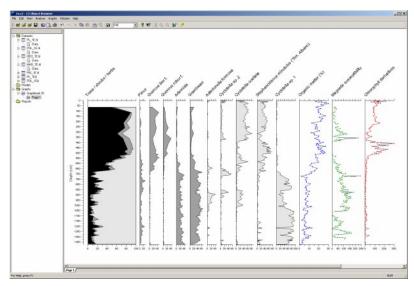


Figure 7: Multi-proxy stratigraphic data from Lago Albano, Italy, covering the last c. 30kyr.

 C^2 comes with a standard Windows installation program that will install the program file, online 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 <u>Stephen.Juggins@ncl.ac.uk</u> http://www.staff.ncl.ac.uk/stephen.juggins