Documentation for psimpoll 4.25 and pscomb 1.03

C programs for plotting pollen diagrams and analysing pollen data

K.D. BENNETT
Quaternary Geology, Department of Earth Sciences
Uppsala Universitet
Villavägen 16
SE-752 36 Uppsala
Sweden

July 1, 2005

Contents

1 Preface

This is the manual for **psimpoll** and **pscomb**, computer programs for the production of diagrams displaying pollen and other stratigraphical data to a standard acceptable as camera-ready copy for publication, with all the features that would be expected in a hand-drawn diagram. It assumes only that you have access to a computer (which may be of any make or operating system), you are basically familiar with the process of manipulating files with a text editor, you have access to an output device with a PostScript interpreter, and you have some data that you wish to plot.

This manual describes **psimpoll** 4.25 and **pscomb** 1.03. Development of **psimpoll** from earlier versions has been as follows, in summary:

4.25 Addition of local regression fitting for age-depth models;

4.10 Addition of facility to handle calibrated radiocarbon ages;

3.10 Addition of correspondence analysis;

3.00 Introduction of colour.

10 November 2003

K.D. Bennett
Palaeobiology
Geocentrum, Uppsala University
Villavägen 16
SE-752 36 Uppsala
Sweden

fax: + 46 18 55 5920
e-mail: Keith.Bennett@geo.uu.se
Anonymous ftp[1]
WWW[2]

[1] See URL ftp://ftp.kv.geo.uu.se
[2] See URL http://www.kv.geo.uu.se/

2 Acknowledgements

psimpoll has been through many versions and changes from early days as a FORTRAN program on an IBM mainframe, through a phase in QuickBASIC on IBM PCs and DOS, and now available on any operating system with an ANSI C compiler. Over this time, it has gradually acquired the features necessary to display the complex range of forms of pollen diagrams, and the means to analyse data and display results conveniently and routinely. I am extremely grateful to friends and colleagues who have assisted in this process by making suggestions for improvements, and tolerating bugs and idiosyncrasies. I thank Steve Boreham, Jane Bunting, Alex Chepstow-Lusty, Petra Day, Julie Fossitt, Janice Fuller, Simon Haberle, Susie Lumley, Jim Ritchie, Helen Roe, Rebecca Teed, Chronis Tzedakis, and Kathy Willis for help with the program, Roger Humphry for help with rate-of-change analyses, and Tom Kelly for advice on programming in `C` and the mysteries of UNIX. John Birks first introduced me to the analysis and plotting of pollen data by computers, and Lou Maher has been a continual source of help, advice, and encouragement for many years. I am deeply indebted to both. If there is any merit in **psimpoll** and **pscomb**, all of these people should share the credit, but I take responsibility for all the errors and deficiencies. I also acknowledge the contribution of journal reviewers and editors who have generated the endless variety of bizarre requirements for the format of diagrams that serve to make a programmer's life unnecessarily difficult. If only it were possible to please all of them all of the time.

31 August 1995
K.D. Bennett
Palaeobiology
Uppsala University
Villavägen 16
S-752 36 Uppsala
Sweden

fax: + 46 18 55 5920
e-mail: Keith.Bennett@geo.uu.se
Anonymous ftp[3]
WWW[4]

[3]See URL ftp://ftp.kv.geo.uu.se
[4]See URL http://www.kv.geo.uu.se/

3 Disclaimer

No portion of these programs or manual may be used or reproduced for use by a commercial facility without the express written consent of the author. The author makes no warranty, express or implied, nor assumes any legal liability or responsibility for the usefulness of any portion of the programs. This copyright message and disclaimer must accompany any reproduced portion of these programs.

Helvetica, New Century Schoolbook, Palatino, and Times are registered trademarks of Allied Corporation.

IBM and PS/2 are registered trademarks of International Business Machines Corporation.

ITC Avant Garde and ITC Bookman are registered trademarks of International Typeface Corporation.

Lotus is a registered trademark of Lotus Development Corporation.

LaserWriter and Macintosh are registered trademarks of Apple Computer Incorporated.

PostScript is a registered trademark of Adobe Systems Incorporated.

Open Desktop and SCO are registered trademarks of The Santa Cruz Operation, Inc. Open Server is a trademark of The Santa Cruz Operation, Inc.

Symantec and THINK C are trademarks of the Symantec Corporation.

Turbo C is a registered trademark of Borland International Incorporated

UNIX is a registered trademark of AT&T Information Systems.

WordStar is a registered trademark of MicroPro International Corporation.

4 Introduction

4.1 Contents

4.2 Background

psimpoll 4.25 and **pscomb** 1.03 are ANSI C programs that translate pollen and other stratigraphical data into PostScript page description files that can be 'printed' on a device with a Postcript interpreter to produce a pollen or other stratigraphical diagram. The aims of these programs are to:

- produce pollen diagrams of publishable quality;
- enable the transmission of pollen diagrams on electronic media (diskettes, e-mail);
- facilitate the use of numerical, multivariate analyses of pollen data by linking such analyses to the production of diagrammatic output as a matter of routine;
- be highly portable, independent of any particular hardware.

Most pollen analysts today have access to computer facilities for carrying out, at least, basic calculations on their raw data and plotting pollen diagrams. Traditional requirements for the presentation of pollen data in graphical form make it difficult to use commercial graphics packages for the purpose, and most pollen analysts use software written specially for this use. For example, in the 1970s John Birks and Brian Huntley developed POLLDATA for calculation and graphical presentation of pollen data on the Cambridge IBM mainframe. This program was successfully transferred to other mainframes, and now has a PC version. More recently, a number of programs for PCs, notably TILIA and TILIAGRAPH (Grimm 1992) have come into general use in many pollen labs around the world (Davis 1993). I have seen and used several such programs (all my Ph.D. thesis work was done with POLLDATA) but over the last few years all my pollen data-crunching and plotting has been done with software that I have written myself: **psimpoll** and **pscomb** are the latest manifestations of that.

Commercial graphics packages cannot be used easily to plot pollen diagrams, but spreadsheets are well-suited to making calculations from raw pollen data. There is little to be gained from writing software for this aspect of pollen data-handling, and **psimpoll** provides none. The most inconvenient part of

routine data-handling is the highly subjective process of constructing a pollen sum and calculating proportions from it. The process requires decisions about each taxon (in the main sum, or not), and for the entire dataset to be present (for the construction of the sum). Spreadsheets provide a convenient way for the analyst to do both. Results can be exported as a text file that may then be read by a plotting program directly, or after modification by a suitable text editor. **psimpoll** provides the first of two steps necessary to plot results: it reads in calculated pollen and ancillary data, annotated to indicate data type (percentages, concentrations, etc), performs certain non-routine analyses, and writes a PostScript page description file, containing the information needed for a PostScript interpreter to produce the pollen diagram. The second step is the passing of that file to the interpreter.

psimpoll reads data from up to four files supplied by the user, and may produce up to nine output files, including a configuration file that saves the current setup for use with subsequent runs. All these files (input and output) are plain text and are readable and modifiable by text editors. The user's input files consist of the main input file (essential), and up to three optional associated files with data on location of zones, radiocarbon ages, and sediment stratigraphy, using the notation of Troels-Smith (1955). **psimpoll** looks for associated files from default or provided file names, uses them if they are present, and carries on without if they are not.

The format of output can be altered by options introduced within the main input file, and by options selected from a menu when **psimpoll** is run. Within the main input file, a two-character code identifies the data (e.g., pollen percentages, concentrations etc), all presented by depth, age, or sample numbers (e.g., for a surface sample dataset). Suitable editing of the dataset enables the user to mark particular samples, pollen types, or individual data values to be ignored, plot different pollen types at different scales, mark certain 'types' as charcoal, rarefaction data, rate-of-change data, or to enable **psimpoll** to recognize a sequence of data for a summary diagram. **psimpoll** can also carry out rarefaction analysis, zonation, principal components analysis, independent splitting of taxa, rate-of-change analyses, and basic statistical description on suitable datasets. All diagram labelling can be modified, and all characters from European languages with Roman script, plus Greek, are available. Thus, although **psimpoll** runs in English, output can be tailored for other languages.

psimpoll will either run a dataset immediately using default options, or the user can change the defaults through a menu. Any changes can be saved for reuse in a configuration file. PostScript output is written to a file by default, but can be directed straight to a printer with PostScript interpreter. Output includes sediment stratigraphy, pollen zones, and radiocarbon dates if the necessary associated files were present when **psimpoll** was run.

psimpoll can handle only one input dataset at a time. It is obviously desirable to be able to combine data from different datasets (sites) to produce combined plots. This facility is provided by **pscomb** (pronounced to rhyme with 'tome', but 'scum' will also do). **pscomb** reads data from PostScript files previously written by **psimpoll**, enables selection of particular pollen types or other

output (e.g., depth axes, zonation, radiocarbon, or sediment columns), and combines them into a single output file. **pscomb** thus allows the simultaneous presentation on a common axis of taxa from many different datasets.

psimpoll is a straightforward plotting program. It works reasonably well, and has most features that a pollen analyst will want. It exists because I have found that it suits me to have available a program that I can tinker with: adding, modifying, and deleting features as necessary. I do not have the time to ensure that the program is completely bug-free, user-friendly, or documented to the standard that would be required for a marketable product. I fix problems as they occur, or when a colleague complains loudly enough. It copes with the kind of data generated in our group, but I have made no effort to include features that might be found useful by others.

psimpoll and **pscomb** are not guaranteed bug- or idiot-proof. I believe they work for all reasonable input, but I have not had a chance to check out all possible combinations of options. There is a degree of error checking, but it is up to users to get their datasets in order. Please notify me about examples of unsatisfactory output so I can prevent the same problem occurring in the future.

4.3 Implementation

psimpoll and **pscomb** have been installed on PCs with processors from 8086, 80286, 80386, 80486, and Pentium families, running under DOS, Windows 3.1, Windows 95, Windows 98, Windows NT, and Windows 2000, and under SCO UNIX, Linux, Sun and Silicon Graphics, and on Apple Mac Plus, Mac LCII, Performa and iMac.

Pollen data is normally collected by a Psion Organiser (Bennett 1990) and transferred to a computer. On PCs, the data may be manipulated using Microsoft Excel or Lotus 123 (or other spreadsheet), and text files suitable for input to **psimpoll** produced with a text editor. A DOS batch file is available that enables the running of **psimpoll** sequentially with each of up to nine files, from any part of the file system (i.e., not necessarily in the same directory as the program). The command is `pp`, and it can (optionally) be followed by a list of up to 9 filenames on the same line (`pp filename1 filename2` etc).

On Macintosh computers, **psimpoll** should be installed within a 'psimpoll' folder, and data files should be placed in the same folder. I have not yet figured out Mac folder and filename conventions to the extent of being able to advise on how to run the program using datasets in other folders.

PostScript output files can be printed under DOS on a LaserWriter with the command:

```
PRINT filename <Enter>
```

(answer `COM1` if asked for a 'list device')

Output files produced on Macintosh can be printed on a LaserWriter connected to an Apple computer using ShowPages.

PostScript files can be viewed using the GhostScript and GhostView[5] family of programs on Windows, Apple, and Unix environments.

The main advantage of using a locally-written program such as **psimpoll** is that changes can be made in response to particular needs. Many of the options in **psimpoll** stem from requests for such features. If you want or need something that is not available now, ask.

4.4 Installation

4.4.1 IBM-compatibles, Apple MacIntosh, Linux

psimpoll and **pscomb** can be obtained, free and gratis, from Uppsala by anonymous ftp[6]. Executable files and documentation are in the sub-directory 'pub/psimpoll'. These files normally contain the current versions of programs and documentation. It is also available at the INQUA public archive at Wisconsin, and may be retrieved by anonymous ftp[7]. DOS, Windows 95 etc, Linux, and Apple executable files are in subdirectory '/pub/inqua'. In either case, files should be downloaded by binary transfer, unpacked and installed on a hard-disc drive.

The same files may be accessed on the web at URL http://www.kv.geo.uu.se/psimpoll.html[8].

The versions of the program that run on window-based systems (Windows 3.1, Windows 95, Apple) use a 'console' window that provides a means to run text-based programs within the windowed environment. This console usually has to be closed explicitly after the end of the program.

4.4.2 Other computers

Contact me (address and other details in the Preface (p. 8)). I will supply the ANSI C source code and you will then need to compile this with your favourite C compiler. It has been compiled successfully on AIX, Sun, and Silicon Graphics operating systems, in addition to those mentioned above.

[5] See URL http://www.cs.wisc.edu/.ghost/index.html
[6] See URL ftp://ftp.kv.geo.uu.se
[7] See URL ftp://geology.wisc.edu
[8] See URL http://www.kv.geo.uu.se/psimpoll.html

4.5 Colour

The file `psimpoll.COL` needs to be available in order for colours to be used with diagrams. The location of the file varies depending on the system. For command line systems (DOS, Unix), place `psimpoll.COL` in the same directory as the data files. For window-based systems (Windows 3.1, Windows 95, Apple), place `psimpoll.COL` in the same directory as the program. If you edit this file on computers with more than one **psimpoll**-user, bear in mind that changes you make may affect other users.

5 Data preparation

5.1 Contents

psimpoll is basically a plotting program. Calculations of percentages and concentrations (but not necessarily accumulation rates), and most other things must be done beforehand. All input files (p. 103) should be in plain text format, but how they are produced is irrelevant to **psimpoll**.

pscomb receives as input files that have been previously output by **psimpoll**. There is therefore no data preparation, except the running of **psimpoll**. The main points to bear in mind when running **psimpoll** if output will be used in **pscomb** are:

- Change any taxa names within **psimpoll** to the form that you want on the **pscomb** diagram. This might, for example, mean including a site name with a taxon name. Use option 't' when selecting taxa interactively (**psimpoll** menu K (p. 50));
- Ensure that taxa destined to be plotted against a common axis really do have the same axes. For example, use menu H (p. 43) in **psimpoll** to plot samples within a time-scale that will be used uniformly across a number of sites.

5.2 Simple use

The easiest way to get started is simply to ignore most of this documentation. Prepare a small dataset, with any data (real or imagined, using or copying from the example main input file (p. 104)). Get **psimpoll** running, with the command `psimpoll filename`. **psimpoll** will run and create a file called `filename.PS`. Then send `filename.PS` to a PostScript printer, as described under Running psimpoll (p. 28).

5.3 Format of main input file

Datasets are structured around 'samples' and 'taxa'. A 'sample' is the basic unit of study, within which various attributes have been measured. Samples

may be stratigraphically related to each other, in which case the term 'level' is often used for them. The measured attributes are 'taxa', and they may be concrete (e.g., the abundance of *Betula* pollen), or abstract (e.g., the rate-of-change of something or other). The 'number of taxa' on line 2 must include them all, except rare cases indicated explicitly below. Any text in an input file may include certain options (for oblique style, accents, superscripts, etc): these are described fully under menu I (p. 43).

The order of entries in a file is important, as described below, but the exact layout of the file is less important. In general, text items (such as taxon names) are considered to be ended by either a comma (','), or by the end of a line. Numerical values, such as the data themselves are generally ended by either one or more spaces or the end of the line. TAB characters have the same effect as spaces. It is thus possible to have either of two basic layouts: one item per line, giving a file with one column, or one taxon per line, beginning with the name, followed by a comma, and then all the numerical values for that taxon separated by spaces. Each layout has its uses, depending on where the data come from. All intermediate combinations of these basic layouts are possible, as long as each item is correctly ended (with comma, space, newline, TAB, as appropriate).

Item 1 A title, any reasonable number of characters, on a line of its own. Must not begin with the character '#', to prevent confusion with TILIA-type files (p. 19).

Item 2 Number of taxa in the dataset.

Item 3 Number of samples in the dataset.

Item 4 A two-character code (first character (p. 22) and second character (p. 23)).

Items 5 onwards Depths, ages, sample numbers, or sample labels. Labels, if used (option 't'), should be presented one per line, or separated by commas, to a maximum length of 39 characters per label. Numbers can be separated by spaces or presented one per line, and may be positive or negative. Following the usual convention of measuring depths from the top of a sequence downwards, positive values increase down the column, and negative values increase upwards. If the first character of the 2-character code (Item 4) is 'm', the sample value should be followed by the sample thickness, in the same units.

Items 6 onwards The data, presented by taxa (i.e., taxon name (maximum 39 characters), followed by values for each sample for that taxon). When using the 'x' option (Item 4), the taxon name must be followed by a label for the units (see above): this can be a space character (i.e., no unit). The data can be presented with each item on a separate line, or with the taxon name followed by a comma, label (if any) followed by a comma, then the data values separated by spaces.

Items 7-14 optional additional data: give one value per sample, except that a single value prefixed by '-' indicates that the value applies to all samples.

Items 7-13 are needed for concentration confidence intervals: Item 14 is additional information for accumulation rate confidence intervals.

Item 7 marker pollen suspension (tablet) concentration.

Item 8 if known, sample standard deviation of marker concentration; otherwise enter '1', and the sample standard deviation is calculated as the square root of the marker concentration (this assumes a Poisson distribution in marker suspensions).

Item 9 volume of marker suspension added or the number of tablets.

Item 10 if known, sample standard deviation of marker suspension added; otherwise enter 1, and the standard deviation is calculated as 0.8% of the volume added. For tablets, enter 0.

Item 11 marker grain count for each sample (must be >100 for calculation of confidence intervals);

Item 12 volume of sediment added.

Item 13 if known, sample standard deviation of sediment volume; otherwise enter 1, and the standard deviation is calculated as 1% of the sediment volume.

Item 14 if known, height of the sediment sample along the core; otherwise enter 1, and it will be assumed to be the cube root of the sediment volume. For a sampler with a circular cross-section, enter the diameter of the sampler.

5.4 Options within main input file

Several optional affects can be achieved by suitable editing of the dataset.

- Samples. Any sample value can be excluded by marking it with 'x', after the number, with no intervening spaces (e.g., `210x`). Use to omit suspect samples without redoing the whole file, or to plot a stratigraphic portion of the diagram (but note the use of menu H (p. 43) for this latter purpose);
- Data. Any data value equal to (by default) -1 will be ignored in producing the plot file. Use where a data value is missing for whatever reason (e.g., charcoal values in every other sample). The default value can be changed in menu Bg (p. 35);
- Taxon names. Several effects are possible here, to allow the mixing of plots of pollen data with certain non-pollen data using reasonable scales and captions, and to select a subset of taxa for plotting. Access to all is by prefixing the taxon name with one or more characters. All labels associated with taxa names can be modified in menu Ik (p. 48). The effects listed below are to do with the type of object being plotted: its recognition, treatment, and labelling of its axes. Additional effects are

possible within the piece of text used as the taxon name. These are all detailed below in the description of menu I (p. 43). Certain taxa have particular input requirements, and, if present, must be identified by prefixes (p. 23). Other prefixes (p. 24) affect labelling and presentation of axes. They can be omitted initially, and given interactively by renaming the taxon (see menu K (p. 50)). Other effects can be achieved by inclusion of characters within taxon names (see menu I (p. 43)).

5.5 Input from a TILIA-type file

psimpoll can recognise and handle input files prepared in TILIA-type format. These include those produced by TILIA itself (general format) (Grimm 1992), and `.p15` files available from NOAA[9]. These files are distinguished by having a '#' as the first character of the first line. They are generated from TILIA (ver. 2) by selecting [`D`] `Save data file` from the TILIA (ver. 2) main menu, then [`A`] `General format`, and answering the prompts for file name, dataset title, and number of decimal places for the data. **psimpoll** can also handle the ascii files of raw pollen data available from NOAA. It does so by calculating percentages from the raw data, using TILIA's taxon category codes to construct a sum. The codes currently defined in **psimpoll** (following those found in NOAA files) are:

A Trees and shrubs;

B Herbs;

C Dwarf shrubs;

D *Equisetum*;

F Pteridophytes;

G, M *Sphagnum*;

J Parasitic plants;

Q Aquatics;

X Indeterminable grains;

The main pollen sum is constructed from A + B + C + D + F + J. The taxa in the groups '*Sphagnum*', 'Aquatics' and 'Indeterminable grains' are calculated as percentages of the main sum plus the sum of their group. These sums are saved and included in the dataset, as taxa called 'Sum ABCDFJ', 'Sum GM', 'Sum Q', and 'Sum X', respectively. They can be reused if the dataset is saved in **psimpoll** format (menu J (p. 50)). Taxa with category codes other than those listed above are left unchanged, and may not plot correctly.

[9] See URL http://www.ngdc.noaa.gov/paleo/paleo.html

The table above can be used to edit the NOAA `.ASC` files appropriately if the category codes used do not coincide with those above.

psimpoll also detects, reads and uses radiocarbon date information from NOAA `.ASC` files, provided that this is detected in a standard format:

```
# Radiocarbon dates:
#   Depth   Thk     Age   SDup   SDlo  Lab no.    Basis  Material
#   159.0   5.0   10170     60     60  TP-313       U    Wood
#   212.0   6.0   10520     60     60  TP-456       U    Charcoal

...

# Dating notes:   Wood = pine
#
```

The phrase 'Radiocarbon dates:' is used to detect the start of the section, and either 'Dating notes:' or a line with just '#' are used to detect the end of the section. If missing, any values for 'Thk', 'SDup' or 'SDlo' are set to zero. 'Depth' and 'Age' must be present.

5.6 Associated files

psimpoll looks for optional associated files for information to include on the diagram and to control writing of the diagram. The names of these are normally derived from the main input file name, and, by default, each should be on the same disc and in the same directory as the dataset file, have the same first four characters of the filename, with the next characters indicating the type of the information file.

Two files save the state of **psimpoll** in its current run for optional re-use in a later run. One, with the filename extension `.CFG`, holds the current configuration of modifiable variables. The second, with the extension `.SCR`, saves the details of interactive plotting. In both cases, the filename extension replaces any extension on the main input filename, and the rest of the file name is as the main input file.

Finally, a file called `psimpoll.COL` contains the details of available colours.

Follow the links below for file formats and examples.

`fileTS` (p. 25) Troels-Smith lithological symbols (Example (p. 105));

`fileC14` (p. 25) Radiocarbon-ages (Example (p. 105));

`fileCAL` (p. 26) Control file for calibrated ages (Example (p. 106));

`xxxx` (p. 26) Calibrated ages (from BCal) (Example (p. 106));

`fileZONE` (p. 26) Zonation data for inclusion in labelled boxes at the end of the diagram (Example (p. 107));

`file.CFG` A complete list of settings from the menu. When it is read in, any filenames that specify a drive letter are changed to have the same drive letter as the main input file. This file is read by default, but may be written from menu O (p. 79) (Example (p. 107));

`file.SCR` A record of the taxa and options selected from interactive plotting. Written by default, but read for re-use by choosing menu P (p. 79) (Example (p. 107))

`psimpoll.COL` A colour palette for **psimpoll**, containing colour definitions on the CMYK system, referenced to an identifier and a name. Colours in the palette are accessed by name in the `TS` file, the `ZONE` file, menu C (p. 35), and interactively (p. 50). They may also be accessed by the identifier from the interactive menu (p. 50), for brevity.

Thus, a file called `a:\dallp` could have associated with it the files `a:\dallTS`, `a:\dallC14`, `a:\dallZONE`, `a:\dallp.CFG`, `a:\dallp.SCR` and `psimpoll.COL`. I recommend using the first four characters of a filename to indicate site (e.g., `dall` for 'Dallican Water'), and the fifth letter to indicate data type (e.g., `p` for percentages, `c` for concentrations, `a` for accumulation rates, and `b` for a concentration dataset which is to be converted to accumulation rates). The same set of associated files will then work for all types of diagrams from the same site, except that the `.CFG` and `.SCR` files are specific to the main input file. The colour palette file can be made specific to a site, or kept general. Avoid filename suffixes. **psimpoll** ignores the `TS` file when datasets are presented by age.

The names of associated files (except for `.CFG`, `.SCR`, and `psimpoll.COL`) can be changed (menu E (p. 41)).

5.7 Output files

psimpoll output may be written to up to 9 different output files. All are text files and can be read by any text editor or word processor (but beware of saving them in anything other than text format). The two PostScript files can be sent directly to PostScript devices for display or printing. The files are:

`.ps` main output file: this is written in PostScript page description language, and contains all the instructions for plotting the diagram. It is not intended for reading by mere humans, but some details of the format of this file, for the curious, are given in Technical matters (p. 85);

`AD` age-depth conversion file: this contains the details of any age-depth conversion, with confidence interval estimates for ages and gradients when these have been calculated. The default file name used is `fileAD` where 'file' is the first four characters in the main input filename. A different name can be given in menu Eb (p. 41);

`AD.ps` age-depth plotted output file: this contains a graphical representation (PostScript format) of the details of any age-depth conversion, with confidence interval estimates for ages when these have been calculated. The default file name used is `fileAD.ps` where 'file' is the first four characters in the main input filename. A different name can be given in menu Ei (p. 41);

`STAT` data analysis file: this may contain the details of results of statistical analyses, zonation, principal components analysis, Fourier analysis, independent splitting of taxa. In some cases this output supplements or repeats information that is also given on screen (e.g., zonation). In other cases, all the results are in this file (e.g., Fourier analysis). The default file name used is `fileSTAT` where `file` is the first four characters of the main input file name. A different name can be given in menu Ec (p. 41);

`DAT` data output file: contains output of the current dataset, incorporating any changes made during the current run of **psimpoll** (e.g., conversion of depths to ages). See menu J (p. 50). The default file name used is `fileDAT` where `file` is the first four characters of the main input file name. A different name can be given in menu Eh (p. 41);

`.CFG` configuration file: see above;

`.SCR` script file: see above;

`ZONE` zonation file: zonation output for the desired number of zones, in a format (p. 26) for immediate input and use by **psimpoll**, is written to this file. An example (p. 107) is available.

`.LOG` log file (`psimpoll.LOG`): written to the same directory as the executable file, with details of some of the parameters used by **psimpoll** during a run. This file is mainly intended to help KDB track errors. It is overwritten with each run.

pscomb has two output files. The main output file as the same format as the **psimpoll** main output file. Additionally there is a log file (`pscomb.LOG`) for the same purpose as the **psimpoll** log file.

5.8 Item 4: First character

% data are presented as proportions, for plotting and labelling as percentages;

a data are pollen accumulation rate data (grains cm-^{2} year-^{1});

b data are as for c (below), but are to be converted to accumulation rate data. A `C14` file must be present;

c data are pollen concentration data by volume (grains cm-^{3});

g data are pollen concentration data by weight (grains g-^{1});

l data are values of the Lomb periodogram (the file was probably written from Fourier analysis within a previous run of the program: see menu Me (p. 69));

m data are non-pollen data, where sample thickness is large, and varies between samples. The data will be plotted as histograms. Each sample value must be the mid-value of the histogram, and must be followed by a value for the histogram width. For example, a sample extending from 310 cm to 320 cm should be given as `315 10`. If the sample width is the same for all samples, then this can be given more easily in menu Dg (p. 40). A label giving the x-axis units must follow each taxon name;

x data are non-pollen data. A label giving the x-axis units must follow each taxon name.

5.9 Item 4: Second character

a data presented by age (years);

d data presented by depth (cm);

f data presented as frequencies (cycles kyr-^{1}), probably output from Fourier analysis in a previous run of the program: see menu Me (p. 69));

m data presented by depth (m);

s data are surface samples, presented by sample numbers;

t data are surface samples, presented by sample labels.

5.10 Prefixes of taxon names for taxa with special input requirements

! Pollen rarefaction data (palynological richness: see Birks & Line 1992). The data here must include, immediately after the name, the count size of standardization, then three values per sample (estimate, lower limit, upper limit, in that order). All three values must be present, even if the sample is considered a missing value (by entering it as (by default) -1: see menu Bg (p. 35)). Plots are labelled 'E(Tn)', where n is a subscript giving the count size for standardization;

< (by itself, or a line beginning with <) Data for summary diagram. May, optionally, be followed (on the same line, and separated by spaces) by values for the grey level or colours to be used in shading the summary diagram. Must be followed, on the next line, by number of taxa in the summary, then the data for each, in the usual format (name, values). Summary counts as one taxon in input Item 2, however many types it includes. By default, taxa in this summary are plotted in a cumulative

style, with changing shades of grey at even intervals from solid black (first type, shade 0.0) towards white (last type, shade 1.0). So, with four taxa in the summary, the default grey shades are 0, 0.33, 0.67, and 1.0. Providing grey values after the < over-rides this default behaviour. Colours for each of the summary taxa can be specified by giving their colour identifiers after the <, each prefixed by a '-'; There can only be one summary diagram in an input file. **pscomb** can be used to combine summary diagrams from the output of multiple input files.

- Ignore this taxon;

* Data are pollen sums, to be printed as values (not labelled), and used to obtain confidence intervals on percentages. Values are printed in a single column, unless there are more than 50 samples or the option in menu Bd (p. 34) is selected, in which case the values are printed in two columns.

5.11 Prefixes that affect labelling and presentation of axes

\ Use as a first character to prevent the second character from being interpreted as indicating a particular taxon type. For example, '@' switches text between Symbol encoding and ISO Latin-1 encoding (see Technical matters (p. 85)), but is also used to indicate a charcoal:pollen 'ratio' (see below). If a taxon name is to be printed entirely in Symbol encoding, it should be prefixed with '\@';

Print this taxon at a different scale from the others. Useful with concentration and accumulation rate data;

& Charcoal data, labelled 'cm^2 cm^{-3}';

% Charcoal data, labelled 'cm^2 cm^{-2} $year^{-1}$';

' Charcoal data, labelled 'cm^2 g^{-1}';

@ Charcoal:pollen 'ratio', labelled 'cm^2 $grain^{-1}$';

+ accumulation rate data if first character of input Item 4 is 'c', or concentration data if first character of input Item 4 is 'a'. Use for a total curve of the other type of data from the main one for the dataset. Plots are labelled 'grains cm^{-3}' or 'grains cm^{-2} $year^{-1}$', as appropriate;

= Data are rate of change data between adjacent pairs of samples (so the last data value will always be missing, and should be given as (by default) -1: see menu Bg (p. 35)), in rate of change units per century. Values plot halfway between adjacent pairs of samples, and are labelled 'cent.$^{-1}$';

* Data are pollen sums, to be printed as values (not labelled);

: data are ratio values (dimensionless), labelled 'ratio'.

6 Format of associated input files

6.1 TS

Example (p. 105).

Item 1 Number of sediment units

Items 2 onwards (for each unit) The Troels-Smith code for the sediment in the form of sediment type (e.g., Ld), then the value for that type (an integer, 1–4, on the next line). Only types scored as 1-4 (not '+') can be shown. Then give the basal depth for the unit. Each unit will thus consist of between 1 and 4 types plus values, then a depth. Possible Troels-Smith symbols are: Ld, Lso, Th, Tbs, Sh, Tb, Ag, As, Ga, Gs, Gg, Dh, Dl, Tl, Lc, ptm, Dg, plus others found useful: Tuf (blank), Hc (Helen's clay), Hs (Helen's silt), Tep (solid black), los (laminated oil shales) and lcs (laminated clays and silt). Other symbols can be added as and when needed.

Items 3 onwards (for each unit) [**Optional**] A colour for the background of the unit, written as text, one colour per line. Any colour in the colour palette is allowed, though it is expected that Munsell colours will normally be used.

6.2 C14

Needed for converting depths to ages in radiocarbon years, or for calculating accumulation rates from concentrations. This file is searched for by default.

Example (p. 105).

Item 1 Number of radiocarbon ages

Items 2 onwards (for each age) Upper depth of dated sample, lower depth of dated sample, age, standard deviation of the age. Each value must be on a separate line. Each age plots as a block labelled with age ± standard deviation. If the upper and lower depths are the same (e.g., for an AMS date on material from a precisely known depth), a cross in a circle is plotted instead. This same symbol is also used to indicate age locations when plotting against age. The 'age' and 'standard deviation' are read in as text, not numbers. Ages that you might want included in the age column of a depth plot, but do not want used for calculation of ages from depths (e.g., a tentative tephra marker, or a reversed age) should be prefixed with a non-numeric character (e.g., present the age as '(4000)'). Give a single space character for the standard deviation to have it ignored and avoid the plotting of a '±' symbol.

6.3 CAL

Needed for converting depths to ages in calendar years derived from BCal[10] calibration), or for calculating accumulation rates from concentrations. This file is searched for if requested in menu Lj (p. 59) (as an alternative to the C14 file). Note that this file acts as a control file, listing the files that themselves contain the calibration results (CAL input files).

Example (p. 106).

Item 1 Number of calibrated ages

Items 2 onwards (for each age) Upper depth of dated sample, lower depth of dated sample, and the name of a file that contains the full calibrated results. Each value must be on a separate line. The use of calibrated ages is complex, and discussed more fully under menu L (p. 53). Each age plots as a block labelled with age, and (if available) ± standard deviation. If the upper and lower depths are the same (e.g., for an AMS date on material from a precisely known depth), a cross in a circle is plotted instead. This same symbol is also used to indicate age locations when plotting against age. It is not currently possible to include other, non-age, items in the CAL file in the way that is possible for the C14 file.

6.4 CAL input

These files should be probability distribution files output from BCal[11]. See Example (p. 106) for more details.

6.5 ZONE

Example (p. 107).

Item 1 Number of zones;

Items 2 onwards (for each zone) Labels, from the top downwards. Use a '_' on both sides of a section of text which should be printed in oblique style (e.g., '_Betula_-Gramineae'), and the same conventions for superscripts and subscripts as for taxa names (see menu I (p. 43)). Maximum length of label is 39 characters;

Items 3 onwards (for each zone) Basal depths or ages, except the lowermost. The top of the highest zone is assumed to be the highest sample, and the base of the lowest zone is assumed to be the same as the lowest sample of the dataset.

[10] See URL http://bcal.shef.ac.uk
[11] See URL http://bcal.shef.ac.uk

Items 4 onwards (for each zone) [**Optional**] A colour for the background of the zone, written as text, one colour per line. . Any colour in the colour palette is allowed. The colour will replace the default background colour for the diagram.

7 Running the programs

7.1 Contents

7.2 Running psimpoll

On DOS and UNIX, ensure that you are in the same directory as the program (normally `psimpoll`, then enter `psimpoll` at the prompt followed by <Enter>. On Macintosh, open the `psimpoll` folder, and select the `psimpoll` application. A name for the dataset file is then requested, and must be given. You will see a menu allowing several parameters to be changed. If a filename is included on the command line (`psimpoll filename`), **psimpoll** runs with that filename right through without allowing selections from the menu but using a `.CFG` file, if present without any more attention from the user (unless any menu options that themselves require prompts are set (e.g., K (p. 50), Ld.3 (p. 56)).

psimpoll commands are not generally case-sensitive: 'Y' = 'y'. Exceptions are the codes used to indicate Troels-Smith symbols and colours: these codes must be given exactly as indicated in the Format of associated files (p. 25) and in psimpoll.COL (p. 110), respectively. Titles, taxa names, etc, will appear exactly as you enter them (w.r.t. case). However, filenames and operating-system commands (such as those given through menu Ud (p. 80)) must follow operating system conventions and may or may not be case-sensitive: UNIX is, DOS/Windows is not.

In all menus, any program default settings are enclosed within '()', and current settings are enclosed within '[]'. Available menu options are detailed below.

At all menus, you must press <Enter> after the letter or number of your choice. This is a change from versions of **psimpoll** before **2.00**, and makes the program more portable.

psimpoll produces a main PostScript output file with a name that by default is `file.PS`, where `file` was the name of the input file, and (after making age-depth transformations) a PostScript file (default name `fileAD.PS`) that displays the age-depth relationship graphically. These output files are text files and can be easily moved around between computers, including mainframes, and via e-mail. They can also be edited: try looking at an output file with an editor (e.g., non-document mode of WordStar). In order to produce either diagram, an output device with a PostScript interpreter is needed. The output files can be printed from a PC connected to a LaserWriter with the command:

`PRINT filename` <Enter> (answer `COM1` if asked for a 'list device')

If running **psimpoll** on any PC connected to a LaserWriter, the step of producing an output file can be avoided by sending output directly to the

LaserWriter. This is done by changing the name of the output file to `COM1`: see menu Ea (p. 41).

PostScript files can be printed using the GhostScript and GhostView[12] family of programs on Windows, Apple, and Unix environments (and others).

7.3 Running pscomb

As with **psimpoll**, change to the directory that includes the program, and run it by entering or selecting `pscomb` [`filename1 filename2 ...`]. If the filenames are omitted, or cannot be given on the command line (e.g., on Macintosh), the names of input files can also be entered from menu Eb (p. 41). The program then offers a menu, where several options can be changed, and then reads in `filename1`, `filename2`, etc, sequentially until all have been read. The program checks each file to ensure that it is a PostScript file, and that it was written by **psimpoll 2.20** or later. Output is directed to a file as for **psimpoll**, and is in the same format. **pscomb** output is thus valid **pscomb** input, and so on, *ad infinitum*.

The same general comments about menus and production of plotted output apply as for **psimpoll** (described above).

[12] See URL http://www.cs.wisc.edu/.ghost/index.html

8 psimpoll menus

After you have entered a filename, **psimpoll** attempts to read that file, determining whether it is a **psimpoll** input file or a TILia-type (TILIA (ver. 2) general format file or a NOAA[13] `.p15` file)

If it is a **psimpoll** input file the title, number of taxa, and number of samples are read, and the main menu appears.

If it is a TILIA-type file, additional information is required. The following message will appear:

```
(Tilia general format detected)
Enter code for dataset type:
'%' for percentage data
'c' for concentration data
'r' for raw data (converted to percentages)
```

Answer appropriately, and a second prompt appears:

```
Enter code for sample type:
'd' for depths (cm)
'm' for depths (m)
'a' for ages
```

The main menu will then be displayed.

In the main menu, and the submenus, one of two actions will result from entering the key letter (or number) of your choice. A further selection will appear, the value of the item will change from one mode to another (described as 'toggling'), or the value will change through a series (described as 'scrolling').

[13] See URL http://www.ngdc.noaa.gov/paleo/paleo.html

8.1 Main menu

```
A Dimensions and scale factors for PostScript plots
B Scaling of taxa and samples
C Changing colours
D Diagram style
E Change or enable files
F Font family (Helvetica)                        [Helvetica]
G Size of selected font (7.1)                    []
H Depth or age range for plot (off)              [off]
I Add/change titles
J Print dataset to file
K Interactive plotting (off)                     [off]
L Conversion of depths to ages
M Data analyses
N Plot confidence intervals (off)                [off]
O Save configuration to a:\dallb.CFG
P Read from script file (off)                    [off]
U Utilities
X Run program
```

Select an item from this list by entering the key letter. Details of the result of each choice, together with the permitted values for any modifiable parameters, may be found by following the links.

8.2 Menu A: scale plot width

```
A Dimensions and scale factors for PostScript plots
a Scale factor for main output file (1.0)        [1.00]
b Scale factor for age-depth output file (1.0)   [1.00]
c Length of print area (mm) (240)                [240]
d Height of diagram (mm) (127)                   [127]
e Paper size (A4)                                [A4]
9 Leave this menu
```

The layout of a plotted page is discussed in Technical matters (p. 85). This menu changes two aspects of the shape of the diagram: scale and dimensions. Scaling is used to adjust plots so that they can be stretched or compressed by a desired amount (e.g. to fit a particular number of pages).

8.2.1 Menu Aa: main output file

The widths of columns for sediment units, zones, radiocarbon ages, and sum values are not affected;

[Allowed range 0.1-10].

8.2.2 Menu Ab: age-depth output file

A value of 1 fills the page completely, and is therefore the maximum allowed. Smaller values may be especially useful if the resulting plot is going to be combined with other data via **pscomb**.

[Allowed range 0.1-1].

8.2.3 Menu Ac: length

Changes the length of the diagram, and then also changes the height and font size to maintain the same proportions. I suggest using 240mm with A4 paper (the default), 339 mm with A3, 226 mm with US Letter, and 287 mm with US Legal. Paper height changes automatically, but can be over-ridden: see Menu Ad (p. 32). Use with caution;

[Allowed range 10-420].

8.2.4 Menu Ad: height

Changes the height of the diagram, measured along the length of the plotted curve, independently of the diagram length. Ordinarily, height changes automatically when changes are made to paper width (see Menu Ac (p. 32)), but this can be over-ridden here. Suggested sizes are 127 mm for A4 (the default), 180 mm for A3 and 131 mm with US Letter and US Legal, to allow room for labelling. Use with caution;

[Allowed range 10-297].

8.2.5 Menu Ae: paper size

Allows change of paper size, currently between A4 paper (297x210mm), A3 (420x297mm), US Letter (279x216mm) and US Legal (356x216). Following selection of a size, diagram width and diagram height are given default values. These can then be overridden in Menu Ac (p. 32) and Menu Ad (p. 32) respectively;

Scrolls between 'A4', 'A3', 'Letter' and 'Legal'.

8.3 Menu B: diagram scaling

```
B Scaling of taxa and samples
a Independent scaling for each taxon (off)       [off]
b Scaling covers excluded samples (off)          [off]
c Use defaults for marks on age/depth scale (on) [on]
d Use two columns for sum values (off)           [off]
e Skip every nth sample (0)                      [0]
f Interpolate to constant sample intervals (off) [off]
g Threshold for missing data values (-0.99)      [-0.99]
9 Leave this menu
```

Allows changes to factors that control various scaling properties:

8.3.1 Menu Ba: independent scaling

By default, a scale is found that will work for all taxa in the dataset, except that for those specifically marked as separate (e.g., charcoal, rarefaction data). This option disables the search for a common scale, and searches for a separate scale for each taxon;

Toggles from 'on' to 'off'.

8.3.2 Menu Bb: scale excluded samples

By default, if samples have been excluded, either because they were marked with an 'x' (see Data preparation (p. 16)) or because they are outside a section of the sequence that has been selected (menu H (p. 43)), then data values in those samples are not considered when scaling plots. This option allows the data values in excluded samples to be considered. It means that when sections of the diagram are picked out, each such section will have plots with the same scaling as the others;

Toggles from 'on' to 'off'.

8.3.3 Menu Bc: labelling on age-depth scale

By default, **psimpoll** determines an appropriate scale with major and minor marks for the age-depth axis. This option enables the user to turn off the default and specify an interval for both major and minor marks. Interval must be given in the same units as those in effect for the age-depth scale (allowing for any conversion to ages). If the default is turned off, the current settings for major and minor marks are shown, and then prompts for new values;

8.3.4 Menu Bd: double column for sums

Values for pollen sums, identified by marking the taxon 'name' with '*', are normally printed as a single column with upto 50 samples, but as a double column for more than 50 samples. Selecting this option forces double-column printing however many samples there are;

Toggles from 'on' to 'off'.

8.3.5 Menu Be: skip every nth sample

Allows omission of samples, regularly spaced. This is included to permit exploration of the sensitivity of certain analyses (notably zonation) to the number of samples in the dataset. A value of '2' causes every other sample to be selected for omission, '3' every third sample, etc. A negative value causes the selected samples to be included, and the rest omitted: a value of -3 causes only every third sample to be included. Thus, '2' and '-2' would both select every other sample, but the two series would have alternating samples. Values of '1' and '-1' are not allowed;

[Allowed range -100-100].

8.3.6 Menu Bf: interpolate to constant sample intervals

Selecting this item produces the following display:

```
6 Interpolate to constant sample intervals (off)  [on]

Enter '0' for on, or '1' for off
```

Allows a new dataset to be generated, based on the input data set, but with interpolated samples evenly spaced. The values at the interpolated samples are based solely on the pair of original samples either side of the interpolated sample (or on one original sample if there is one with exactly the same age or depth value as the desired interpolated sample). If 0 is selected, you will see the following submenu:

```
A Interpolation interval in age-depth units (1)   [100]
B Start level for interpolation (0)               [0]
C Stop level for interpolation (0)                [10000]
D Interpolation method (nearest pair)             [nearest pair]
9 Leave this menu
Q Return to main menu
```

Here you can select the interpolation interval, using the same units as the input dataset or age units (if a conversion to age is being done), specify where

the interpolation should start and stop, and determine the interpolation method. The two available methods are 'nearest pair', and 'line fitting'. The first finds a pair of samples that lie either side of the interpolated level (however far away they are), and interpolates between them. The second uses all levels upto half of the interpolation interval either side of the interpolated level, fits a line through these points, and obtains the desired value for the interpolated level from the line. For the line-fitting method only, if there is only 1 (or 0) level available, the value for the interpolated level is given as for a missing data (-1, by default: see Menu Bg (p. 35)).

8.3.7 Menu Bg: threshold for missing data

Allows change to the value of the negative number used to identify missing values. By default, this is -0.99. Anything more negative than this (e.g. -1) is used to mark a missing value. The value can be changed to any other negative number. This may, however, cause undesirable results for existing data sets that use -1, so use with caution.

[Allowed range: any negative real number].

8.4 Menu C: changing colours

This menu has two forms, one for each of the two PostScript output files, main (p. 35) and age-depth (p. 37). Toggling item `a` switches between the two forms.

Available colours for any part of either diagram are those stored in the colour palette.

8.5 First form: for main output file

```
C Changing colours
a Output file (main)                              [MAIN]
b Background colour (white)                       [white]
c Foreground colour (black)                       [black]
d Colour of plots (black)                         [black]
e Colour of title (black)                         [black]
f Colour of age-depth axis (black)                [black]
g Colour of 14C column (black)                    [black]
h Colour of sediment column foreground (black)    [black]
i Colour of zone column foreground (black)        [black]
j Colour of x-axes (black)                        [black]
k Use colours for sediment units (on)             [on]
l Use colours for zones (on)                      [on]
9 Leave this menu
```

8.5.1 Menu Ca: background colour

Toggles from colours for main output file to colours for age-depth output file.

8.5.2 Menu Cb: background colour

Colours the entire background of *both* plots. Default is 'white', which does not print.

8.5.3 Menu Cc: foreground colour

Colour of 'ink' for the lines and lettering of the plot, establishing the default that may be over-ridden for any particular part. Default is 'black'.

8.5.4 Menu Cd: colour of plots

Colour of the foreground of plots. Colours can also be selected for individual types: see menu K (p. 50). Default is 'black'.

8.5.5 Menu Ce: colour of title

Colour of the title. Default is 'black'.

8.5.6 Menu Cf: colour of age-depth axis

Colour of the foreground of the age-depth axis, at both sides of the plot. Default is 'black'.

8.5.7 Menu Cg: colour of the 14C column

Colour of the foreground of the 14C column, including the colour used to fill the blocks marking the position of dated samples.

8.5.8 Menu Ch: colour of the sediment column foreground

Colour of the foreground of the sediment column: the 'ink' used to mark the Troels-Smith symbols. Colours for the background of each sediment unit are read from the `TS` file, and are used if the switch at menu Ck (p. 37) is 'on'.

8.5.9 Menu Ci: colour of the zone column foreground

Colour of the foreground of the zone column: the 'ink' used to mark the labels and zone lines. Colours for the background of each zone are read from the `ZONE` file, and are used if the switch at menu Cl (p. 37) is 'on'.

8.5.10 Menu Cj: colour of x-axes

Colour of the foreground of all the x-axis scales and labelling of the diagram.

8.5.11 Menu Ck: use colours for sediment units

Controls whether the colours for sediment units, read from the `TS` file, are used. These colours fill the background of the sediment column, replacing the main background colour (menu Cb (p. 36)).

Toggles from 'on' to 'off'.

8.5.12 Menu Cl: use colours for zones

Controls whether the colours for zones, read from the `ZONE` file, are used. These colours fill the background across the entire width of the diagram, replacing the main background colour (menu Cb (p. 36)). The coloured zones are (invisibly) divided into blocks corresponding to the part of each zone that forms the background for each taxon. This means that a taxon selected by **pscomb** from a plot with coloured zones will also have a coloured background.

Toggles from 'on' to 'off'.

8.6 Second form: for age-depth output file

```
a Output file (main)                                [AGE-DEPTH]
b Background colour (white)                         [white]
c Colour of main curve (black)                      [black]
d Colour of confidence interval curves (black)      [black]
e Colour of points (black)                          [black]
f Colour of depth axis (black)                      [black]
g Colour of x-axis (black)                          [black]
9 Leave this menu
```

8.6.1 Menu Ca: background colour

Toggles from colours for age-depth output file to colours for main output file.

8.6.2 Menu Cb: background colour

Colours the entire background of *both* plots. Default is 'white', which does not print.

8.6.3 Menu Cc: colour of main curve

Colour of the main plotted curve. Default is 'black'.

8.6.4 Menu Cd: colour of confidence interval curve

Colour of both confidence interval curves. Default is 'black'.

8.6.5 Menu Ce: colour of points

Colour of the points (crosses) marking the location of data used in the calculation of the curve. Default is 'black'.

8.6.6 Menu Cf: colour of depth axis

Colour of the y-axis, labelled with the depth scale. Default is 'black'.

8.6.7 Menu Cg: colour of x-axis axis

Colour of the x-axis, labelled with the calculated time-scale. Default is 'black'.

8.7 Menu D: diagram style

```
D Diagram style
a <Experimental> Tufte style (off)                [off]
b Diagram type (outline)                          [outline]
c x10 exaggeration (off)                          [off]
d Shade of grey for filling plots (0)             [0.000]
e Angle of rotation for taxa names (45)           [45]
f Join-up across missing levels (off)             [off]
g Width of bars for histograms (0)                [0.0]
h Place a dot next to low values (off)            [off]
i Upper value for placing dot (0.5)               [0.5]
j Include radiocarbon age locations (on)          [on]
k Factor to adjust width of lines (1)             [1.0]
l Resolution of curve in age-depth plots (1000)   [1000]
9 Leave this menu
```

Changes aspects of the style of the diagram:

8.7.1 Menu Da: Tufte style

Experimental approach to the layout of pollen diagrams, following the ideas of Tufte(1983). Incomplete;

Toggles from 'on' to 'off'.

8.7.2 Menu Db: type

The diagram type may be 'outline' (lines along curves, with data points connected back to the base line), or 'filled' (shaded between curve and baseline). Curves plotted with confidence intervals (including rarefaction values) cannot be plotted 'filled'. This option can be selected for individual types: see menu K (p. 50);

Toggles from 'outline' to 'filled'.

8.7.3 Menu Dc: × 10 exaggeration

Adds a curve at × 10 beyond the principal curve. This option can be selected for individual types: see menu K (p. 50);

Toggles from 'on' to 'off'.

8.7.4 Menu Dd: shading

Allows selection of the grey shade used in plots that are filled. This option has no effect if 'outline' plots are selected (menu Db (p. 39)). The available shades run continuously from 0.0 (black) to 1.0 (white). Plots are outlined in black, so completely white will be 'visible'. This option can also be selected for individual types: see menu K (p. 50);

[Allowed range 0-1].

8.7.5 Menu De: name angle

Changes the angle for names of taxa and headings, in degrees anticlockwise from horizontal;

[Allowed range 0-90].

8.7.6 Menu Df: join-up across samples

Controls whether curves are joined up across missing samples;

Toggles from 'on' to 'off'.

8.7.7 Menu Dg: histogram width

Changes the width of histogram bars, using the same width for all samples. This should be used merely as a graphical device, except in the rare situation that all samples really did have the same width. For varying histogram widths, the widths have to be entered with the sample values (using 'm' or 'x' as the first character of the two-character code in Item 4 of the input (see Data preparation (p. 16)). The units for histogram width are the same as the units of the samples, allowing for any conversion to ages;

[Allowed range 0-1000].

8.7.8 Menu Dh: bullet

Adds a solid dot next to low values that are non-zero but less than a default value of 0.5% (or equivalent after scaling for non-percentage data). The default value can be changed from menu Di (p. 41). These symbols are not placed by certain special curves (such as sums, rarefaction, rate-of-change, charcoal, etc). This option can also be selected for individual types: see menu K (p. 50);

Toggles from 'on' to 'off'.

8.7.9 Menu Di: bullet threshold

Changes the value of the upper limit for placing a solid dot next to a curve (see menu Dh (p. 40)). The value can also be selected for individual types: see menu K (p. 50);

[Allowed range 0-1,000,000].

8.7.10 Menu Dj: Include radiocarbon age locations

By default, the locations of radiocarbon age determinations are plotted along side the left hand age or depth scale (if they are available). This option toggles this feature off.

Toggles from 'on' to 'off'.

8.7.11 Menu Dk: Factor to adjust width of lines

Factor to vary widths of lines from the default (taken to be 1). This value is device dependent, and some experimentation may be needed to achieve the desired output. High-resolution devices are capable of finer lines, and the line width may need to increased considerably before there is much noticeable difference in the output;

[Allowed range 0-100].

8.7.12 Menu Dl: Resolution of curve in age-depth plots

The graph in age-depth plots is made up of a series of short straight line segments. This option controls how many there are. The default is 1000, but this may be time-consuming to generate. Fewer segments means lower resolution: the graph may become visibly step-like. More segments means a smoother curve.

[Allowed range 10-10000].

8.8 Menu E: files

```
E Change or enable files
a Main PostScript output file                          [a:\dallb.ps]
b Output file for age-depth conversion                 [a:\dallAD]
c Output file for data analyses                        [a:\dallSTAT]
d Input file with zone details                         [a:\dallZONE]
e Input file with 14C ages                             [a:\dallC14]
f Input file for BCal calibrated ages                  [a:\dallCAL]
g Input file with sediment description                 [a:\dallTS]
h Output file for data                                 [a:\dallDAT]
i PostScript output file for age-depth plot            [a:\dallAD.ps]
9 Leave this menu
```

Selecting Ea-Ei allows you to add or change the name of an output or input file. Defaults based on the input file name are used (here based on an input file called a:\dallb).

File names given under Eb, Ec, Eh and Ei are only used if appropriate options are requested in menus L (p. 53), M (p. 59), and J (p. 50).

File names given must be legal under the operating system you are using (DOS, UNIX, whatever). If the input files cannot be found, the program will continue without using them. If output files already exist, they will be overwritten with new data.

For Ea, when using a PC connected to a LaserWriter, the main output file name may be given as `COM1`, which causes output to be directed straight to the printer.

8.9 Menu F: font family

```
F Font family (Helvetica)                              [Helvetica]

Serif fonts:                                           Sans serif fonts:
B ITC Bookman                                          A ITC AvantGarde
C Courier                                              H Helvetica
N New Century Schoolbook                               E Helvetica Narrow
P Palatino
T Times

Enter letter indicating new font name
```

Choose one of the letters to determine which font family you want to use for labelling diagrams. Plain text style is used by default, but italics can be selected by using a '_' before and after the text concerned (see Data preparation (p. 16)). It is normal to use a *sans serif* font for labelling

diagrams intended for publication.

8.10 Menu G: font size

```
G Size of selected font (7.1)                    []
May cause text and diagram to overlap
Enter new font size (units of 1/72 inches) <Enter>
```

Font size is adjusted depending on the chosen font. Use this option to make further changes. The first figure given is the default for the current font, and the figure in the square brackets is any value selected by the user to over-ride the default. You may get undesirable effects from a change to a size which is too large or too small. The size chosen applies to taxa names and labelling text. The title is larger by a factor of 1.73. Font size is also adjusted proportionately to the length of diagram selected in menu Ac (p. 32).

[Allowed range 0-36]

8.11 Menu H: extreme values for depth or age in plots

```
H Depth or age range for plot (off)              [off]
Enter '0' for on, or '1' for off
```

Here you can specify minimum and maximum values for depths or ages of samples that you would like to select for plotting. After entering 0, you will be prompted first for a minimum value, and then a maximum value. These values must be in the units (age or depth) that will appear on the plot. Thus, Dallican Water sediments have a depth range of 390 cm to 720 cm, so the depth range 400-600 can be selected for depth plots. But if the depths are being converted to ages by **psimpoll**, then the values selected would have to be in radiocarbon years (3000-7000, for example), and if the ages are to be plotted in calendar years AD or BC (menu Li (p. 58)), then the values selected would have to be negative (years AD) or positive (years BC) as appropriate.

The program checks that the maximum value given is greater than the minimum value, but nothing more. Values given can be part of the range of the dataset, or overlap it. Thus, it is possible to obtain plots from different sites with the same age or depth scale to facilitate comparison. Datasets for numerical analyses (menu M (p. 59)) are affected by this option. Analyses will be carried out only on the selected portion of the sequence.

The allowed range for both maximum and minimum values is -2000-10,000,000.

8.12 Menu I: labelling

This menu has allows access to different headings depening on the settings from menu L (p. 53), affecting items e, f, g, h ,i. The default, below, shows the menu with default settings in menu L (p. 53) (i.e. plotting by depth). The alternatives appear if plotting against age menu La (p. 54) by radiocarbon dates or by calibrated dates menu Li (p. 58). These must be set first to give access to the appropriate headings here.

```
I Add/change titles
a Label for base of diagram                    []
b Subtitle                                     []
c Zonation heading                             []
d Dendrogram heading                           []
e Sediment column heading                      []
f Age column heading                           []
g Age-depth column heading (line 1)            [Depth]
h Age-depth column heading (line 2)            [cm]
i Timescale column heading                     []
j Age-depth plot x-axis                        []
k Age-depth plot y-axis                        [Depth (cm)]
l Modify graph unit labels
9 Leave this menu
```

Adds or modifies pieces of text around the diagram. All are printed in the same font and at the same size as taxa names. All text on the diagram can be controlled by the user, either through input within any of the input files or from this menu. All characters alphabet are available, plus the Greek alphabet. The following characters may be used to achieve useful effects in any label (including text read from input files):

_ Use to mark sections (beginning and [optionally] end) of a section of text that you wish to appear in oblique style. The default font at the start of printing each label is roman style, and this default comes into effect for each one: there is therefore no need to cancel the oblique style at the end of a label, (e.g., '_Betula');

{} Use to bracket any piece of text that should appear as a superscript;

[] Use to bracket any piece of text that should appear as a subscript;

\ Use to modify the preceding character with a following character, to enable use of accented characters and a few special characters (see Table (p. 47)). These characters should cover all those used by European languages: please inform KDB of any missing characters that should be included;

@ Use to bracket sections of text that should appear with Symbol encoding (see Technical matters (p. 85)). This is intended to enable the use of Greek characters, but may be useful for certain other effects.

Equivalence between lower case plain text characters and the Greek alphabet is shown in a separate Table (p. 48). Other symbols are also available, including mathematical symbols, arrows, etc., and these can be included by entering codes directly: see Table (p. 48);

\# Switches font to normal size, roman style. Used (automatically) between title and subtitle in **psimpoll**, but may be useful in **pscomb** titles to produce the same effect;

$ Use to include the name of the main input file within a label. I like this incorporated in the subtitle (I1) for test plots;

| Use to include the time and date within a label. I like this incorporated in the subtitle (I1) for test plots.

The labels that can be modified are:

Ia Text at base of diagram e.g., '`Percentages of total ...`'. Default is '`%`' for percentage diagrams, but nothing for other types of plot;

Ib Subtitle, appended to the main title. Default is nothing;

Ic Labelling for column of zonation (if any), printed centrally, at the same angle as for taxa names. Default is nothing;

Id Labelling for dendrogram produced by CONISS or CONIIC (if any), printed centrally, at the same angle as for taxa names. Default is nothing;

Ie Labelling for column of sediment stratigraphy (if any), printed centrally, at the same angle as for taxa names. Default is nothing;

If Labelling for column with radiocarbon ages (if any), printed centrally, at the same angle as for taxa names. Defaults depend upon settings in menu L (p. 53):

La = 'off'; second character of code on fourth input line is not 'a'; (p. 23) No value (unused)

La = 'on' or second character of code on fourth input line is 'a'; Li = 'off'; Lh = 'off' (p. 58) Default '`{14}C ages (yr BP)`' (which prints with a superscripted '14').

La = 'on' or second character of code on fourth input line is 'a'; Li = 'off'; Lh = 'on' (p. 58) Default '`{14}C ages (yr AD/BC)`' (which prints with a superscripted '14').

La = 'on' or second character of code on fourth input line is 'a'; Li = 'on'; Lh = 'off' (p. 58) Default '`Cal. ages (yr BP)`'.

La = 'on' or second character of code on fourth input line is 'a'; Li = 'on'; Lh = 'on' (p. 58) Default '`Cal. ages (yr AD/BC)`'.

Ig First line of label for age-depth column printed centrally and horizontally. Defaults depend upon settings in menu L (p. 53):

La = 'off'; Li = 'off' (p. 58) Depends upon the second character of code on fourth input line (p. 23):

a 'Age';

d, m 'Depth';

s 'Sample';

f 'Cycles';

t Default is no label;

La = 'on'; Li = 'off' or 'on'; Lh = 'off' or 'on' (p. 58) Default 'Age'.

Ih Second line of label for age-depth column printed centrally and horizontally. Defaults depend upon settings in menu L (p. 53):

La = 'off'; Li = 'off' (p. 58) Depends upon the second character of code on fourth input line (p. 23):

a 'yr BP' or yr AD/BC, depending on whether AD/BC option has been selected (see menu Lh (p. 58));

d 'cm';

f 'kyr{-1}' (prints with superscripted '-1')';

m 'm';

s, t Default is no label;

La = 'on'; Li = 'off' or 'on'; Lh = 'off' (p. 58) Default 'yr BP'.

La = 'on'; Li = 'off' or 'on'; Lh = 'on' (p. 58) Default 'yr AD/BC'.

Ii Labelling for time-scale column heading (if any), printed centrally, at the same angle as for taxa names. Defaults depend upon settings in menu L (p. 53):

Lh = 'off' (p. 58) No value (unused)

Lh = 'on'; Lj = 'off'; Li = 'off' (p. 58) Default 'Age ({14}C yr BP)' (which prints with a superscripted '14').

Lh = 'on'; Lj = 'on'; Li = 'off' (p. 58) Default 'Age (Cal. yr BP)'

Lh = 'on'; Lj = 'off'; Li = 'on' (p. 58) Default 'Age ({14}C yr AD/BC)' (which prints with a superscripted '14').

Lh = 'on'; Lj = 'on'; Li = 'on' (p. 58) Default 'Age (Cal. yr AD/BC)'

Ij Labelling for x-axis of age-depth plot, printed centrally. Defaults depend upon settings in menu L (p. 53):

Lh = 'off' (p. 58) No value (unused)

Lh = 'on'; Lj = 'off'; Li = 'off' (p. 58) Default 'Age ({14}C yr BP)' (which prints with a superscripted '14').

Lh = 'on'; Lj = 'on'; Li = 'off' (p. 58) Default 'Age (Cal. yr BP)'

Table of accented and other special characters for use in labels
What to type in order to have the indicated character displayed on the final diagram. Examples are shown where this is possible within the limitations of LaTeX.

Type: to get:
`a\'` á acute accent on any letter
`a\u` ă breve accent on any letter
`z\v` ž caron accent on any letter
`c\,` ç cedilla with any letter
`a\^` â circumflex on any letter
`a\"` ä dieresis on A, a, E, e, I, i, O, o, U, u
`a\‘` à grave accent on A, a, E, e, I, i, O, o, U, u
`a\-` ā macron accent on any letter
`e\;` ę ogonek with any letter
`n\~` ñ tilde on A, a, N, n
`o\e` œ French ligature (lowercase)
`O\E` Œ French ligature (uppercase)
`s\s` ß German sharp S
`o\*` ő long Hungarian umlaut on any letter
`d\h` ð Icelandic eth
`D\h` Ð Icelandic Eth
`p\h` þ Icelandic thorn
`P\h` Þ Icelandic Thorn
`l\/` ł Polish suppressed-l
`L\/` Ł Polish suppressed-L
`a\e` æ Scandinavian ligature
`A\E` Æ Scandinavian ligature
`a\o` å Scandinavian a-ring
`A\o` Å Scandinavian A-ring
`o\/` ø Scandinavian o-slash
`O\/` Ø Scandinavian O-slash
`\251` © Copyright symbol
`\261` ± Plus-minus symbol

Lh = 'on'; Lj = 'off'; Li = 'on' (p. 58) Default '`Age ({14}C yr AD/BC)`' (which prints with a superscripted '14').

Lh = 'on'; Lj = 'on'; Li = 'on' (p. 58) Default '`Age (Cal. yr AD/BC)`'

Ik Labelling for y-axis of age-depth plot, printed centrally, rotated vertically. Default '`Depth (cm)`';

Il Modifies the labels for the units of plots, affecting both general footnotes (menu Ia (p. 45)), and labelling of individual curves. After selecting this option, a submenu (p. 48) allows a choice of labels for modification.

Table of Greek equivalents of English alphabet characters in Symbol encoding for use in labels
What to type in order to have the indicated character displayed on the final diagram. Examples are shown where this is possible within the limitations of LaTeX, or names of characters otherwise.

Type:	to get:
A a	A α
B b	B β
G g	Γ γ
D d	Δ δ
E e	E ϵ
Z z	Z ζ
Q q	Θ θ
I i	I ι
K k	K κ
L l	Λ λ
M m	M μ
N n	N ν
X x	Ξ ξ
O o	O o
P p	Π π
R r	R ρ
S s	Σ σ
T t	T τ
U u	Υ υ
F f	Φ ϕ
C c	X χ
Y y	Ψ ψ
W w	Ω ω

8.12.1 Menu Ik: submenu for choice of labels to be modified

```
A Concentration (volume) units                  [ grains cm{-3}]
B Concentration (weight) units                  [ grains g{-1}]
C Accumulation rate units                       [ grains cm{-2} year{-1}]
D Charcoal concentration units (volume)         [ cm{2} cm{-3}]
E Charcoal accumulation rate units              [ cm{2} cm{-2} year{-1}]
F Charcoal concentration units (weight)         [ cm{2} g{-1}]
G Charcoal:pollen units                         [ cm{2} grain{-1}]
H Rate of change units                          [ cent.{-1}]
I Ratio label                                   [ ratio]
J Fourier output values                         [ power]
K Total dispersion                              [ Total dispersion]
9 Leave this menu
```

Select the letter or number concerned to change a label. The new label can

Table of character codes in Symbol encoding for use in labels

What to type in order to have the indicated character displayed on the final diagram. Examples are shown where this is possible within the limitations of\ LaTeX, or names of characters otherwise.

Type:	to get:		
”	∀	’	∋
\243	≤	\245	∞
\247	♣	\250	♢
\251	♡	\252	♠
\253	↔	\254	←
\255	↑	\256	→
\257	↓	\260	°
\261	±	\263	≥
\264	×	\265	∝
\266	∂	\267	•
\270	÷	\271	≠
\272	≡	\273	≈
\274	. . .	\275	|
\276	—	\277	carriage return
\300	ℵ	\301	ℑ
\302	ℜ	\303	℘
\304	⊗	\305	⊕
\306	∅	\307	∩
\310	∪	\311	⊃
\312	⊇	\313	⊅
\314	⊂	\315	⊆
\316	∈	\317	∉
\320	∠	\321	∇
\322	registered serif	\323	copyright serif
\324	trademark serif	\325	∏
\326	√	\327	·
\330	¬	\331	∧
\332	∨	\333	⇔
\334	⇐	\335	⇑
\336	⇒	\337	⇓
\340	⋄	\341	⟨
\342	registered sans serif	\343	copyright sans serif
\344	trademark sans serif	\345	∑
\361	⟩		

include any of the usual effects, as described in menu I (p. 43).

8.13 Menu J: print dataset

```
J Print dataset to file
a Print dataset to file (off)                    [off]
b Specify format of output                       [psimpoll]
9 Leave this menu
```

Controls whether to print the main dataset to the data output file (menu Eh (p. 41)) in order to provide output of values calculated during a run of **psimpoll**. The dataset may be in a form to be used immediately as **psimpoll** input, or for input to WACALIB (Line *et al* 1994). It includes all changes made during the current **psimpoll** run (conversion of depths to ages, concentrations to accumulation rates, calculation of palynological richness and rates of change, sums calculated during conversion of TILIA raw data files to percentages).

If radiocarbon dates have been read from a TILIA raw data file, these will be written to a C14 file in the right format to be read in to **psimpoll**. However, an existing C14 file will not be over-written.

Select 'a' to toggle printing of a dataset between 'on' and 'off', and select 'b' to determine the dataset type.

8.14 Menu K: interactive plotting

```
K Interactive plotting (off)                     [off]
```

Controls whether the user can select taxa individually for plotting (on), or taxa are plotted in the order found in the data file (off);

Toggles from 'on' to 'off'.

When running under this option, the program provides a screen listing of all taxa in the dataset, each preceded by a number. Any initial '-' (to mark a taxon that should be skipped by non-interactive plotting) is removed. The first six characters of each taxon name are given where the dataset includes less than 110 taxa, the first three characters for less than 153 taxa, and the number only for less than 237 taxa. The option is not available for 238 or more taxa (because there is not enough room on the screen). Previously 'plotted' taxa are marked by a hyphen between the taxon name and its number, and listed by number beneath the display of taxa names.

For example:

```
Dallican Water, Lunnasting, Shetland

 0-_Betul   1-_Coryl   2-_Pinus   3 _Querc   4 _Ulmus   5 _Alnus   6 _Fraxi
 7 _Junip   8 _Salix   9 _Empet  10-_Callu  11 Cypera  12 Gramin  13 _Filip
14 _Poten  15 _Artem  16 Crucif  17 Caryop  18 _Jasio  19 _Rumex  20 _Polyp
21 _Pteri  22 Filica  23 _Huper  24 :AP:NA  25 <>      26 !Palyn  27 =Rate
28 *Main   29 Potamo  30 _Isoet  31 _Isoet  32 _Myrio  33 *Aquat  34 _Sphag
35 *Sphag  36 *Indet

Enter 'Q', 'N', or 'number[Bx][Cx][D][E][F][Gx][Hx][O][Px][Sx][T]'

Taxa plotted so far:                                  0.50 pages plotted
0 1 2 10
```

You select a taxon for plotting by entering its number, or a number range, followed [optionally] by the characters indicated above in square brackets. Any other characters, including spaces, are ignored. Effects are cumulative along the sequence if contradictory options are given. For example, `0fo` is equivalent to `0o`. A number range is given by entering `n1-n2`, where `n1` and `n2` are the first and last numbers of the desired range. The optional characters enable you to over-ride defaults set for the dataset for a particular taxon (or all taxa in a range):

B plot a solid dot next to the curve where values are non-zero but less than a default value of 0.5% (or equivalent after scaling for non-percentage data). A value 'x' can be added to change the default value. This should be in percentage units or the same units as the taxon being plotted;

Cx Change colour for current plot. The value given after the letter must be either a number which is a colour identifier from the colour palette file (p. 16) (e.g. `C142`), or a question mark (`C?`). If the identifier is given, the corresponding colour is used. If a question mark is given, a prompt displays the current colour, and a asks for a new one (which should be the name of a colour in the palette file);

D switch off × 10 exaggeration (menu Dc (p. 39));

E switch on × 10 exaggeration (menu Dc (p. 39));

F switch on plotting of filled curves (menu Db (p. 39)), using the grey shade of filling selected in menu Dd (p. 40), or solid as default. If a colour has been selected, the curve will be filled in that colour mixed with an amount of black according to the value of the grey shade;

Gx switch on plotting of filled curve and selection of grey shade (menu Dd (p. 40)), with 'x' set to a real number. Available shades range from 0.0 (black) to 1.0 (white). The value is given immediately after the letter (e.g., `G0.75`. The way that decreasing grey shades for the default summary diagram are calculated is given in Data preparation (p. 16),

and can be extended to the taxa that make up the summary types by using this option. Shaded plots are outlined, so pure white is visible even with a white background. If a colour has been selected, the grey value is added to the colour. A value of 0.0 always produces a black plot;

Hx switch on plotting of current curve as a histogram. The value for the histogram width is given immediately after the letter (e.g., `H0.25`), in the current age-depth units. This value over-rides any settings made in menu Dg (p. 40). This option is ignored if applied to a sum value. Sum values can be plotted as histograms if the '*' that defines them as sums is removed from the taxon name. This can be done by combining a histogram switch with a change of taxon name (e.g. `H0.25T`);

O switch on plotting of outline curves (menu Db (p. 39));

Px Change the line thickness for the current plot. The value is given immediately after the letter (e.g., `P10`). The default value is 1, unless changed in menu Dk (p. 41). The result is dependent on the resolution of the output device, so some experimentation may be needed to get a desired effect;

S smooth the data before plotting. The default is 3-point smoothing by fitting a regression line (2-term polynomial) through each data point, plus one point on either side. If a value is given for 'x', immediately after 'S', then that value is used for the number of points to smooth across. 'x' should be an odd number. If it is 5 or more, then a 4-term polynomial (cubic) is fitted to the points. As with any smoothing, data points are lost at both ends of the sequence. The smoothed data replace the original values in memory, so do not use this option with the same taxon more than once per run of **psimpoll**. Smoothed values are printed to the data analysis file, if this is available. Values at the ends of the sequence appear as '-1'. If confidence intervals have been requested (menu N (p. 75)), these will not be calculated;

T change the 'name' of the taxon. This is intended for use in the production of PostScript output files that will be used in **pscomb**. For example, if you want a plot that compares the *Betula* record from several sequences, you might produce a PostScript output file using this option to replace the name '*Betula*' by the site name, and then add a main title in **pscomb** which indicates that all the plots are of *Betula* from the named sites.

X extend the plotting limits for the taxon. By default, these are the maximum and minimum values found in the data for the taxon concerned. These values can be extended to higher and lower values, respectively. This feature can be used to produce plots of a given taxon to the same scale as plots for the same taxon in other datasets, for example. The code 'X' is saved in the .SCR file, but the new values are not (and so will be prompted for even when reading from a script file).

Alternatively, instead of a number, enter 'N' to force the start of a new page, or 'Q' to end the plot. As usual, **psimpoll** is not case-sensitive about any of these characters.

By default, all options selected are written to a `.SCR` file. If option P (p. 79) was selected from the main menu, then options will be read from a previous `.SCR` file, and not from the keyboard.

Setting interactive plotting to 'off' does not necessarily make **psimpoll** completely non-interactive. Some options from other menus require action from the user. These are:

- Age-depth modelling by least-squares (L4.3 (p. 56)) or SVD (L4.4 (p. 56)) with the number of terms set at 0 (L5 (p. 57));
- Zonation (Mc (p. 62)) with the desired number of zones set at 0 (Mc.1 (p. 64));
- Data analyses with the threshold for inclusion of a taxon set to be <=0 (Mi (p. 74)).

8.15 Menu L: age-depth conversion

Enables calculation of sediment ages from depths, provided data are presented by depth, and either a `14C` or `CAL`file is available. This menu allows control of the process of constructing an age-depth model. The model is constructed before the writing of any 'plotting' results to the main output file, and the results are plotted on an age-depth PostScript output file (menu E (p. 41)) and summarized in a plain text file (menu E (p. 41)) suitable for incorporation into a spreadsheet. It is thus possible to run your model with the program in interactive mode (menu K (p. 50)), quit when the list of taxa appears for 'plotting', and examine the results in either output file.

Age-depth models may be based on either radiocarbon age determinations (the default) or calibrated ages. Use with radiocarbon ages directly is simpler, and their input is described under Format of associated input files (p. 25).

8.15.1 Calibrated ages

Use with calibrated ages is more complex, because of the nature of these ages. There is no one-to-one conversion of radiocarbon ages onto a calibrated calendar year timescale. The calibration curve is irregular, and the result of mapping a radiocarbon age, with its symmetrical normal probability distribution, onto the irregular calibration curve is an irregular probability distribution, which cannot be treated statistically in the same way as radiocarbon ages. **psimpoll** appraoches the problem through the output generated by the BCal[14] calibration system. Other calibration programs are available (see Radiocarbon Web-info[15]), but BCal offers two advantages:

[14] See URL http://bcal.shef.ac.uk
[15] See URL http://www.c14dating.com/

- It uses a statistical approach that makes it possible to take account of other dates during the calibration of each individual date. For example, if it is known that one date is younger then the other, because of their stratigraphic relationship, it is possible to use this restriction to reduce the range of probabilities for calibrating either date. Individual dates can also be calibrated, with essentially the same result as from any other calibration program.
- It provide output in a convenient tabulated form, for further use.

The probability distribution of a calibrated date is irregular. **psimpoll** handles this by using either the mode or the weighted average mean as the value for the calibrated age. The mode has no standard deviation associated with it, but the weighted average does. If confidence intervals are requested (menu N (p. 75)), these are obtained by drawing random numbers from the irregular distribution, if the mode is being used, or from the normal distribution defined by the weighted mean and its standard deviation.

The use of calibrated ages is determined here, in menu L, but some of the details are controlled in menu N (p. 75), and discussed there.

When menu L is selected, the following submenu appears:

```
L Conversion of depths to ages
a Convert depths to ages (off)                      [off]
b Specify an age for the top sample                 [0+/-50]
c Specify an age for the basal sample               []
d Select an age model                               [1]
e Select number of terms for polynomial age model [0]
f Change laboratory error multiplier (1.00)         [1.00]
g Include timescale with depth axis (off)           [off]
h Convert all ages to AD-BC scale (off)             [off]
i Use calibrated ages from BCal (off)               [off]
j Mode or weighted average with calibrated dates  [MODE]
9 Leave this menu
```

Lb-Lf, Lh and Li are relevant only if La is 'age', Lg is 'on', or the second letter of the code in input Item 4 of input is 'b' (conversion of concentration data to accumulation rate data). Lj is relevant only if Li is 'on'. If sample values are not converted to ages, any output file for age-depth data (menu Eb (p. 41)) is not used.

8.15.2 Menu La: convert depths to ages

Enables conversion of sample values given as depths to ages, using the age model characteristics set up in the rest of the menu. The program then works as if the original data were age data in all respects. If values are to be plotted

as depths, the age modelling is still relevant for the calculation of accumulation rates and their confidence intervals. Headings for age or depth scales are chosen automatically here, but may be over-ridden in menu I (p. 43). This option and the addition of a timescale to diagrams plotted by depth (menu Lg (p. 58)) are mutually incompatible: setting one of them on will cause the other to be turned off.

This option toggles between 'off' and 'on'.

8.15.3 Menu Lb: top age

The user is prompted for values for an estimated age and then estimated error for the top sample of the sequence. A value for the age of the top sample may be given, but the default (0±50) will cover most purposes, and a value of < -2000 (I suggest using -9999) means that there is no age estimate available (or that a radiocarbon age falls exactly at the top sample). This usage is different from that for the basal age (Lc (p. 55)) for historical reasons: it may change. The allowed range is -10000 to 100,000, with an error between 0 and 2000. If no value is given for the top sample, ages calculated for samples above the uppermost age are obtained by extrapolation.

A value can still be entered here, and will be used, even if calibrated dates are being used (menu Lj (p. 59)). It will always be treated statistically as a normal distribution.

8.15.4 Menu Lc: basal age

The user is prompted for values for an estimated age and then estimated error for the basal sample of the sequence. No age is assumed for the basal sample, unless one is given here: the allowed range is 0 to 10,000,000, with an error between 0 and 20,000. If no value is given for the basal sample, ages calculated for samples below the lowermost age are obtained by extrapolation.

A value can still be entered here, and will be used, even if calibrated dates are being used (menu Lj (p. 59)). It will always be treated statistically as a normal distribution.

8.15.5 Menu Ld: age model

Offers the following choices for the age model to be used:

```
1 Linear interpolation between dates
2 Cubic spline interpolation
3 General linear line-fitting by weighted least-squares
4 General linear line-fitting by singular value decomposition
5 Curve-fitting by Bernshtein polynomial
6 Loess smoothing

Enter age model
```

Ld.1 Linear interpolation is a good standby, and surprisingly hard to beat, but confidence intervals are poor.

Ld.2 Cubic spline is a form of curve fitting, based on 4-term polynomials, that passes through all the given points. It can be interesting, but is highly susceptible to wild swings even short distances outside the range of given dates.

Ld.3 The age-depth model is a polynomial curve, fitted using 'normal' equations, of the type
$y = a + bx + cx^2 + dx^3$ etc.,
where y is estimated age, and x is sediment depth. Having selected one of these options, you may need to indicate how many terms you want in the polynomial (menu Le (p. 57)). A two-term model (i.e., $y = a + bx$) is identical to linear regression. The number should be in the range from 2 up to the number of age estimates available, inclusive, or else 0 to obtain the results of χ^2 from attempts to fit curves using first 2, then 3, up to the maximum number of terms in turn. You then need to choose which number of terms to use. In general, the idea is to use as few terms as possible, consistent with a low value for χ^2. To help evaluate χ^2, a measure of the goodness-of-fit is also given. This number is the probability that results as bad as yours would have been obtained if the selected model and number of terms is the right one. This value should, ideally, exceed 0.05, but often the best available value will be much less than this. You should notice that χ^2 values become lower with more terms used: a better fit with more terms. Occasionally, the goodness-of-fit will increase slightly as the number of terms increases. This is because the number of degrees of freedom decreases with increasing numbers of terms.

Ld.4 As for Ld.3 (p. 56), except that curves are fitted using singular value decomposition (SVD): see Press *et al.* (1992) for details. SVD should always give results at least as good as the normal equations: with some datasets it may do better. Try both.

Ld.5 The age-depth model is constructed using a Bernshtein polynomial, also known as a Bézier curve. This is a curve that passes smoothly through or near all the points, and is fitted by successive approximation. It is constrained so that it must pass through both endpoints (the highest

and lowest ages). It will always fall within a polygon that bounds all the points. The curve uses all points for estimating an age for any given depth, so changing any point influences the entire curve. I find that it tends to fit close to most points, but will effectively bypass an outlier. The curve also tends to be rather stiff, but this has the advantage that it does not behave wildly in extrapolated portions (cf. the spline in menu Ld.2 (p. 56), for example). Newman and Sproull (1981) discuss the properties of these curves in more detail.

Gradients on this curve are obtained by numerical differentiation, using a central difference formula.

Ld.6 The age-depth model is constructed by local regression. This is an approach that fits curves to noisy data by a multivariate smoothing procedure: fitting a linear or quadratic function of the variables in a moving fashion that is analogous to how a moving average is computed for a time series. The method is described by Cleveland (1979) and Cleveland and Devlin (1988).

Do not use age models blindly. Some fitted curves can do strange things (especially the spline), others will simply not provide reasonable fits (usually through inadequate estimation of errors on ages, or peculiarly variable sediment accumulation patterns). Before using these models to interpret pollen data, check the PostScript output file. Is it reasonable? Are there portions where the curve provides a negative accumulation rate? What happens in extrapolated portions of the curve? Use the confidence interval option with concentration data to see the sample standard deviations of the calculated sediment accumulation rates. Compare curve output with linear interpolation output.

Graphical and data output is provided even where accumulation rates are negative. This may look strange, especially on the main pollen diagram.

8.15.6 Menu Le: number of terms for polynomial age models

This enables choice of the number of terms for the age models of Ld.3 (p. 56) and Ld.4 (p. 56).

8.15.7 Menu Lf: laboratory error-multiplier

Use to increase the errors on the age estimates by a constant factor ('laboratory multiplier'). Experiment with this to see the consequences for the goodness of fit through age models Ld.3 (p. 56), Ld.4 (p. 56), or Ld.6 (p. 57). A good start would be to use a factor of 2.0.

[Allowed range 0-5]

8.15.8 Menu Lg: timescale with depth scale

Enables the addition of a timescale next to a depth scale. The time-scale is obtained via the age models set elsewhere in menu L. Timescales are obtained by an iterative process for all age models except linear interpolation, and may be slow (especially with the Bernshtein polynomial). During calculation, numbers are printed to the screen to indicate progress, counting up towards the number of tick marks on the scale. This option and the plotting of diagrams by age (menu La (p. 54)) are mutually incompatible: setting one of them on will cause the other to be turned off. However, a separate age-depth plot will be produced, displaying the model used to determine the timescale.

This option toggles between 'off' and 'on'.

8.15.9 Menu Lh: AD/BC scale

Enables plotting of age scale in AD/BC units without changing the `C14` file. Age input for all variables is assumed to be in years before AD 1950, and 1950 years are subtracted from the input. The scale heading is adjusted, but can be over-ridden from menu I (p. 43).

This option toggles between 'off' and 'on'.

If sediment accumulation rates are found to be negative, the program stops after writing the output file for age-depth data (if any was requested). This is to enable the user to examine the output, and find out exactly how the line-fitting went wrong.

You may find that no model produces, by itself, a reasonable approximation of the age-depth relationship in your sequence. In this case, produce your list of ages, perhaps by combining output from two or more models over different sections of the core, and incorporate it into your main input file, presenting the data by age rather than depth. However, if you do this, the program will not be able to calculate accumulation rates: you would also have to do that yourself, and no error estimates would be possible.

8.15.10 Menu Li: Calibration

Enables reading of BCal calibrated data, instead of the default radiocarbon data. The direct effect is that the program searches for a `CAL` file instead of a `C14` file, and preceeds from there.

This option toggles between 'off' and 'on'.

Details of the input format of the `CAL` file are described under Format of associated input files (p. 26). The name of the `CAL` file can be changed in menu E (p. 41).

8.15.11 Menu Lj: Mode or weighted average with calibrated dates

Calibrated dates have an irregular probability distribution, and are thus difficult (impossible) to represent with a single number, unlike radiocarbon dates. Two possibilities are available: to use the mode (year with the highest probability), or a weighted average of the distribution as a single value for use on plots. Menu Lj, here, allows the choice of which to use.

Toggles from 'MODE' to 'WTAVE'.

Results of modelling are always presented graphically as a 95% confidence interval band either side of their mean (which will be identical, or nearly so, to the weighted average). If menu Lj is set to 'MODE', the dates will be apparently off-center from this distribution,.

8.16 Menu M: data analyses

```
M Data analyses
a Rate-of-change analysis (off)                  [off]
b Rarefaction analysis (off)                     [off]
c Zonation (off)                                 [off]
d Principal components analysis (off)            [off]
e Correspondence analysis (off)                  [off]
f Fourier analysis (off)                         [off]
g Statistical description of data (off)          [off]
h Independent splitting of taxa (off)            [off]
i Minimum value for inclusion of a taxon         [0.05]
j Recalculation of dataset proportions (on)      [on]
9 Leave this menu
```

Enables various numerical analyses on the input data, ranging from summary statistics to complex multidimensional analyses. Except for independent splitting, analyses are carried out before the writing of any 'plotting' results to the main output file. It is thus possible to run these analyses with the program in interactive mode (menu K (p. 50)), and then quit when the list of taxa appears for 'plotting'. Where appropriate, results will be in the data analysis file selected in menu Ec (p. 41).

For analyses other than rarefaction analysis and independent splitting, the program searches through the dataset for taxa that have a maximum value that exceeds a certain threshold (menu Mg (p. 74)), and (optionally) then asks the user whether to include that taxon or not. Answering 'A' at any point in the sequence of taxa causes all remaining taxa with values that exceed the threshold to be included. Answering 'M' includes all taxa until a sum value is encountered (identified by a * as the first character): the sum value itself and all following taxa are excluded. If the threshold value is zero, the user is asked about each taxon in the dataset. I cannot think of a less cumbersome way to

eliminate taxa from consideration. The resulting reduced dataset may then be recalculated as proportions of the taxa included (the default) or left in the same form as input (menu Mi (p. 75)).

8.16.1 Menu Ma: rate of change

Rate-of-change analysis produces this display:

```
0 Rate-of-change analysis (off)                    [off]

Enter '0' for on, or '1' for off
```

Rate-of-change analysis was introduced by Jacobson *et al.* (1987) and Jacobson and Grimm (1988), and further explored by Bennett & Humphry (1995). It involves measuring the dissimilarity between adjacent pairs of samples, and then relating that to the temporal difference between the samples. The methods of Jacobson *et al.* (1987) and Jacobson and Grimm (1988) are not well-explained, but it appears that they first smoothed their pollen data, then interpolated to give pollen spectra at even intervals of time along the sequence, then measured dissimilarities. The method used here is to measure the dissimilarity between adjacent samples, and divide this measure by the temporal difference between the samples (Bennett *et al.* 1992). This is simpler, and involves much less tinkering with the data. However, dissimilarity measures are non-linear (in a sequence of three samples, the dissimilarity between samples 1 and 3 will not be the same as the sum of the dissimilarities between 1 and 2 and between 2 and 3). This should not be a serious problem where the time difference between samples is approximately constant, or changing smoothly and slowly.

The dataset can be smoothed before doing the analysis. The number of points smoothed can be 3, 5, 7, or 9. Smoothing is carried out by a least-squares line-fit, with 2-terms (straight line) for 3 points, or 4-terms (cubic polynomial) for the others.

Results are available for immediate plotting, and will be included with the rest of the dataset for saving to the data analysis file (menu J (p. 50)). The 'taxon name' given is name of the dissimilarity measure used (see below). Depths must be either presented as ages (see Data preparation (p. 16)) or converted to ages (menu La (p. 54)) in order to obtain rates of change. This means that on the **psimpoll** run that does the calculations, the output will have to be plotted against age.

If 0 is selected, the following display appears:

```
1 Dissimilarity measure (1)                    [1]
2 Smoothing parameter (0)                      [0]
3 Modelling of rate of change (off)            [off]
9 Leave this menu
Q Return to main menu
```

Ma.1 enables choice of dissimilarity measures:

```
1 Chord distance
2 Information statistic
3 Chi-squared coefficient 1
4 Chi-squared coefficient 2
5 Manhattan metric
6 Euclidian distance
7 Angular separation
8 Standardized Euclidian distance
Enter number <Enter>
```

There are many dissimilarity measures (Prentice 1980), but currently I have implemented just these eight after experiments by Roger Humphry (Bennett & Humphry 1995). Choice 1 is the classic measure, and the one used by Jacobson *et al.* (1987), Jacobson & Grimm (1988), and Bennett *et al.* (1992). Try them all.

Ma.2 allows choice of the number of points smoothed (see above).

Ma.3 allows modelling of rates of change in order to assess confidence in the results. A submenu (see menu Mc.7 (p. 66)) allows change of various parameters connected with this. Results, printed to the data analysis file, give the standard deviation for each calculated rate of change based only on the uncertainty of the original pollen counts. Errors associated with the age-depth modelling are incorporated if available (request confidence intervals in menu N (p. 75)), but there is not yet any plotted output.

8.16.2 Menu Mb: rarefaction

Rarefaction analysis produces this display:

```
1 Rarefaction analysis (off)                   [off]

Enter '0' for on, or '1' for off
```

It is a matter of common observation that different pollen spectra have different numbers of pollen types, but it is also obvious that the number of taxa must have some relation to the number of pollen grains counted, and this number is never constant from one sample to another. Rarefaction analysis (introduced to pollen analysis by Birks & Line [1992]) provides a way of determining the palynological richness of a given pollen spectrum if a certain other number of grains had been counted (which must be less than the actual number counted). The pollen count is thus effectively standardized to a single sum, and it is then possible to compare the richness between samples. If the size of standardized count is kept constant between sequences, and other conditions hold (similarity of pollen taxonomy between analysts, for example), then it may be possible to compare richness between sequences.

In order to use this option, the main input file must:

- include all pollen types counted within the main sum;
- include values for the pollen sums (distinguished by a '*' in the usual manner), placed immediately after the last pollen type from within that sum;
- not include any types outside the sum earlier in the main input file than the sum values.

If 0 is selected, you will be prompted:

```
Specify the count size for standardization       [0]
Enter number <Enter>
```

The default value is the lowest pollen sum in the dataset. Otherwise, 250 is a good number, and should be achievable for all datasets.

Results are available for immediate plotting, and will be included with the rest of the dataset for saving to the data analysis file (menu J (p. 50)). If samples are marked for skipping, the results for those samples are set as -1. When plotted, the curve will be joined up across them, or a break left, depending on the setting of menu Df (p. 40).

8.16.3 Menu Mc: zonation

Zonation produces this display:

```
2 Zonation (off)                                 [off]

Enter '0' for on, or '1' for off
```

This section carries out zonation of pollen stratigraphic data, by splitting using binary or optimal techniques (each by minimizing sum-of-squares or information content criteria for making the splits), or by agglomeration using constrained cluster analysis, based on squared Euclidian dissimilarity (CONISS), developed by Grimm (1987), or on information content dissimilarity (CONIIC).

The binary approach splits the data set into successively small groups by splitting existing zones. Results for any given number of zones are thus an extension of results for all results with fewer zones. The optimal approach starts afresh for each successive number of splits, and thus there will not necessarily be any correspondence between results for division into diferent numbers of zones. The principles are discussed fully in Birks & Gordon (1985). Optimal splitting can be very time-consuming. However, in principle and in practice it is more satisfactory than binary splitting, and I now use it routinely (Bennett *et al.* 1992).

CONISS and CONIIC are based on cluster analysis, with the constraint that clusters must be based on agglomeration of stratigraphically adjacent samples. See Grimm (1987) for details of CONISS. CONIIC differs in using the information content statistic (see Prentice (1980) instead of sqaured Euclidian distance to form the initial dissimilarity matrix.

Results, as defined zones, are written to the `ZONE` file, unless the data have been shuffled (see menu Mc.6 (p. 65)), and will be used and plotted on the current diagram. If the data are not being shuffled and a `ZONE` file is not available, zonation will not be carried out. A data analysis file may be set up (menu Ec (p. 41)) and, if so, will include details of the numerical side of the zonation. Results from CONISS are laid out more or less as in Grimm (1987, Appendix), except that depth labels are used instead of markers for depths. No judgement is made about subzones: that remains a subsequent matter for the analyst (Grimm 1987).

For CONISS and CONIIC, a dendrogram showing the formation of the clusters, is drawn at the end of the plot, after the final column of zones. If there is insufficient room on the final page, it will be plotted on a new page. It can be moved and (re-)combined with other plot components using **pscomb**.

It would be usual to apply a threshold value (menu Mi (p. 74)), and recalculate the dataset for analysis to proportions of the sum of types included (menu Mj (p. 75)).

The output tables mark, with '*', zonations with variances considered to be interpretable by exceeding values generated by a broken-stick model of the distribution of variance amongst the various components, as has been done with the eigenvalues of a PCA (Legendre & Legendre 1983; Jackson 1993). The idea is that each successive 'split' accounts for part of the total variance in the dataset, and this part is tested against the portion expected from a broken-stick model. Strictly, I suspect that this may not be applicable to optimal splitting, since each level of zonation is begun anew, and is not part of a previous level of zonation. The zonations marked may thus be considered to

have some statistical validity. See Bennett (1996) for more details.

If 0 is selected, you will see the following submenu:

```
1 Number of zones wanted                          [10]
2 Method for obtaining zones                      [1]
3 Name for output file with zone details          [a:\dallZONE]
4 Prefix for zone labels                          [D-]
5 Data transformation                             [0]
6 Use a randomized dataset (off)                  [off]
7 Modelling of zonation (off)                     [off]
9 Leave this menu
Q Return to main menu
```

Follow the links shown for more details.

Mc.1 enables you to choose how many zones you want. If 0 is given, the program prints summary results. For the splitting methods, these are in terms of the decreasing variance with successive numbers of zones, and allows a choice of which the user wants. The program continues splitting until 10 splits have been made, and then offers a choice of selecting one of those, or looking at the next 10. For CONISS, summary results are in terms of the increase in 'dispersion' after combining two clusters. This should be large for the division into two zones (actually the last agglomeration), and decrease for more zones (earlier agglomerations). Again, results are presented in batches of 10 'clusterings' with a choice of using one, or continuing with the next batch.

[Allowed range 0-(number of samples - 2)]

Mc.2 enables choice from the five possible methods for making splits:

```
1 Binary splitting by sums-of-squares
2 Binary splitting by information content
3 Optimal splitting by sums-of-squares
4 Optimal splitting by information content
5 Constrained cluster analysis by sum-of-squares (CONISS)
6 Constrained cluster analysis by information content (CONIIC)
Enter method
```

I recommend Mc.2.4 (optimal splitting by information content) as first choice, but try them all. CONISS and CONIIC are faster, but note that they work by agglomeration rather than splitting. Comparison of results between CONISS/CONIIC and the splitting methods may give an indication of the robustness of results.

Mc.3 allows a change of output file for zonation results. This is the same as

the file used to read in zonation details, which is done by the program after any zonation calculation. So, the file name given here will contain results from the current run of the program, overwriting any existing results in that file, and will then be used to label diagrams. The file name should be changed here (or in menu Ed (p. 41)) if you do not want to lose previous results.

Mc.4 determines the prefix to be used for zone labels. The default is 'X-', where 'X' is the first character of the title (Item 1 of the main input file). Zone labels are centred vertically and horizontally in boxes at the end of the plot.

Mc.5 allows data transformations before the zonation (by any method) is carried out. Selecting this option produces the menu:

```
Transformations available:

0 None
1 Square-root
2 Standardization to zero mean and unit variance
3 Normalization of variables to unit length
4 Centred log ratio / logarithm
Enter number <Enter>
```

The first three options are from CONISS (Grimm 1987), which should be consulted for details of their use. For proportional data, Mc.5.4 uses the transformation recommended by Aitchison (1986) for compositional data, but otherwise uses the transformation $\log(x + 1)$. If the program crashes, you may be using a form of transformation that is incompatible in some way with the method used for zonation.

Mc.6 produces a shuffle of the samples in the dataset being analysed. The shuffling of samples is effective only for zonation, and not for other analyses carried out in the same run of **psimpoll**, nor for the main dataset. The purpose of doing it is to establish what the background change in variance is between consecutive splits. You should find that the variance change is more or less constant, and similar to the change between splits in the unshuffled dataset when the number of splits is large. It can thus be used to help determine how many splits are worth considering, and where the splitting is working on variation that is really noise. Results are written to the data analysis file, but not to the `ZONE` file.

Shuffling may be used for all types of zonation in menu Mc.2 (p. 64), but I am not convinced of its usefulness for CONISS.

The shuffle is controlled by random numbers picked using a minimal standard random number generator (menu N3.1 (p. 77)), and uses the same initial seed as would be used in obtaining random numbers for the purpose of estimating confidence intervals (menu N4 (p. 78)).

Mc.7 enables modelling of zonation, controlled by the following menu:

```
Options to control modelling:
0 Switch on/off (off)                          [off]
1 Number of models to generate (10)            [10]
2 Random number generator to use (1)           [1]
3 Seed for random number generator             [TIMER]
9 Leave this menu
Q Return to main menu
Enter number <Enter>
```

The results of zonation depend to some extent on the reliability of the input data. However, if it is possible to obtain any kind of confidence interval for the results, I am unaware of it, although the broken stick model (discussed above) helps. One solution is to 'model' the data, using the uncertainties associated with the input data (which can be expressed as confidence intervals) to generate alternative datasets, and carrying out zonation on those. I have implemented this in a simple way for proportional data. Such data are distributed binomially, so I draw random numbers from a binomial distribution, which requires knowledge of the observed proportion for a given type in a given sample, and the sum on which that calculation was based. In order to carry out the modelling, therefore, a sum must be available (marked by '*' at the beginning of the 'taxon' name: see Data preparation (p. 16)), which must follow in the dataset all types included within the sum.

Types selected for inclusion in the sum are recalculated, and the sum itself reduced in proportion to the total abundance of the types excluded. Then the zonation is run a number of times which should be as many as possible (e.g., 100): the default is 10 (menu Mc.7.1). Results are then accumulated and printed to the data analysis file (which must be available). A particular number of zones must be designated via menu Mc.1 (i.e., the number of zones cannot be set to 0). Output varies depending on whether zonation is binary or optimal splitting: CONISS output resembles that from binary splitting. For binary splitting, the output shows which zones boundaries are indicated and at what proportion of the number of models, in order of occurrence. For optimal splitting, the output shows which complete patterns of splits occurred, and the proportion of the models that these were found with. In all cases, the output concludes with results of zonation from the actual data for comparison.

Zonation is expensive in computer time anyway: running many models may take an inordinately long time. Run this option overnight, or use a mainframe.

Menu Mc.7.2 allows choice of random number generator, and menu 2.7.3 allows choice of the seed used. These submenus are identical to those described under menu N3, and the same variables are used to store the choice of random number generator and seed.

8.16.4 Menu Md: principal components

Principal components analysis produces this display:

```
3 Principal components analysis (off)            [off]

Enter '0' for on, or '1' for off
```

This is the classic way to reduce multidimensional data to a low number of dimensions (Birks & Gordon 1985). First, the method measures the similarity between taxa (correlations or covariances), then the resulting matrix is subjected to the PCA procedure, looking for major directions of variation within the data set.

Results are printed out to the data analysis file (see menu Ec (p. 41)), if one is available. The output includes the basic matrix, eigenvalues, taxa loadings, and sample scores. These can, of course, be extracted and plotted in some appropriate manner. It would be usual to apply a threshold value (menu Mi (p. 74)), and recalculate the dataset for analysis to proportions of the sum of types included (menu Mj (p. 75)).

The table of eigenvalues marks, with '*', those eigenvalues considered to be interpretable by exceeding eigenvalues generated by a broken-stick model of the distribution of variance amongst the various components (Legendre & Legendre 1983; Jackson 1993).

If 0 is selected, the following submenu appears:

```
1 Specify the matrix to use                      [3]
2 Specify the number of axes for output          [6]
3 Data transformation                            [0]
4 Modelling of pca (off)                         [off]
9 Leave this menu
Q Return to main menu
```

While PCA is running, just so that you know something is happening, some numbers are printed to the screen. During calculation of the basic matrix, these are the numbers of the current taxon being analysed. During the PCA itself, the number is the number of 'sweeps' through the input matrix to achieve convergence. This number ought to increment to between 6 and 10. The output file includes a figure for the number of 'jacobi rotations': this should be between $3n^2$ and $5n^2$, where n is the number of taxa in the input matrix.

Md.1 produces this submenu:

```
Specify the matrix to use                              [3]
1 Linear correlation
2 Rank correlation
3 Covariance
Enter number <Enter>
```

Allows choice of basic matrix. Md.1.3 (covariance matrix) is the usual favourite, but the others work as well. Rank correlation is experimental: see any statistical text for how this differs from linear correlation.

Md.2 produces this submenu:

```
Specify the number of axes for output                  [6]
Enter number <Enter>
```

There are as many axes as there are taxa in the dataset, but usually only the first few are useful. Use this option to choose how many you would like to see.

Md.3 produces the submenu for data transformations: this is identical to that described for zonation (menu Mc.5 (p. 65)). Square-root transformation is routinely used with covariance matrices for the standard pollen data PCA. For proportional data, centred log ratio) is probably the most desirable statistically, using a method recommended by Aitchison (1986) specifically for compostional data such as pollen percentages. It works by transforming the data values to `log(value / geometric mean of all values in the sample)`, and then calculating covariances in the usual way to derive the basic matrix. I know of no use of this in the palynological literature (yet!). Because this method takes log values, zero data entries cannot be used. For proportional data, I treat zeroes as 'trace zeroes': values that are not absolute zero, but only seen as zero at the detection level of the count. I use a value of 0.001 as the rounding error ($0.5 \times \frac{1}{500}$), assuming that 500 is a typical pollen count. For other data types, the centred log ratio method calculates $\log(x+1)$. I have not yet tried PCAs with the other two transformation methods (taken from CONISS [Grimm 1987]: see under zonation (p. 62)).

Md.4 produces the same submenu as menu Mc.7 (p. 66), controlling options for modelling principal components analysis. The same general considerations apply with respect to running a number of models with data drawn from binomial distributions corresponding to the observed data.

Output is currently confined to including a standard deviation to each eigenvalue, obtained from the eigenvalues from modelling runs. As with zonation, modelling principal components analysis may be extremely expensive in computer time

8.16.5 Menu Me: Correspondence analysis

Correspondence analysis produces this display:

```
e Correspondence analysis (off)            [off]

Enter '0' for on, or '1' for off
```

This type of multivariate analysis was developed to cope with enumerated data. It resembles principal components analysis, but is based on a different type of similarity matrix. Like other forms of ordination, results from correspondence analysis often plot along a marked curve, or horseshoe, when seen in two dimensions. Detrended Correspondence Analysis is correspondence analysis with a subsequent procedure to remove this tendency (i.e. detrend it). This part of the program is based on the FORTRAN program DECORANA, written by Hill 1979), with corrections from Oksanen & Minchin 1997).

Results are printed out to the data analysis file (see menu Ec (p. 41)), if one is available. The output includes eigenvalues, taxa loadings, and sample scores. These can, of course, be extracted and plotted in some appropriate manner.

It is often desirable to apply a threshold value (menu Mi (p. 74)), and recalculate the dataset for analysis to proportions of the sum of types included (menu Mj (p. 75)).

If 0 is selected, the following submenu appears:

```
a Detrended correspondence analysis (off)      [off]
b Downweighting of rare types (off)            [off]
c Rescaling of axes                            [off]
d Number of times to rescale axes              [4]
e Rescaling threshold                          [0]
f Number of segments                           [26]
g Data transformation                          [0]
h Modelling of ca (off)                        [off]
9 Leave this menu
Q Return to main menu
```

Menu Mea: Detrended correspondence analysis The default form of analysis is correspondence analysis (also known as reciprocal averaging). This option turns on detrending to remove the 'horeshoe' effect and allow rescaling of the axes.

This option toggles 'on' and 'off'.

Menu Meb: Downweighting of rare types This option enables a transformation that gives rare types a lower value, reducing their significance in the analysis. All values for a taxon are reduced if the overall level of the frequency if the taxon is less than 0.2 times the overall frequency of the most abundant taxon. The measure of frequency used here is the square of the sum of all the values for the taxon divided by the sum of the squares of the values.

This option toggles 'on' and 'off'.

Menu Mec: Rescaling of axes This deals with the compression effect of points at the ends of the axes, and rescales to remove the compression. The number of times to rescale is set in Med (p. 70)

This option toggles 'on' and 'off'.

Menu Med: Number of times to rescale axes This deals with the compression effect of points at the ends of the axes. The effect of re-scaling may be incomplete at the first attempt, but 4 attempts have been found (by M. Hill) to work well. It is not advised to change this parameter (without reading Hill's DECORANA manual).

Rescaling must be turned 'on' for this option to take effect (see Mec (p. 70)).

[Allowed range 0-20]

Menu Mee: Rescaling threshold The rescaling threshold is the length of an axis, in standard deviation units, above which it will be rescaled: there is probably no need to change this.

[Allowed range 0.0-100.0]

Menu Mef: Number of segments CA axes are detrended by dividing them into segments, and then equalizing the scores in each segment. The default number of segments to use is 26. Users are not advised to change this parameter. However, if detrending is not fully effective (visible as a systematic relation between the first and second axes), try raising the parameter to the maximum (46).

[Allowed range 10-46]

Meg produces the submenu for data transformations: this is identical to that described for zonation (menu Mc.5 (p. 65)). Square-root transformation

effectively upweights minor taxa relative to more abundant taxa.

Within correspondence analysis itself, an additional transformation is available, the downweighting of rare types. See Meb (p. 69)

Standardization and centred log ration produce negative values, which cannot be used in correspondence analysis. These transformations should not be selected (the program will exit with an error message such as `sqrte: Assertion 'x>=0' failed`).

Meh produces the same submenu as menu Mc.7 (p. 66), controlling options for modelling correspondence analysis. The same general considerations apply with respect to running a number of models with data drawn from binomial distributions corresponding to the observed data.

Output is currently confined to including a standard deviation to each eigenvalue, obtained from the eigenvalues from modelling runs. As with zonation, modelling correspondence analysis may be extremely expensive in computer time

8.16.6 Menu Mf: Fourier

Fourier analysis produces this display:

```
4 Fourier analysis (off)                              [off]

Enter '0' for on, or '1' for off
```

Fourier analysis converts the data from the time domain to the frequency domain. This form of time series analysis is normally only available for samples that are evenly spaced in time, which is far from true for most pollen data. However, a technique devised by Lomb to generate a periodogram works with unevenly sampled data. I have implemented this, on an experimental basis, from Press *et al.* (1992). Results are written to the data analysis file, giving a string of values for the importance of each taxon at a range of frequencies. The numbers can be plotted by 'cutting' the results out of that file and creating a new file that can be presented to **psimpoll**. The number of taxa depends on how many you selected when the program ran. The number of 'samples' (now 'frequencies') is 5 × the number of samples in the original dataset. The 2-character code for input Item 4 is '`lf`', resulting in appropriate labelling (see Data preparation (p. 16)).

This option may run slowly on older computers, perhaps 10 minutes per taxon for a dataset with 80 samples on a 386 processor. The length of time increases as n^2 where n is the number of samples.

If 0 is selected, you will see the following submenu:

```
1 Data transformation                              [0]
9 Leave this menu
Q Return to main menu
```

Me.1 allows a data transformation before Fourier analysis is carried out. The submenu is the same as that described for zonation (menu Mc.5 (p. 65)), and discussed for principal components analysis (menu Md.3 (p. 68)).

8.16.7 Menu Mg: statistical description

Six basic analyses are available:

- calculation of mean, standard deviation, skewness, and kurtosis for the values presented for each pollen type, with results written to the data analysis file (menu Ec (p. 41)). Values of skewness and kurtosis significantly different from zero (the value for a perfect normal distribution) are marked;
- calculation of a weighted average, or 'optimum' location of each pollen type in time (or depth), together with a standard deviation, or 'tolerance'. This is a measure that combines the extent of occurrence of a taxon over time, weighted by its abundance. It is included here, experimentally, as a potentially better statistic for the location of a taxon in a sequence than such subjective measures as 'rational' or 'empirical' limits etc. The idea is borrowed from the statistics used to define optima of diatoms along pH gradients: I have substituted time (depth) for the environmental gradient (Birks *et al.* 1990). Output in the data analysis file gives, for each taxon, the optimum, the tolerance, and the location (depth or age) of the taxon's maximum value, and the value of that maximum;
- calculation of a 'runs' test. If abundances of a pollen type along a sequence were randomly distributed with time, then the length of 'runs' (series of samples all successively greater (or lower) than the previous sample) will follow a predictable pattern. We can count the lengths of observed runs (how many runs of a single sample, runs of two samples, etc), and compare the resulting distribution with the distribution for random samples. The result is a runs test, and gives a measure of whether the overall pattern of change is, or is not, significantly different from random. The output gives the runs statistic, udr, and the goodness-of-fit, q. q greater than 0.05 indicates that the sequence of runs for the taxon is not significantly different from random.
- calculation of a Linear Correlation Coefficient. This is a statistic that can be used for detecting trends in the values for taxa. If a taxon's values increase continuously along the sequence, its correlation coefficient

(r) will be high and significant. If there are no trends in the data, r will be low (near zero). The sign of r indicates the direction of the trend. In the normal situation where depth (time) values increase numerically down a sequence, r will be positive for taxa whose values are highest towards the base (i.e., decrease upwards). Conversely, r will be negative for taxa whose values increase up the sequence. Results (in the data analysis file) include for each taxon r, the significance level for the difference between r and zero, and Fisher's z, which can be used further to investigate differences between values of r. For 95% confidence values, look for values of 'signif' that are less than 0.05. See Press *et al.* (1992) for further details.

- calculation of Spearman's Rank Correlation Coefficient for each taxon. This is a non-parametric statistic, making no assumptions about the underlying distribution of the data. It is a means of detecting 'trend' in the dataset. The coefficient is based on the rank order of the values for the taxon concerned and assessing the degree to which the rank order is the same as the ordering of the samples (by depth or time). If the taxon values increase continuously along the sequence, the correlation coefficient will be high and significant. If there are no trends in the data, the coefficient will be low (near zero). The sign of the correlation coefficient indicates the direction of the trend. In the normal situation where depth (time) values increase numerically down a sequence, the coefficient will be positive for taxa whose values are highest towards the base (i.e., decrease upwards). Conversely, the coefficient will be negative for taxa whose values increase up the sequence. Results (in the data analysis file) include for each taxon the Correlation Coefficient 'r(s)', its significance 'signif.', the sum-squared difference of ranks 'D', the number of standard deviations by which D differs from zero 'num of sd', and the significance level for the difference from zero. For 95% confidence values, look for values of 'signif' that are less than 0.05. See Press *et al.* (1992) for further details, and Gauthier (2001) for some simple examples of use in environmental research.

- calculation of Kendall's Tau. This is another statistic that can be used for detecting trends in the values for taxa. It is even more non-parametric than Spearman's, being based only on the relative ordering of values: higher, lower, or the same. If the taxon values increase continuously along the sequence, Tau will be high and significant. If there are no trends in the data, Tau will be low (near zero). The sign of Tau indicates the direction of the trend. In the normal situation where depth (time) values increase numerically down a sequence, Tau will be positive for taxa whose values are highest towards the base (i.e., decrease upwards). Conversely, Tau will be negative for taxa whose values increase up the sequence. Results (in the data analysis file) include for each taxon Kendall's Tau, the number of standard deviations by which Tau differs from zero 'num of sd', and the significance level for the difference from zero. For 95% confidence values, look for values of 'signif' that are less than 0.05. See Press *et al.* (1992) for further details.

Output from all three analyses is sent to the data analysis file (selected from menu Ec (p. 41));

This option toggles 'on' and 'off'.

8.16.8 Menu Mh: independent splitting

Walker & Wilson (1978) argued that individual curves for pollen types should be considered independently of each other, especially where the data were pollen accumulation rates, and therefore statistically independent. They presented an approach, with FORTRAN programs by Y. Pittelkow, for making independent splits of the records for individual taxa in a sequence. The approach involves first identifying portions of the record that are effectively non-zero sequences, distinguishing them from other portions of effectively zero sequences. This may be termed presence-absence splitting. The approach then looks at the non-zero sequences, and splits them quantitatively with a maximum likelihood method (Walker & Wilson 1978). Pittelkow's programs implemented the approach in two stages. In **psimpoll** I have combined these, making presence-absence splits first, and then looking for quantitative splits within non-zero sequences. The detection of either sort of split is a statistical matter, depending on a level of significance that is a function of the number of samples (Walker & Wilson 1978, Table 2) and there may be no splits of either sort for any given taxon. Splits are ignored if they would create 'zones' of less than 4 samples. Walker & Pitelkow (1981) present results using the method from three late Quaternary sequences. Walker & Wilson (1978) present curves fitted to portions of the record of particular taxa, identified from presence-absence and quantitative splitting. However, this was not included in Pittelkow's programs, and I have not implemented it in **psimpoll**. Instead, I give means and standard deviations for each identified segment of the record for each taxon.

Results are printed to the screen (possibly too fast to read when plotting interactively), and to the data analysis file (which must have been specified: menu Ec (p. 41)). Lines marking the location of splits are marked across the plot of each taxon. These are normal lines for quantitative splits, but dashed lines for presence-absence splits. If pollen zones are also being incorporated, these are indicated only by boxes at the righthand end of the plot: marker lines for the zones are not drawn across the plot.

Independent splitting is carried out as the program is 'plotting' each taxon to the main output file. It is thus necessary to go through with this stage in order to get any output in the data analyses file. As presently set up, the process will be carried out on all 'taxa' except summary data, pollen sums, and rate-of-change results. I leave it to the user to decide whether it is useful or reasonable for items such as palynological richness!

This option toggles 'on' and 'off'.

8.16.9 Menu Mi: threshold values

Most types in a typical full pollen dataset occur at low frequencies, contributing little that is statistically-useful (Birks & Berglund 1979). It is customary, therefore, to limit datasets to those types that reach a certain minimum value somewhere within the sequence. This option enables you to select what that value should be. For proportional data, 0.05 (= 5%) is set as a default. No default is set for other data types. If the value is negative, the interactive selection of types to include in analyses is disabled. Use a very small negative number to include everything without being asked (e.g., -0.00001). Note that if the recalculation option is set on (menu Mj (p. 75)), the dataset used in analyses will be a proportion of the sum of **everything** that you allow to be included: sums, palynological richness, aquatics, etc. If accumulation rates are being calculated by the program, this is done before producing the reduced dataset for any analysis. Accordingly, the threshold set should be appropriate for the calculated accumulation rates, not the input concentrations values.

8.16.10 Menu Mj: recalculation

This option controls whether the reduced dataset used for analyses is recalculated as proportions of the sum of types included, or left in the same form as the original dataset. It would be normal to recalculate for zonation or principal component analyses, for example, but the approach of Walker & Wilson (1978) for independent splitting of each taxon was intended to work with pollen accumulation rate data. Recalculation would also render the statistical summary (menu Mg (p. 72)) useless;

This option toggles 'on' and 'off'.

8.17 Menu N: confidence intervals

```
N Plot confidence intervals (off)                [off]

Enter '0' for on, or '1' for off
```

This menu controls the calculation of confidence intervals for age estimates and sediment accumulation rates, and also allows the calculation of confidence intervals for data values (which, in certain cases, need the confidence intervals for sediment accumulation rates). Certain extra information must be available to obtain confidence intervals. For % data, pollen sums must be available (marked '*': see above), and must be positioned in the input file after all of the taxa included in those sums. If no sum is found among the pollen taxa after a particular pollen type in the file, no confidence intervals will be calculated for that pollen type. Note that separate sums must be given for taxa in the main

sum, aquatics, *Sphagnum*, indeterminables etc, if separate sums were used to calculate their percentages. The method of calculation follows Maher (1972). For concentrations ('c' in input Item 4 of main input file), additional data are needed (input Items 7-13, as above), and for accumulation rates ('b' in input Item 4), a `C14` file must be present, and additional data for Items 7-14. These Items are described in Data preparation (p. 16).

Confidence intervals for sediment age estimates and sediment accumulation rates are obtained, mostly, by simulation from the available radiocarbon dates and their associated errors (Bennett 1994), or calibrated dates. Values are obtained using means and standard deviations of radiocarbon ages, from 100 simulations by default, or by modelling from the irregular distributions of probabilities of calibrated ages. These may take a few minutes to calculate. The output file (see E2 (p. 41)) includes sample standard deviations of age and accumulation rates. These need to be divided by the square root of the number of simulations to obtain standard deviations of the mean, for comparison with radiocarbon age errors.

Confidence intervals are 'exact' (i.e., obtained by calculation, rather than simulation) for age estimates from linear interpolation, and for the gradient with 2-term polynomial line-fitting.

Using the linear interpolation method (menu L1 (p. 54)), you should find that you get wider confidence intervals where the radiocarbon dates are closest together. The line-fitting methods should not show this effect.

If 0 is selected, the following submenu, given below in full, appears (option 1 only appears for accumulation rate data):

```
1 Fine-scale variation in sediment accumulation  [0.30]
Default = 0.3 per 5 years
2 Method for calculating confidence intervals    [1]
3 Method for generating random numbers           [1]
4 Seed for random number generator               [TIMER]
5 Number of simulations                          [100]
6 Include negative accumulation rates in models  [off]
7 Attempts to model dates without -ve acc. rates [1000000]
9 Leave this menu
```

8.17.1 Menu N1: fine-scale variation

For accumulation rate data ('b' in input Item 4) only, the program prompts for a sample standard deviation of the rate of sediment accumulation over the very short time periods for accumulation of the sediment thickness represented in individual samples. This value is completely unknowable, but some assumptions can be made. The default is 30% of the calculated sediment accumulation rate. You can accept this, enter another value, or disable it by

entering 0.

[Allowed range 0-1]

8.17.2 Menu N2: method

This option affects the method of calculating confidence intervals for concentrations and accumulation rates:

```
1 Modification of Maher's lognormal distributions
2 Propagation of errors
```

N2.1 this method follows Maher (1981), with additional use of propagation of errors methods (Parratt 1961). It produces confidence intervals that are asymmetric (i.e., wider above the calculated value).

N2.2 this method follows a propagation of error method (Parratt 1961; Squires 1985) throughout the calculation. Confidence intervals are almost identical to those obtained by Maher's method, but are symmetric about the calculated point. This means that the lower interval may extend to the negative side of the axis.

8.17.3 Menu N3: random number generation

This option changes the method used to generate random numbers needed to produce confidence intervals by simulation:

```
1 Minimal standard generator
2 Combination of three congruential generators
3 Single congruential generator
4 Subtractive generator
5 Pseudo-DES generator
6 FSU ULTRA generator
7 Quick and dirty generator
8 long-period generator
```

These methods produce uniform random variates between 0 and 1, which are then transformed into the desired distribution. Press *et al.* (1992) discuss the generation of the uniform variates. Here, they are transformed into the normal distribution using the method of Leva (1992). Method 2 is probably the best for general use, 7 is the fastest, 5 is the slowest, but very good, in terms of the properties of the numbers it generates. Method 6 is claimed to be best

generator ever (Zaman & Marsaglia 1991), and method 8 to have an exceptionally long period (Marsaglia & Zaman 1994). (The time taken taken to generate these numbers is small relative to the time doing other calculations.) Although 7 is the fastest, it is the generator provided by the compiler, and should be viewed with deep mistrust (Press *et al.* 1992). It is included because it is fast, and may be good enough for most purposes. The generators work by different methods, so between them they should cover a wide range of likely behaviour. Try different ones to satisfy yourself that results are not an artefact of a particular generator.

8.17.4 Menu N4: random number seed

This option changes the initial seed for random number generation. By default, the seed is taken from the system date and time, and is the number of seconds since the beginning of 1970. The value used for each run is included in the output (menu Eb (p. 41)) so a run can then be repeated exactly by giving here the seed used previously with the same generator and same number of simulations. Enter either 'TIMER' (to use the default) or any positive number not greater than 2147483647.

8.17.5 Menu N5: number of simulations

This option changes the number of simulations. The number of simulations affects the accuracy of estimates of confidence intervals for ages and deposition times. The standard error of the sample standard deviations obtained (and listed in the output file) is $0.5s \times \frac{\sqrt{2}}{\sqrt{n}}$where s is the sample standard deviation, and n is the number of simulations: this works out as $0.071s$ for 100 simulations.

[Allowed range 10-1000]

8.17.6 Menu N6: Include negative accumulation rates in models

Radiocarbon ages and calibrated dates both have a probability distribution, normal in the first case, irregular in the second. Although the dates are known to be in order, because of a stratigraphical relationship, these distributions may (often) overlap. This means that drawing from these distributions can produce modelling sequences of ages that include negative accumulation rates. We can allow this, or we can require that all the modelled series of ages are always positive all the sequence (i.e. each lower modelled date is older than the previous one). Menu N6, here, either allows negative accumulation rates, or prevents them (the default). There can be problems when they are turned off, discussed further in menu N7 (p. 79).

Toggles from 'off' to 'on'.

8.17.7 Menu N7: Attempts to model dates without -ve acc. rates

As discussed under menu N6 (p. 78), there can be a problem with excluding negative accumulation rates from modelled age-depth sequences. Each modelled age must be older than the one below it. If, by chance, an age is taken from the old tail of the one of the age distributions, then next down must be older still, and this can get increasingly more difficult, especially where age probability distributions overlap substantially. The default action is to try 1,000,000 times at each age to obtain a kodelled value older than the previous value. If this fails, that particular modelled sequence is abandoned, and the whole sequence re-done. The number of attempts should be as high as possible, limited by the amount of time available, which is dependent on the speed of the computer system being used. The default works well for me, but I recommend trying other values, especially if your sequence of ages has substantial overlaps. The upper allowed value is the maximum allowed value for a 'long integer', and may be even higher (but not lower) than the value shown below.

[Allowed range 1-2147483647 (at least)]

8.18 Menu O: save configuration file

```
O Save configuration to a:\dallb.CFG
```

Selecting this option saves the current configuration of items in the menu to a `.CFG` file. The `.CFG` file is specific for a particular input file and is saved on the same disc(ette). It is read automatically when that dataset is run again. The `.CFG` is text, and can be altered with an editor. I don't recommend it, but see the example (p. 107) for details. Any previous `.CFG` is overwritten.

8.19 Menu P: read script file

```
P read from script file
```

Selecting this option, in association with menu K (p. 50), reads the details of a previous run of interactive plotting from a `.SCR` file, including any changes of taxa names. The `.SCR` file is specific for a particular input file and is saved on the same disc(ette). If the file cannot be found, then the option is switched off, and a `.SCR` file is written instead. A `.SCR` file is written by default for each interactive run of **psimpoll**, so you must use the usual file manipulation tools

if you wish to protect a .SCR file from overwriting. The .SCR file is text, and can be altered with an editor: see the example (p. 107). After this option has been selected, the plotting part of the program runs without further attention from the user, but the screen updates in the same way.

8.20 Menu U: utilities

```
U Utilities
a Leave program without writing output file
b Display psimpoll version and compilation date/time
c Make coffee
d Show memory usage in psimpoll.LOG                [off]
e Display references for this program
9 Leave this menu
```

This menu allows sundry useful pieces of behaviour.

Ua Leave the program. No output file is created, and no analyses carried out;

Ub Display lines such as:

```
psimpoll version 4.25
compiled by ANSI C compiler 11:07:54 Jul 1 2005
```

If the phrase includes 'non-ANSI', certain aspects of the behaviour of the program may not be as expected;

Uc Self-explanatory;

Ud Show memory usage in psimpoll.LOG. This can be (very) verbose, and thus create large files. It can be useful when problems occur, but is turned off by default.

Toggles from 'off' to 'on'.

Ue Display a list of references used in writing the program.

8.21 Menu X: run program

```
X Run program
```

Entering 'X' runs the program. It can then be stopped by typing the relevant break characters for the system being used (e.g., CTRL-C on DOS) or, more crudely, resetting the system.

9 pscomb menus

A main menu appears immediately after running **pscomb**, with no intermediary prompts. This menu, and associated submenus, are based on those used by **psimpoll**, and wherever possible have the same key letters. Fewer options are available in **pscomb** because most of the work of preparing datasets for output has already been done by **psimpoll**. In all menus, the term 'box' refers to a distinct unit of PostScript code, and may be any of age or depth axes, sediment columns, radiocarbon age columns, zonation columns, or plots of individual taxa. **pscomb** selects 'boxes' from **psimpoll** output, and allows the user to pick and mix them into a new PostScript file.

9.1 Main menu

```
E Change or enable files
F Font family (Helvetica)                    [Helvetica]
G Size of selected font (7.1)                []
I Add/change titles
K Interactive plotting (on)                  [on]
U Utilities
X Run program
```

Select an item from this list by entering the key letter. Details of the result of each choice, together with the permitted values for any modifiable parameters, may be found by following the links.

9.2 Menu E: files

```
E Change or enable files
1 Main PostScript output file                [comb.PS]
2 Input file(s)
9 Leave this menu
```

E1 modifies the main output file. The default is `comb.ps`, which will be in the same directory as the program. This will normally need to be changed. Any file of the same name that already exists will be overwritten.

E2 allows addition of input file names, in addition to any that have been given on the command line, up to a total of no more than 10. Some operating systems (e.g., DOS, UNIX), permit command line arguments, and on these systems input file names may be given by either route (or both). On operating systems that do not support command line arguments (e.g., Macintosh), using this menu is the only way to provide input file names.

All file names given must be legal under the operating system that you are using.

9.3 Menu I: labelling

```
I Add/change titles
1 Title                                          []
2 Label for base of diagram                      []
9 Leave this menu
```

pscomb deals with the main pieces of diagrams ('boxes'), extracted from one or more input files. It cannot, therefore, know or make assumptions about general diagram characteristics such as titles. This menu provides facilities for adding a main title (I1), and adding a general footnote (I2). The text given for either of these items may include special characters to achieve useful effects, just as for **psimpoll** menu I (p. 43) Note especially the use of '#' within a main title to separate it into large and normal font sections (title and subtitle), as is possible in **psimpoll**.

The default for both labels is to print nothing.

9.4 Menu K: interactive plotting

```
K Interactive plotting (on)                      [on]
```

Controls whether the user can select taxa individually for plotting (on), or taxa are plotted in the order found in the input files, running sequentially through each input file in the order given on the command line (off).

Toggles from 'on' to 'off'.

When running under this option, the program provides a screen listing of all boxes in the dataset, each preceded by its number. The first six characters of each box name are given where the dataset includes less than 110 boxes, the first three characters for less than 153 boxes, and the number only for less than 237 boxes. The option is not available for 238 or more boxes (because there is not enough room on the screen). Previously 'plotted' boxes are marked by a hyphen between the box number and the box name, and listed by box number beneath the display of names. Numbers in the extreme left margin indicate a file which is started within the row, or the last file to be begun within the row if more there is more than one.

Box names are derived from the names of the taxa in the original **psimpoll** datasets, except for:

`C14C` Radiocarbon age column;

`Dend` Dendrogram from zonation by CONISS or CONIIC.

`LSSc` Depth or age scale from the left hand side of the plot;

`LTSc` Timescale from the left hand side of the plot;

`RSSc` Depth or age scale from the right hand side of the plot;

`SedC` Troels-Smith sediment description column;

`Zone` Zonation column.

Each input Postscript file will thus have a 'LSSc' and either a 'RSSc' box or a 'Zone' box on every page. There will be one copy per page of any of those boxes on this list that occur at all.

For example:

```
pscomb 1.00

74 boxes
1  0-LSca     1 SedC     2 C14C     3-_Betul   4 _Coryl   5 _Pinus   6 _Querc
   7 _Ulmus   8 _Alnus   9 _Fraxi  10 _Junip  11 _Salix  12 _Empet  13 _Callu
  14 Cypera  15 Poacea  16 _Filip  17 _Poten  18 _Artem  19 Brassi  20 Caryop
  21 _Jasio  22 Zone    23 LSca    24 SedC    25 C14C    26 _Rumex  27 _Polyp
  28 _Pteri  29 Filica  30 _Huper  31 <>      32 _Sphag  33 &Charc  34 Zone
2 35 LSca    36 SedC    37 C14C    38-_Betul  39 _Coryl  40 _Pinus  41 _Querc
  42 _Ulmus  43 _Alnus  44 _Junip  45 _Salix  46 _Empet  47 _Callu  48 Cypera
  49 Poacea  50 _Filip  51 _Poten  52 _Artem  53 Caryop  54 _Jasio  55 _Campa
  56 RSca    57 LSca    58 SedC    59 C14C    60 _Rumex  61 _Ranun  62 Apiace
  63 _Sedum  64 _Saxif  65 _Plant  66 _Polyp  67 _Pteri  68 Pterop  69 !Palyn
  70 <>      71 _Sphag  72 &Charc  73 RSca

Enter 'Q', or 'N': see documentation for details

Taxa plotted so far:                                0.31 pages plotted
 0 3 38
```

You select a box for plotting by entering its number, or a number range. Any other characters, including spaces, are ignored. A number range is given by entering `n1-n2`, where `n1` and `n2` are the first and last numbers of the desired range. Alternatively, instead of a number, enter 'N' to force the start of a new page, or 'Q' to end the plot. As usual, **pscomb** is not case-sensitive about either of these characters.

9.5 Menu U: utilities

```
U Utilities
a Leave program without writing output file
b Display pscomb version and compilation date/time
c Make coffee
d Show memory usage in pscomb.LOG                  [off]
9 Leave this menu
```

This menu allows sundry useful pieces of behaviour.

Ua Leave the program. No output file is created, and no analyses carried out;

Ub Display lines such as:

```
pscomb version 1.03
compiled by ANSI C compiler 11:07:54 Aug 07 1993
```

If the phrase includes '`non-ANSI`', certain aspects of the behaviour of the program may not be as expected;

Uc Self-explanatory;

Ud Show memory usage in pscomb.LOG. This can be (very) verbose, and thus create large files. It can be useful when problems occur, but is turned off by default.

Toggles from 'off' to 'on'.

10 Technical matters

10.1 Contents

10.2 Programming language

psimpoll and **pscomb** are written in ANSI `C`. The `C` language has been around since about 1972, developed first of all in the Bell Laboratories by Dennis Ritchie. Several dialects appeared subsequently, but the language is now standardized around the recommendations of the American National Standards Institute (ANSI) and the International Standards Organization (ISO). This should mean that the source code can be compiled to run on any hardware that has an ANSI-conforming `C` compiler. The source code was prepared with text files on an IBM PC, and has been compiled using the strict ANSI-compliant options of the following compilers:

Borland Turbo C++ 3.0 on IBM-compatible PC
Free Software Foundation gcc under SunOS 4.1 UNIX on Sun4m
cc under IRIX 5.1 UNIX on IP22
cc under SCO Development System on IBM-compatible PC
Symantec THINK C on Macintosh
Linux 2.0.18 - 2.2.24
Mac Os X 10.1-10.3

The use of ANSI `C` rather than QuickBASIC (as versions of **psimpoll** before `2.00`) produces a smaller executable file that runs faster and has source code that is portable.

10.3 Computers

psimpoll has been installed successfully under DOS on IBM PS/2 30, Viglen 386SX20, Viglen 486DX266, Elonex 386 SX-T, Toshiba T1200XE, and several Pentium PCs, under Macintosh on Mac Plus, Mac Classic, Mac LC II, and Performa computers, and under UNIX on Sun, IRIX, and AIX machines. It has also been installed to run under UNIX through SCO Open Desktop /

Open Server on a Viglen 486DX266, and Linux on Pentium-based machines. I would be interested to hear of other installations.

10.4 PostScript

The plotting of pollen diagrams has always been a highly individual process, dependent on locally available hardware and software. Because different printers and plotters run from different sets of control sequences, it has not been possible to transport the information needed to plot a pollen diagram except as the data itself or as a finished, camera-ready diagram. In principle, two users with identical hardware and software could exchange a plotter file, but in practice this is rarely done.

The PostScript language provides a way to describe any piece of graphics or text in a text file (with ASCII coding) that can be read and interpreted on many printers (especially, but not exclusively, laser printers). Software packages now often include a PostScript driver which can either send output directly to the printer/plotter, or generate an ASCII file with the commands to generate the work on any output device with a PostScript interpreter. Since a driver is available in many packages, and the interpreter is available in many output devices, the ASCII file provides a useful way to link the two without both being present at the same site. As an ASCII file, a PostScript file can be easily moved by e-mail.

PostScript is probably most readily available, as standard, in the Apple LaserWriter printers, but can added to many other printers by purchase of an additional card. The Apple LaserWriter is an excellent laserprinter, and it can be connected to PC-compatibles through a COM port, but it may be necessary to tinker with the wiring to overcome the incompatibility of Apple and IBM (see, for example, Glover 1989, Fig. 5).

PostScript is a 'simple interpretative programming language' (Adobe 1990, pp. 345-678), which describes the appearance of a page, whether text, graphics, or both. Adobe own the copyright in the operators and specification of PostScript, but allow anyone to use them, under certain conditions. A typical file will have a header, such as `%!PS-Adobe-3.0`, definition of sundry parameters for the page (e.g., font, character size), and definition of operators (to reduce repetition in the body of the page). `%%`, followed by a keyword, introduces a standard set of 'document structuring conventions', which are usually comments. Campbell (1993) gives a brief introduction to some commonly used commands. Operands precede their operators, so `(ABC) show` prints `ABC` at the current location. `showpage` prints the current page and prepares for the next. For graphics, `20 40 lineto` is one segment of a 'path', from the location `0 0` to `20 40` in the current co-ordinates. Text and graphics alike can be moved around on the page, rotated, and scaled. PostScript files are ASCII, so they can be edited using an appropriate editor (e.g., `EDIT` in DOS 5, or 'non-document' mode of WordStar), and I do this with text files to include special characters, but there is not too much that can be done with this kind of fiddling. More significant is the possibility of writing PostScript

drivers in any programming language, and thus producing output from almost any software that can be printed on any output device with a PostScript interpreter.

Use of PostScript thus has considerable potential for improving the ready portability of pollen diagrams, whether for publication or as part of information exchange, and that is why **psimpoll** and **pscomb** produce their output as PostScript files.

10.5 Fonts and character encoding

Characters are stored in computer systems as numbers, usually integers of less than 256 (using base 10). The character number coding is reasonably stable for numbers less than 128, using the ASCII system, although other codings do exist. Beyond 128, matters become complex, and vary between systems. PostScript can use several types of encoding, of which two are important for text characters. The default is 'Adobe Standard' encoding, and the other is 'ISO Latin 1' encoding. These produce identical output characters up to number 128 (except that a hyphen in 'Standard' is a dash in 'ISO Latin 1'), but are different for higher numbers. **psimpoll** uses ISO Latin 1 encoding, because this has most of the accented characters needed. But Standard encoding is needed for Polish suppressed-l and French ligature characters (see Table (p. 47)). In the beginning section of the PostScript file, there is a section that defines a series of fonts (F1, F2, etc) in terms of the font selected by the user, and converts the encoding from the default into ISO Latin 1. Then another series of fonts (Fa, Fb, etc) are defined without conversion, thus using the default Standard encoding. When a piece of text is written, one of the fonts is selected, and this will nearly always be from the F1 series. But if you need Polish suppressed-l or French ligature, the appropriate font from the other series will be used automatically.

The exact amount of space taken up by printed characters needs to be known in order to calculate the space needed for text in labels (e.g., zone labels). This is done with data from Adobe font metrics (AFM) files, using the versions included with WordStar 5, Sun UNIX, and via anonymous ftp from unix.hensa.ac.uk. These files also give the heights of characters, and this information is used for correctly locating certain accents.

Many accented characters are available as single characters in ISO Latin 1 encoding, but others need to be created by combining the unaccented character with the accent (e.g., the long Hungarian umlaut). Positioning the acccent accurately needs information on the width of the character and accent, height of the character, and the height of the base of the accent when printed normally. With these data, the character is printed, then the current position of the 'pen' is 'back-spaced' to the point where it should start drawing the accent centred over (or under) the character, raising the 'pen' position if necessary so that the accent will not overprint the character, and finally the accent is drawn. Then the pen position updates to the location it had after drawing the character.

The different fonts in the two series are for different sizes (title size, basic, and subscript / superscript), and for roman style and oblique style.

psimpoll uses 'Symbol' to obtain special characters, notably the circle-and-cross character in radiocarbon age columns, and also the Greek alphabet.

10.6 Compatibility

The format of main data files has not changed from previous versions of **psimpoll**, and the `C14`, `TS`, and `ZONE` associated files are also unchanged. The `ZONA` file has been abandoned, from **psimpoll** version 2.21 onwards, since **psimpoll** can now calculate zones itself, and generate a zonation file with age data included. However, the scope of the `.CFG` file has been extended considerably, so that old versions will no longer work, and may cause the program to crash, or at least work in unexpected ways. `.CFG` files created by versions of **psimpoll** earlier than `2.20` should be deleted before running the current version. **pscomb** requires input from PostScript files created by **psimpoll** version 2.20 or later.

10.7 Page layout

The default layout for a page described with PostScript is described in terms of co-ordinates with an origin (0, 0) in the bottom left corner of a page held vertically ('portrait'). The default units are 1/72 inch. Diagrams produced with **psimpoll** are laid out along the long axis of a page, so the diagram 'origin' is the upper left corner when the page is held horizontally ('landscape').

It is assumed, by default that the paper being used is A4, with length 297 mm, and width 210 mm (this can be changed in menu Ae (p. 32). Locations on the page are measured in units of 0.01 mm, necessitating a scale factor in the PostScript file. The default usable area of the page is 240 mm long and 240 / square-root(2) = 169.7 mm high, centred on the page. The length can be changed in menu Ac (p. 32). The height is adjusted automatically to maintain the same proportion, but can be changed itself (menu Ad (p. 32)). For US paper, with dimensions 11" × 8", the usable paper length should be given as 226 mm, and the usable height as 123 mm.

Within this usable area, **psimpoll** fills the space with 'boxes', building up from left to right until the space is full (at a rate that depends on the scale factor from menu A). A 'box' may be an axis, a descriptive column (e.g., sediment), or a plotted curve. Each box is preceded by a translation that shifts the origin along the page. The instructions for the image of each box are then given relative to this new origin. The default proportion of the width of the paper occupied by the diagram height is 75% of the usable area, or 127 mm. This figure can be modified in menu Ad (p. 32). The location of captions is fixed relative to the diagram, so they will move closer in if the height of the

diagram is reduced.

Within the program, all text is rotated 90° to bring it from the natural horizontal position of portrait orientation to the correct natural position for landscape orientation. Rotations specified in menu De (p. 40) (default 45°) are additional to this rotation, and control the angle of rotation from the horizontal position of landscape orientation.

The numbers in the PostScript output file giving locations on the page are thus in units of 0.01 mm. The first of each pair is the x-axis co-ordinate, increasing from the left edge of a page held in portrait location (i.e., in the direction of down the depth axis of the resulting diagram). The second number of each pair is the y-axis co-ordinate, increasing from the lower edge of a page held in portrait location (i.e., in the direction of along the resulting diagram). The co-ordinate origin shifts along the page, in the y-axis direction, for the plotting of each box.

Within the PostScript output file, boxes are demarcated by a series of comments that enable **pscomb** to break the file back into boxes and rebuild a new file by combining boxes from different files. The comment lines begin with '%', which means that the PostScript interpreter ignores them, and they are interpreted by **pscomb** as follows:

`%e` Stop interpreting the input file. Further lines are read in, but ignored until `%s` is encountered. Used to exclude the original setup and new page instructions;

`%i` Occurs within a box, and marks the beginning of lines that **pscomb** should ignore. Used to exclude the instructions that draw zone lines across the page, for example;

`%j` Occurs within a box, and marks the end of lines that **pscomb** should ignore.

`%n` Marks the start of a new box (ending the previous box), and is followed on the same line by six characters that give the name of the box, as displayed on the screen while the file is being read and in the interactive menu;

`%s` Start interpreting the input file. Lines at the beginning of the file, and after `%e` are ignored until `%s` is encountered;

`%t` Immediately followed by a number that gives values for the translation along the page of the box in units of 0.01 mm. This enables **pscomb** to accumulate the proportion of a page that it has filled with boxes, and to determine whether it has enough room for the next one.

pscomb produces a new output file by combining boxes from the input files with new set of the document structuring conventions, instructions for ending and starting pages, and adding titles and general footnotes.

10.8 Known problems and limitations

- (DOS only) **pscomb** will crash, complaining about calloc or realloc memory allocation problems if the size of the input files plus the size of the program exceeds available conventional memory (normally about 500kB). It may be possible to work around this by using **pscomb** on individual input files to reduce them to the minimum size necessary.
- Smoothing works after printing of dataset to a file, so smoothed data cannot be output.
- The program does not assume that it is being run on a computer that supports ASCII character coding. However, PostScript interpreters do assume that input files are ASCII. So, output from a non-ASCII computer might need to be translated before presentation to a PostScript printer.
- The handling of output to screens assumes that screens are 25 lines high, with 80 characters per line. The program should work correctly with screens of other dimensions, but the appearance may be odd.
- It is not possible to have more than two columns of numbers in sum columns.
- Sediment columns cannot be plotted against age.
- Script file names are associated with input file names. In future, they may be associated with output file names in order to connect a particular output file more directly with the commands used to create it.
- The anglo 'point' is used as decimal separator. Arguably, **psimpoll** should be able to use the European 'comma', depending on locale, but unfortunately this is not currently possible.

11 Error messages

11.1 Contents

11.2 General information on errors

Running **psimpoll** involves an interaction between a user and the software, and an interaction between a computer and the software. It is the aim in writing the software to be aware of problems that might arise in either of these interactions and anticipate them. Where possible, some corrective action is taken, and the program will continue. In other situations, this is not possible, and the program will stop with a `Program ending` message. The sections below include messages that might arise when the program stops unexpectedly, and give more detail of why they might have occurred. Other messages may be given by the operating system, depending on why the failure occurred, and which operating system is being used. Problems also occur in other ways, but enabling **psimpoll** to continue. These may give an explanatory message on the screen if the program realised something odd was happening, or else may be evident as unsatisfactory output.

Errors are likely to fall into one or more of the following categories:

- The program could not interpret the input file(s) in some way: incorrect specification of numbers of taxa or samples, wrongly ordered, incomplete. This is by far the commonest source of problems. Failure by user, or failure to communicate what the program expected. Please check the former first!
- The program did not anticipate some (perfectly legitimate) combination of options or some type of (perfectly legitimate) data input. Program incomplete: complain;
- The program failed to do something this documentation claimed it should do. This may be evident as a crash, or erroneous output. Complain.

If you encounter an error, whether it produces a crash, or just wrong output, please follow the following steps:

1. Try to replicate the error, to ensure that the problem was not simply through a typing mistake at the terminal;

2. Note carefully when the error appears, by reading messages that appear on the screen, and taking a look in any output files. **psimpoll** is moderately verbose, and this is one of the uses for all that stuff. Note the message itself, especially any preceding information on file and line number(see below);
3. Check all data input;
4. Explore the behaviour of the problem: does it occur with other inputs? If so, when does it appear, and when not?
5. Bring evidence to KDB: include all input files.

Crashes can appear in several ways. Wherever possible, you will get an error message, usually from the program, occasionally from the operating system. However, it is also possible that the computer will just 'lock up': no action evident at all. I have made **psimpoll** verbose to help detect this. Or other strange things can happen: unexpected rebooting, running of disc drives, etc. Watch for screen output that includes garbage characters as a warning of a problem. All these circumstances should be reported to KDB. If the program does crash, you may need to reboot your computer.

11.3 List of messages

A list of messages that precede an unexpected ending follows. It is ordered strictly alphabetically, in some cases starting from after a phrase of the form '`ppmain.c 288`', which gives information on the program file and line number that detected the error. To see this, run **psimpoll**, but give an erroneous file name. Unless otherwise stated, any of these messages might occur with either **psimpoll** or **pscomb**. Select the message of interest to get more information.

- agedepth (timescale): failed to converge on a value for t after 10000 iterations (p. 96)
- asin (p. 96)
- atan (p. 96)
- atan2 (p. 96)
- bernshtein: failed to converge on a value for t after 10000 iterations (p. 96)
- bus error (p. 96)
- CA: residual bigger than tolerance (p. 96)
- calloc: memory allocation error (p. 96)
- core dump (p. 96)
- Data set for analyses has zero samples (p. 96)

- Error estimates on ages must be greater than zero, cannot complete analysis. Check values in C14 file (p. 97)
- exp (p. 97)
- Fatal error detected by operating system: please report to KDB SIGABRT: Abnormal termination (p. 97)
- Fatal error detected by operating system: please report to KDB SIGFPE: Floating point exception (p. 97)
- Fatal error detected by operating system: please report to KDB SIGILL: Invalid function image (p. 97)
- Fatal error detected by operating system: please report to KDB SIGSEGV: Invalid access to storage (p. 97)
- Fatal error detected by operating system: please report to KDB SIGTERM: Termination request received (p. 97)
- fclose: error in closing a file (p. 98)
- fgetc: no character read (end of file?) (p. 98)
- fgets: no string read (end of file?) (p. 98)
- First character of two-character code on line 4 of input must be one of %, A, B, C, G, L, M, or X. See documentation for details. The character in your file is 'c' (p. 98)
- fopen: failed to open file (p. 98)
- fscanf: unexpected end of file (p. 98)
- getbuf: buffer is overfull. Check that individual input items do not exceed allowed number of characters. (p. 98)
- getbuf: unexpected end of file. (p. 98)
- getc: reading error (end of file?) (p. 99)
- indsplit: more than the allowed maximum of 20 splits found. Please report to KDB (p. 99)
- Input box is greater than 63000 characters long (p. 99)
- int conversion: floating point number cannot be represented as integer (p. 99)
- lfit (gaussj): singular matrix, cannot complete analysis (p. 99)
- linint_s: two depths are identical or reversed (p. 99)
- loess: extrapolation beyond date limits is not allowed. Define an age for top and basal samples in menu L. (p. 99)
- log10 (p. 99)

- log (p. 99)
- Memory allocation error. Cannot load COMMAND, system halted (p. 99)
- Number of samples is zero or negative (p. 99)
- Number of sum taxa is greater than 10 (p. 99)
- Number of taxa is zero or negative (p. 100)
- pow (p. 100)
- ran4_s: this function is somewhat non-portable Output on your computer appears unreliable Use one of the other random number generators (p. 100)
- Rate of change analysis: the dissimilarity measure in use has exceeded its theoretical maximum upper limit (p. 100)
- realloc: attempt to reallocate to a negative amount of space (p. 100)
- realloc: memory allocation error (p. 100)
- scanf: input failure (p. 100)
- Second character of two-character code on line 4 of input must be one of A, D, M, S, or T. See documentation for details. The character in your file is 'c' (p. 100)
- Sediment accumulation rate(s) have been calculated as negative This will result in age reversals, and/or negative pollen accumulation rates. Check output files for details, and reconsider age model (p. 100)
- Segmentation fault (p. 101)
- SIGINT: interactive attention signal from user (p. 101)
- spline_s: two depths are identical or reversed (p. 101)
- splint: bad xa[] input, cannot complete analysis. Check depths in C14 file (p. 101)
- sqrt (p. 101)
- strcat: total string length too long (p. 101)
- strftime: failed to fill array (p. 101)
- svdfit (svdcmp): no convergence in 30 iterations of its, abnormal exit (p. 101)
- svdfit (svdcmp): zu() needs augmenting with extra zero rows, abnormal exit (p. 101)
- switch: unexpected value (p. 101)
- SYSTEM: (p. 101)

- TILIA format: line-length unexpectedly short. Check input file! (p. 102)
- time: calendar time not available (p. 102)
- Unexpected end-of-file reached: data set is incomplete (p. 102)
- ungetc: error pushing character back (p. 102)
- vprintf/vfprintf: output error (p. 102)

11.4 Error messages

11.4.1 Explanation of messages

- agedepth (timescale): failed to converge on a value for t after 100 iterations Timescales are generated by an iterative process that should have converged to give an answer within a specified accuracy after 16-20 iterations. In this case, the process has not converged after 100 iterations. Please report to KDB.
- asin Program tried to obtain arc sine of a number < -1 or > +1. Probably a program error: report to KDB.
- atan Program tried to obtain arc tangent of a number < -pi / 2 or > pi / 2. Probably a program error: report to KDB.
- atan2 Program tried to obtain arc tangent of y / x < -pi or > pi. Probably a program error: report to KDB.
- bernshtein: failed to converge on a value for t after 100 iterations **psimpoll** message. Bernshtein polynomials are solved by convergence that should be complete within 20 iterations. This one has failed. Check the output file, and if the problem occurred with samples beyond the range of the radiocarbon age points, trying fixing the end points with age estimates from linear interpolation. It is also possible that the initial guesses **psimpoll** makes for t were too far out, so the way these are obtained needs to be modified, or an age-depth range given in menu H (p. 43) caused problems. See KDB.
- bus error UNIX error message. The program has crashed completely. Report to KDB.
- CA: residual bigger than tolerance Correspondence analysis error message. The program has failed to converge on a solution. Check all input values and transformation options.
- calloc: memory allocation error Program was unable to make the allocation of memory for an array. Check that you have given the numbers for taxa, samples, zones, radiocarbon ages, etc, correctly (these numbers determine the size of memory requested from the computer for arrays). Consider the possibility that your computer's memory is not adequate, then report to KDB.
- core dump UNIX error message. The program has crashed completely. Report to KDB.
- Data set for analyses has zero samples **psimpoll** message. The program produces a reduced dataset for numerical analysis by eliminating samples outside any defined section (menu H (p. 43)), and samples marked for omission. After doing this, there were no samples left. The most likely problem is that you requested a section of the dataset by age, but the samples are presented by depth and were not converted. Please rerun.

- Error estimates on ages must be greater than zero, cannot complete analysis. Check values in C14 file **psimpoll** message. The program has read a standard deviation for an age from your `C14` file as 0 or negative, and this caused a problem in the calculation of an age model. If you want an 'age' to be excluded from analysis, make sure it begins with a non-numeric character (i.e., `(4000)` not `4000`). Otherwise, check the file carefully.

- exp Attempt to calculate an exponential failed because the value was too high. Probably a program error: please report to KDB.

- Fatal error detected by operating system: please report to KDB SIGABRT: Abnormal termination This is a serious crash, brought about by the program trying to do something that the computer could not handle. All such conditions should be trapped by the program, so please report this one to KDB immediately, even if you know of something horrible you did that might have caused it to behave this way. The computer may need rebooting to clear memory.

- Fatal error detected by operating system: please report to KDB SIGFPE: Floating point exception This is a serious crash, brought about by the program trying to use a number outside the range that the computer can handle (too large or divide by zero). All such conditions should be trapped by the program, so please report this one to KDB immediately, even if you know of something horrible you did that might have caused it to behave this way. The computer may need rebooting to clear memory.

- Fatal error detected by operating system: please report to KDB SIGILL: Invalid function image This is a serious crash, brought about by the program trying to do something that the computer could not handle. All such conditions should be trapped by the program, so please report this one to KDB immediately, even if you know of something horrible you did that might have caused it to behave this way. The computer may need rebooting to clear memory.

- Fatal error detected by operating system: please report to KDB SIGSEGV: Invalid access to storage This is a serious crash, brought about by the program trying to do access memory in some way that the computer could not handle. All such conditions should be trapped by the program, so please report this one to KDB immediately, even if you know of something horrible you did that might have caused it to behave this way. The computer may need rebooting to clear memory.

- Fatal error detected by operating system: please report to KDB SIGTERM: Termination request received This is a serious crash, brought about by the program trying to do something that the computer could not handle. All such conditions should be trapped by the program, so please report this one to KDB immediately, even if you know of something horrible you did that might have caused it to behave this way. The computer may need rebooting to clear memory.

- fclose: error in closing a file An error was reported to the program when it tried to close a file. May occur if a disc that you are writing to is full (so please check this), but probably a program error: report to KDB.
- fgetc: no character read (end of file?) An error was reported to the program when it tried to read a character from a file. Check that your input files are complete.
- fgets: no string read (end of file?) An error was reported to the program when it tried to read text from a file. Probably a program error: report to KDB.
- First character of two-character code on line 4 of input must be one of %, A, B, C, G, L, M, or X. See documentation for details. The character in your file is 'c' **psimpoll** message. Check line 4 of your main input file, and ensure that the first character on the line matches one on this list. The character read by the program is indicated at the end of the message. And check that the file name you entered was the main input file you intended.
- fopen: failed to open file An error was reported to the program when it tried to open a file. Usually through mistyping the name of the main input file, but might occur if two files had the same name (from menu E (p. 41)).
- fscanf: unexpected end of file An error was reported to the program when it tried to read from a file. Probably a program error: report to KDB.
- getbuf: buffer is overfull. Check that individual input items do not exceed allowed number of characters. This mesage is preceded by more detail of what the program was trying to read when the problem occurred, what it actually did read, and, where possible, suggestion of what needs to be corrected. This information should help to locate the problem, but note that the 'real' problem may lie earlier in the file than the input line that was the immediate source of the difficulty. getbuf is a function that handles the reading of many input items. Its internal buffer is 160 characters long, which is ample for all reasonable and correctly formatted input. The 'individual input items' referred to may be values or labels, which may occupy lines by themselves or be parts of lines with other values. Check your input files carefully to make sure that you haven't omitted spaces (to separate numbers) or commas (to separate labels), and that the values given for the number of taxa, number of samples, and the two-character code on main input line 4 are correct for the data given. See the output file `psimpoll.LOG` for the contents of the buffer, which may help work out where the problem occurred, along with information on the screen at the time.
- getbuf: unexpected end of file. The end of a file was reached unexpectedly. This messge should have been preceeded by message with details of the file and line number concerned. Check that the file has all the input items expected.

- getc: reading error (end of file?) An error was reported to the program when it tried to read from a file. Probably a program error: report to KDB.
- indsplit: more than the allowed maximum of 20 splits found. Please report to KDB **psimpoll** message. Your data set has generated more than 20 splits in one of the taxa. This is more than anticipated would ever occur. Report to KDB and ask for the program to be modified.
- Input box is greater than 63000 characters long **pscomb** message. There is a limit of 63000 characters for the size of an input box, and one of those in your PostScript input file exceeds this. Please report to KDB, and this will encourage him to provide a greater limit.
- int conversion: floating point number cannot be represented as integer A floating point number was too large to be converted to an integer. Probably a program error: please report to KDB.
- lfit (gaussj): singular matrix, cannot complete analysis **psimpoll** message. One of the matrices generated during construction of an age model by weighted least-squares cannot be analysed. This may be a data error, or just bad luck. Check the data, and if that doesn't work, try using singular value decomposition (menu Ld.4 (p. 56)).
- linint_s: two depths are identical or reversed **psimpoll** message. Problem occurred while trying to obtain linearly interpolated ages. Check depth values in `C14` file, and ensure that the ages given for the top sample (defaults to $0±50$), and basal sample (no default) are reasonable (see menu L (p. 53)).
- loess: extrapolation beyond date limits is not allowed. Define an age for top and basal samples in menu L This technique for age-depth modelling cannot extrapolate. Check and, if necessary define, ages for the top and bottom sample (which then get treated as if they were dates) so that there will be no extrapolated vales.
- log10 Program tried to obtain base 10 logarithm of 0 or a number < 0. Probably a program error: report to KDB.
- log Program tried to obtain base e logarithm of 0 or a number < 0. Probably a program error: report to KDB.
- Memory allocation error. Cannot load COMMAND, system halted DOS error message, probably occurring after one of the program messages. The computer will need to be rebooted before you can do anything else. Probably a program error: report to KDB.
- Number of samples is zero or negative The number of samples specified on line 3 of the main input file (or elsewhere in a Tilia file) has been given as zero or negative. This needs to be corrected.
- Number of sum taxa is greater than 10 The number of taxa included in a summary diagram is limited to 10, which was expected to be enough under all circumstances. If you really want more than 10, and can

convince me that this will work when plotted, I will make the necessary change.

- Number of taxa is zero or negative The number of taxa specified on line 2 of the main input file (or elsewhere in a Tilia file) has been given as zero or negative. This needs to be corrected.

- pow Program tried to obtain x to the power of y with x negative and y not an integer, or an overflow occurred (number too large for the computer to handle). Probably a program error: report to KDB.

- ran4_s: this function is somewhat non-portable Output on your computer appears unreliable Use one of the other random number generators The 'pseudo-DES' random number generator (ran4_s) is not completely portable. It works fine on Macintosh, IBM PCs, and the UNIX systems that I have tried, but the version of **psimpoll** you are using has been compiled on some other system. As the message says, use another generator. This one may be unreliable on your system.

- Rate of change analysis: the dissimilarity measure in use has exceeded its theoretical maximum upper limit **psimpoll** message. Some of the dissimilarity measures used in rate of change analysis have a theoretical upper bound, and this will never be exceeded with correct input and correct programming. There has been a failure in one of these areas. Check that you are using only proportional data as input. Then report to KDB.

- realloc: attempt to reallocate to a negative amount of space Failure of attempt to reallocate memory that was previously allocated because the new size of memory requested is negative. Probably a program error: report to KDB.

- realloc: memory allocation error Failure of attempt to reallocate memory that was previously allocated. Check the lengths of text items that you have included in (especially) the main input file to see if they are really excessive. In **pscomb** it may mean that a box from one of the input files is more than 31,000 characters long. But probably a program error: report to KDB.

- scanf: input failure An error was reported to the program when it tried to read from the keyboard. Probably a program error: report to KDB.

- Second character of two-character code on line 4 of input must be one of A, D, M, S, or T. See documentation for details. The character in your file is 'c' **psimpoll** message. Check line 4 of your main input file, and ensure that the second character on the line matches one on this list. The character read by the program is indicated at the end of the message. And check that the file name you entered was the main input file you intended.

- Sediment accumulation rate(s) have been calculated as negative This will result in age reversals, and/or negative pollen accumulation rates. Check output file (if any) for details, and reconsider age model

psimpoll message. The program was asked to calculate accumulation rates, and finds that the model in use produces negative rates. The program continues, and the full gory details will be in the age-depth output files if you requested them. Note that the pollen diagram will still be plotted against age, but will have overlapping samples.

- Segmentation fault UNIX error message. The program has crashed completely. Report to KDB.
- SIGINT: interactive attention signal from user User has interrupted normal running of program.
- spline_s: two depths are identical or reversed **psimpoll** message. Problem occurred while trying to obtain an age model with a spline. Check depth values in `C14` file, and ensure that the ages given for the top sample (defaults to 0±50), and basal sample (no default) are reasonable (see menu L (p. 53)).
- splint: bad xa[] input, cannot complete analysis. Check depths in C14 file **psimpoll** message. Spline curve-fitting is a two stage process. splint is the second stage, and it has found that two input depth values are identical. Check input from your `C14` file.
- sqrt Program tried to obtain the square-root of a number < 0. Probably a program error: report to KDB.
- strcat: total string length too long A concatenated 'string' of characters has become too long for the amount of memory allocated for it. Probably a program error: report to KDB.
- strftime: failed to fill array A function that formats representations of data and time for output has failed. Probably a program error: report to KDB.
- svdfit (svdcmp): no convergence in 30 iterations of its, abnormal exit **psimpoll** message. The program could not complete analysis of your proposed age-depth model using singular value decomposition. Check the data in your `C14` file, and try using a weighted least-squares age-depth model. If this fails, report to KDB.
- svdfit (svdcmp): zu() needs augmenting with extra zero rows, abnormal exit **psimpoll** message. Program error, failing to trap analysis of age-depth model by singular value decomposition with more terms in the model than ages. Report to KDB.
- switch: unexpected value Program does not allow for all possible cases in a 'switch' statement. This is one likely error message if your `.CFG` file is out of date. If deleting any `.CFG` file does not remove the problem, report to KDB.
- SYSTEM: The text on lines beginning like this is an error message from the computer system, and is therefore specific to that system. Note what it says, and report to KDB. Information from **psimpoll** or **pscomb** should be on the next line down, and more details may be found by looking up that message elsewhere in this section.

- TILIA format: line-length unexpectedly short. Check input file! The program recognised a file as TILIA format, but found that one of the lines with taxa names was too short. There should be at least 16 characters on these line, with a letter indicating the (sub)sum group as character 16. Taxon names begin at character 18. It is acceptable (but strange?) to have no taxon name, but the (sub)sum value must be present. Check the input file, make sure that it is indeed TILIA format, and look at the contents of the list of taxa names.

- time: calendar time not available The program attempts to obtain the calendar time from the operating system, but this appears not to be available for some (unknown) reason. Please report to KDB.

- Unexpected end-of-file reached: data set is incomplete An input file does not have as many items as it should have. If the file named immediately above this message is a `.CFG` file, it is probably out of date. Use the example (p. 107) `.CFG` file in to work out which features are set in your file, then delete it, and reset them from the menus on a new run of **psimpoll**. If the named file is not a `.CFG` file, check its length against the number of items that are indicated at the top (number of taxa, number of samples, number of zones, etc., as relevant).

- ungetc: error pushing character back This function is used when reading 'garbage' characters from a file. The first non-garbage character is 'pushed back', so that it can be read by the next input routine. There has been a problem with the 'push back' process, probably a programming error. Report to KDB.

- vprintf/vfprintf: output error An error has been reported by a function that 'prints' data from a variable length list of parameters. Program error: report to KDB.

12 Sample input files

12.1 Contents

The sample **psimpoll** input files below illustrate most of the features described above. Items are numbered on the left to correspond to the item numberings used in Data preparation (p. 16), and should be compared with them. Item numbering should, of course, be omitted from actual files.

12.2 Data presented as proportions

```
1 Dallican Water, Lunnasting, Shetland
2 8
3 4
4 %d
5 720 716 712 708
6 _Betula, 0.2004 0.0674 0.1635 0.3001
6 -_Juniperus communis, 0.0118 0.0019 0.0115 0.0058
6 Cyperaceae, 0.1886 0.0520 0.1885 0.0996
6 _Potentilla_ t., 0.0000 0.0058 0.0096 0.0029
6  <
  4
  Sum Trees & Shrubs, 0.4401 0.2505 0.2788 0.4286
  Sum Ericales, 0.0255 0.0039 0.0654 0.1400
  Sum Herbs, 0.5226 0.7380 0.5058 0.2771
  Sum Filicales, 0.0118 0.0077 0.1500 0.1544
6 !Palynol. richness
  250
  20.1672 17.0265 23.3079
  19.4393 15.7504 23.1283
  20.3516 17.5684 23.1348
  14.7937 12.1803 17.4070
6 =Rate of change, 0.366249 0.420106 0.236846 -1
6 *Main sum, 509 519 520 693
```

12.3 Data presented as concentrations for conversion to accumulation rates

```
1 Dallican Water, Lunnasting, Shetland
2 3
3 4
4 bd
5 720 716 712 708
6 _Betula, 6595 2702 10670 48175
6 #Sum Trees & Shrubs, 14095 9882 18202 68788
6 &Charcoal, 0 0.399084 0.055816 0.064629
7 -49333
8 -1
9 -0.5
10 -1
11 763 639 393 213
12 -0.5
13 -1
14 -1
```

12.4 TS file

```
1 3
2 Ld
  4
  708
2 Ld
  2
  Ag
  2
  718
2 Ag
  3
  Ga
  1
  720
3 5YR6/4
  5YR3/4
  5Y5/1
```

12.5 C14 file

```
1 4
2 433
  439
  1565
  65
2 478
  486
  3350
  70
2 504
  504
  Hekla?

2 698
  710
  9350
  90
```

12.6 CAL file

```
1 4
2 433
  439
  dallCL01
2 478
  486
  dallCL02
2 698
  710
  dallCL03
```

12.7 CAL input file

Normally written by BCal (so line-numbering omitted). But can be created in by hand. Content consists solely of a series of pairs of values. The first of each pair is an age in calendar years (negative earlier than 0BP), and the second is a probability. The pairs and the values by separated by spaces, TABs or newlines. A very simple (but useful) can be:

```
-1200 1.0
```

which means that the age is 1200 with 100% probability. More usually, the file will look like this (topmost lines only):

```
-6868 1.982710762154017E-5
-6867 1.982710762154017E-5
-6866 0.0
-6865 0.0
-6864 0.0
-6863 0.0
-6862 0.0
-6861 0.0
-6860 0.0
-6859 1.982710762154017E-5
-6858 1.982710762154017E-5
-6857 0.0
-6856 1.982710762154017E-5
-6855 0.0
-6854 1.982710762154017E-5
-6853 3.965421524308034E-5
-6852 1.982710762154017E-5
```

12.8 ZONE file

```
1 4
2 D-4
  D-3
  D-2
  D-1
3 436.000
  547.000
  706.000
4 red
  yellow
  green
  blue
```

12.9 .SCR file

```
0
1b
2e
3f
4g.2
5g.2eb
```

12.10 .CFG file

A .CFG file includes the current values for all variables that can be modified by the menus. An example is shown below, continuing on the next page. Numbers and text beginning down the left margin of the page are the file contents. Labels at the right are the names of the **psimpoll** variables that hold these numbers. Some string variables may be empty, and where that happens in this particular file there is no entry in the left margin. Lines beginning '...' should be read as continuing from the line above. A separate list of the variables (p. 113) gives more information about the type of quantity each contains.

```
0.002000                                gl.page.minpl
24000.000000                            gl.page.wide
0.750000                                gl.page.prop
0                                       coll.style
0                                       coll.ix10
a:\dallAD                               agedep.fname
a:\dallp.PS                             outmain.fname
H                                       gl.font.code
0.000000                                gl.font.ufac
45                                      coll.rot
```

```
                                        head.base
                                        head.subtitle
                                        head.zone
                                        head.sed
0 -1                                    coll.join  intfg
-1 0.000000 50.000000 0.000000 50.000000 3 5 1.000000
                                ... age.flag age.up age.upsd age.base
                                ... age.basesd age.model age.npterm age.c14k
a:\dallZONE                             zone.fname
a:\dallC14                              inage.fname
a:\dallTS                               ints.fname
0 0.300000 1                            ci.flag ci.cvsed ci.n
1                                       gl.mran
TIMER                                   gl.seed
100                                     ci.nran
0 0                                     rarer.flag rarer.n
-1 0.050000 5 5 D-        zoner.flag  num.thresh  zoner.n zoner.meth  zoner.pre
0 1                                     roc.flag  roc.diss
{14}C ages (yr BP)                      head.age[C14BP]
Depth                                   head.ps1[DEPTH]
cm                                      head.ps2[DEPTH]
0                                       num.stat
a:\dallSTAT                             stat.fname
0                                       pca.flag
3                                       pca.matrix
6                                       pca.naxes
0                                       four.flag
0                                       lev.adsc
0                                       pca.tran
0                                       zoner.tran
-1                                      coll.indspl
0                                       zoner.ransam
-1                                      gl.recalc
0.500000                                coll.shade
-1 400.000000 700.000000                lev.udata lev.umax lev.umin
0                                       coll.all
0                                       ipdat
0                                       roc.smooth
0.000000 0                              coll.hist coll.indscale
grains cm{-3}                           sunit[0]
grains g{-1}                            sunit[1]
grains cm{-2} year{-1}                  sunit[2]
cm{2} cm{-3}                            sunit[3]
cm{2} cm{-2} year{-1}                   sunit[4]
cm{2} g{-1}                             sunit[5]
cm{2} grain{-1}                         sunit[6]
cent.{-1}                               sunit[7]
ratio                                   sunit[8]
power                                   sunit[9]
0                                       four.tran
```

```
0                                   coll.script
0                                   coll.bull
0.500000                            coll.lval
0                                   age.timx
Age ({14}C yr BP)                   head.time[C14BP]
0                                   gl.sum2col
0 10                                roc.model.ci roc.model.n
0 10                                pca.model.ci pca.model.n
0 10                                zoner.model.ci zoner.model.n
0.000000 0.000000                   lev.majmark lev.minmark
0                                   coll.skip
0                                   age.adbc
0                                   gl.tufte
-1                                  coll.c14scale
a:\dallDAT                          dat.fname
0                                   ipform
1.000000                            coll.penwid
0                                   coll.interp
1.000000                            lev.interp
0.000000                            lev.intstart
0.000000                            lev.intstop
a:\dallAD.ps                        adplot.fname
0.002                               gl.adpage.minpl
Age ({14}C yr BP)                   head.adxaxis[C14BP}
Depth (cm)                          head.adyaxis
1                                   gl.palette.back
0                                   gl.palette.fore
0                                   coll.colpl
0                                   coll.coltit
0                                   coll.coldep
0                                   coll.col14c
0                                   coll.colts
0                                   coll.colzone
0                                   coll.colxax
-1                                  gl.colts
-1                                  gl.colzone
-0.99                               gl.neglev
0                                   agerange.colpl
0                                   agerange.colsdpl
0                                   agerange.col14c
0                                   agerange.coldep
0                                   agerange.colxax
0                                   lev.inttype
21000.000000                        gl.papwid
29700.000000                        gl.papht
0                                   ca.flag
0                                   ca.tran
0                                   ca.iweigh
0                                   ca.irescfg
4                                   ca.iresc
```

```
0                                       ca.dca
26                                      ca.mk
0.000000                                ca.resth
0 10                                    ca.model.ci ca.model.n
dallCAL                                 incal.fname
0                                       age.cal
Age (Cal. yr BP)                        head.adxaxis[CALBP]
Cal. ages (yr BP)                       head.age[CALBP]
{14}C ages (yr AD/BC)                   head.age[C14AD]
Cal. ages (yr AD/BC)                    head.age[CALAD]
Age ({14}C yr AD/BC)                    head.adxaxis[C14AD]
Age (Cal. yr AD/BC)                     head.adxaxis[CALAD]
Age                                     head.ps1[C14BP]
yr BP                                   head.ps2[C14BP]
Age                                     head.ps1[CALBP]
yr BP                                   head.ps2[CALBP]
Age                                     head.ps1[C14AD]
yr AD/BC                                head.ps2[C14AD]
Age                                     head.ps1[CALAD]
yr AD/BC                                head.ps2[CALAD]
0                                       ci.negar
-1                                      gl.calval
1000000                                 ci.looplim
Age (Cal. yr BP)                        head.time[CALBP]
Age ({14}C yr AD/BC)                    head.time[C14AD]
Age (Cal. yr AD/BC)                     head.time[CALAD]
                                        head.dendro
 Total dispersion                       sunit[10]
1000                                    gl.adres
0                                       age.ngterm
0                                       gl.showmem
```

12.11 psimpoll.COL file

This file contains the colour palette used by the program, and consists of colours defined using the CMYK system, and keyed to an identifier and a name. Subsequent use of the colours needs only the name (in most cases) or the identifier (rarely).

```
# psimpoll 3.00 colour palette
# K.D. Bennett 1998
#
# Colours are made up of the four components of the CMYK scheme
# (Cyan Magenta Yellow blacK) on a scale from 0.0 (none of that component)
# to 1.0 (full expression of that component)
#
# The first line of values must contain a single value giving the number
```

```
# of calours described
#
# Then each colour is described on a line by itself, identified by a number,
# which may be of any positive value from 0 to 2147483647
#
# The order of values is:
# identifier cyan magenta yellow black name
#
# All lines beginning '#' are ignored (comment lines)
#
# It is recommended that the default settings are not changed
#
220
0 0.0 0.0 0.0 1.0 black
1 0.0 0.0 0.0 0.0 white
2 1.0 0.0 0.0 0.0 cyan
3 0.0 1.0 0.0 0.0 magenta
4 0.0 0.0 1.0 0.0 yellow
5 0.0 1.0 1.0 0.0 red
6 0.0 1.0 0.4 0.0 ruby
7 0.0 0.4 1.0 0.0 orange
8 1.0 0.0 1.0 0.0 green
9 0.4 0.0 1.0 0.0 lime
10 1.0 0.4 0.0 0.0 blue
11 0.4 1.0 0.0 0.0 purple
12 1.0 1.0 0.0 0.0 violet
13 1.0 0.0 0.4 0.0 turquoise
14 0.5 0.75 1.0 0.0 brown
15 0.0 0.0 0.0 0.5 grey
# Munsell colours follow, taken from the Munsell to CMYK conversion program
# available from Munsell (R) at http://www.munsell.com
180 0.20 0.21 0.20 0.17 N8/0
171 0.19 0.20 0.20 0.34 N7/0
160 0.19 0.20 0.20 0.48 N6/0
150 0.19 0.20 0.20 0.59 N5/0
140 0.19 0.20 0.20 0.67 N4/0
130 0.19 0.20 0.20 0.73 N3/0
120 0.19 0.20 0.20 0.75 N2.5/0
10068 0.20 0.64 0.74 0.12 10R6/8
10066 0.20 0.53 0.61 0.21 10R6/6

# Many lines cut out for conciseness in this the manual

5253 0.19 0.24 0.35 0.54 5Y5/3
5252 0.19 0.23 0.30 0.56 5Y5/2
5251 0.19 0.21 0.25 0.57 5Y5/1
5244 0.19 0.24 0.33 0.63 5Y4/4
5243 0.19 0.23 0.30 0.64 5Y4/3
5242 0.19 0.22 0.27 0.65 5Y4/2
5241 0.19 0.21 0.24 0.66 5Y4/1
```

```
5232 0.19 0.21 0.24 0.71 5Y3/2
5231 0.19 0.20 0.22 0.72 5Y3/1
52252 0.19 0.21 0.23 0.74 5Y2.5/2
52251 0.19 0.20 0.21 0.74 5Y2.5/1
```

13 Variables saved in .CFG files

A list of the variables saved in .CFG files follows, in alphabetical order, explaining: what the variable does; which part of the menu sets the variable; type of quantity (integer [int], real number [real], or multiple characters [string]); default setting when **psimpoll** is run without a .CFG file; and allowed range. Variable names that include a period are members of structures. Integer variables that can take only values 0 or -1 are 'ON' (i.e., the relevant action is performed) when set at -1.

adplot.fname: output file for age-depth plots; Eh; string; a:\dallAD.ps; any text [must be legal on the operating system being used, but that is not checked by the menu].

age.adbc: flag for plotting age scale in AD/BC units rather than BP; Li; int; 0; 0 or -1.

age.base: age for base of lowest sample in sequence; Lc; real; 0; 0-10^7.

age.basesd: standard deviation for age of lowest sample in sequence; Lc; real; 50; 0-2000.

age.c14k: laboratory error multiplier for radiocarbon ages; Lg; real; 1; 0-5.

age.cal: flag for reading ages from CAl file instaead of C14 file; Lj; int; 0; 0 or -1.

age.flag: flag for conversion of depths to ages; La; int; 0; 0 or -1.

age.model: selection of age model; Ld; int; 1-5 (see menu).

age.ngterm: number of terms used in Gaussians for age.model 6; Lf; int; 0-12.

age.npterm: number of terms used in polynomials for age.model 3 or 4; Le; int; 0-12.

age.timx: flag for addition of a timescale next to the depth scale; Lh; int; 0; 0 or -1.

age.up: age for uppermost sample; Lb; real; 0; -1000-100000.

age.upsd: standard deviation for age of uppermost sample; Lb; real; 50; 0-2000.

agedep.fname: output file for age-depth conversion results; Eb; string; a:\dallAD; any text [must be legal on the operating system being used, but that is not checked by the menu].

agerange.coldep: colour identifier for depth axis; Cf (age-depth); long; 0; none.

agerange.colpl: colour identifier for main curve; Cc (age-depth); long; 0; none.

`agerange.colsdpl`: colour identifier for confidence interval curves; Cd (age-depth); long; 0; none.

`agerange.col14c`: colour identifier for data points; Ce (age-depth); long; 0; none.

`agerange.colxax`: colour identifier for x-axis; Cg (age-depth); long; 0; none.

`ca.dca`: flag for detrended correspondence analysis; Mea; int; 0; 0 or -1, entered as 1 or 0, respectively.

`ca.flag`: flag for correspondence analysis; Me.0; int; 0; 0 or -1, entered as 1 or 0, respectively.

`ca.iresc`: number of times to rescale axes in correspondence analysis; Med; int; 4; 0-20.

`ca.irescfg`: flag for rescaling axes in correspondence analysis; Mec; int; 0; 0 or -1, entered as 1 or 0, respectively.

`ca.iweigh`: flag for downweighting rare taxa in correspondence analysis; Meb; int; 0; 0 or -1, entered as 1 or 0, respectively.

`ca.mk`: number of segments used in detrended correspondence analysis; Mee; int; 25; 10-46.

`ca.model.ci`: flag for modelling correspondence analysis; Meh.0; int; 0; 0 or -1, entered as 1 or 0, respectively.

`ca.model.n`: number of models to carry out for correspondence analysis; Meh.1; int; 10; 1-1000.

`ca.tran`: method of transformation of dataset before CA; Meg; int; 0; 0-4 (see menu).

`ci.cvsed`: annual sediment variability; N1; real; 0, but 0.03 if the data is presented as concentrations for conversion to accumulation rates; 0-1.

`ci.flag`: flag for calculation of confidence intervals; N; int; 0; 0 or -1, entered as 1 or 0, respectively.

`ci.looplim`: number of attempts to achieve an age-depth model without negative accumulation rates before restarting; N7; long int; 1000000; 1-LONG_MAX (which is at least 2147483647).

`ci.n`: method for calculating confidence intervals; N2; int; 1; 1 or 2 (see menu).

`ci.negar`: flag for allowing negative accumulation rates in age-depth modelling; N6; int; 0; 0 or -1, entered as 1 or 0, respectively.

`ci.nran`: number of simulations for calculation of confidence intervals; N5; int; 100; 10-1000.

`coll.all`: flag for considering all samples for scaling; Bb; int; 0; 0 or -1, entered as 1 or 0, respectively.

`coll.bull`: flag for placing a solid dot (bullet) next to samples with low values; Dh; int; 0; 0 or -1, entered as 1 or 0, respectively.

`coll.c14scale`: flag for enabling or disabling the inclusion of a column of radiocarbon age locations; Dj; int; -1; 0 or -1, entered as 1 or 0 respectively.

`coll.coldep`: colour identifier for age-depth axis; Cf (main); long; 0; none.

`coll.colpl`: colour identifier for plots; Cd (main); long; 0; none.

`coll.coltit`: colour identifier for title; Ce (main); long; 0; none.

`coll.colts`: colour identifier for sediment column; Ch (main); long; 0; none.

`coll.colxax`: colour identifier for x-axes; Cj (main); long; 0; none.

`coll.colzone`: colour identifier for zone column; Ci (main); long; 0; none.

`coll.col14c`: colour identifier for 14C column; Cg (main); long; 0; none.

`coll.hist`: value for histogram widths for all samples; Dg; real; 0; 0-1000.

`coll.indscale`: flag to enable independent scaling; Ba; int; 0; 0 or -1, entered as 1 or 0, respectively.

`coll.indspl`: flag for independent splitting; Mh; int; 0; 0 or -1.

`coll.interp`: flag for interpolating samples; Bf; int; 0; 0 or -1.

`coll.ix10`: flag for \times 10 exaggeration; Dc; int; 0; 0 or -1, entered as 1 or 0, respectively.

`coll.join`: flag for join-up across missing samples; Df; int; 0; 0 or -1.

`coll.lval`: upper value for placing a solid dot (bullet) next to a curve; Di; real; 0.5; 0-10^6.

`coll.penwid`: width of pen lines; Dk; real; 0; 0-100.

`coll.rot`: angle of rotation for taxa names; De; int; 45; 0-90.

`coll.script`: flag for reading from a script file; P; int; 0; 0 or -1.

`coll.shade`: value set for the grey shade in filled plots; Dd, K; real; 0; 0-1.

`coll.skip`: value of n for omission of every nth sample; Be; int; 0; -100-100.

`coll.style`: diagram style; Db; int; 0; 0 or 1.

`dat.fname`: output file dataste; Eg; string; none; any text [must be legal on the operating system being used, but that is not checked by the menu].

`four.flag`: flag for Fourier analysis; Mf; int; 0; 0 or -1.

`four.tran`: method of transformation of dataset before Fourier analysis; Mf.1; int; 0; 0-4 (see menu).

`gl.adpage.minpl`: distance on y-axis of age-depth plot for unit time; Ab; real; 0.002; 0.0002-0.02 [appears as 0.1-10 on menu].

`gl.adres`: resolution of age-depth curve; Dl; int; 1000; 10-10000.

`gl.calval`: flag for using mode or weighted average with calibrated dates; Lk; int; -1; 0 or -1, entered as 1 or 0, respectively.

`gl.colts`: flag for colouring sediment units; Ck (main); int; -1; 0 or -1, entered as 1 or 0, respectively.

`gl.colzone`: flag for colouring zones; Cl (main); int; -1; 0 or -1, entered as 1 or 0, respectively.

`gl.font.code`: code for font to use for lettering; F; int; 'H'; any of A, B, C, E, H, N, P, or T (see menu).

`gl.font.ufac`: user-imposed size for font; G; real; 0; 0-36 (units of 1 / 72 inches).

`gl.mran`: method for obtaining random numbers; N3, Mc.7.2, Md.4.2; int; 1; 1-6 (see menu).

`gl.neglev`: threshold for excluding missing data; Bg; real; -0.99; any negative real number between 0 and -FLT_MAX.

`gl.page.minpl`: distance on y-axis for 1%; Aa; real; 0.002; 0.0002-0.02 [appears as 0.1-10 on menu].

`gl.page.prop`: height of diagram as a proportion of page height; Ad; real; 0.75; 1000 / `gl.page.high` to `gl.papht` / `gl.page.high`, where `gl.page.high` = `gl.page.wide` / square-root(2) [appears on menu as 10-297].

`gl.page.wide`: page width in units of 0.01mm; Ac; real; 24000; 1000-42000 [appears as 10-420 on menu].

`gl.papht`: paper height in units of 0.01mm; Ae; real; 29700; 29700, 42000, 27940 or 35560 [appears as A4, A3, Letter or Legal on menu].

`gl.papwid`: paper width in units of 0.01mm; Ae; real; 21000; 21000, 29700, 21590 or 21590 [appears as A4, A3, Letter or Legal on menu].

`gl.palette.back`: background colour identifier; Cb (both); long; 1; none.

`gl.palette.fore`: foreground colour identifier; Cc (main); long; 0; none.

`gl.recalc`: flag for whether datasets should be recalculated to proportions of the sum of the values included before analysis; Mj; int; -1; 0 or -1.

`gl.seed`: seed for random number generator; N4, Mc.7.3, Md.4.2; string; 'TIMER'; either 'TIMER' or a positive number of 10 digits or fewer, and less than 2147483647.

`gl.showmem`: flag for wrting details of memory allocation to logfile; Ud (main); int; -1; 0 or -1, entered as 1 or 0, respectively.

`gl.sum2col`: flag for printing sum values in a double column; Bd; int; 0; 0 or -1. `gl.tufte`: flag for switching to Tufte diagram style; Da; int; 0; 0 or -1.

`head.adxaxis[C14BP]`: label for x-axis of age-depth plot in radiocarbon years BP; Ij (but access depends on settings in menu L); string; 'Age ({14}C yr BP)'; any text up to 75 characters.

`head.adxaxis[CALBP]`: label for x-axis of age-depth plot in calendar years BP; Ij (but access depends on settings in menu L); string; 'Age (Cal. yr BP)'; any text up to 75 characters.

`head.adxaxis[C14AD}`: label for x-axis of age-depth plot in radiocarbon years BP; Ij (but access depends on settings in menu L); string; 'Age ({14}C yr AD/BC)'; any text up to 75 characters.

`head.adxaxis[CALAD]`: label for x-axis of age-depth plot in calendar years AD/BC; Ij (but access depends on settings in menu L); string; 'Age (Cal. yr AD/BC)'; any text up to 75 characters.

`head.adyaxis`: label for y-axis of age-depth plot; Ik; string; 'Depth (cm)'; any text up to 75 characters.

`head.base`: text at base of diagram; Ia; string; none; any text up to 75 characters.

`head.age[C14BP]`: heading for age column in radiocarbon years BP; If (but access depends on settings in menu L); string; '{14}C ages (yr BP)'; any text up to 75 characters.

`head.age[CALBP]`: heading for age column in calibrated years BP; If (but access depends on settings in menu L); string; 'Cal. ages (yr BP)'; any text up to 75 characters.

`head.age[C14AD]`: heading for age column in radiocarbon years AD/BC; If (but access depends on settings in menu L); string; '{14}C ages (yr AD/BC)'; any text up to 75 characters.

`head.age[CALAD]`: heading for age column in calibrated years AD/BC; If (but access depends on settings in menu L); string; 'Cal. ages (yr AD/BP)'; any text up to 75 characters.

`head.dendro`: heading for dendrogram produced by CONISS or CONIIC; Id; string; none; any text up to 75 characters.

`head.ps1[DEPTH]`: first line of heading for depth axis; Ig (access depends upon settings in menu L); string; 'Depth' if data are presented by depth, 'Sample' if data are presented as surface samples; any text up to 75 characters.

`head.ps1[C14BP]`: first line of heading for age axis (if present) for ages in radiocarbon years BP; Ig (access depends upon settings in menu L); string; 'Age'; any text up to 75 characters.

`head.ps1[CALBP]`: first line of heading for age axis (if present) for ages in calendar years BP; Ig (access depends upon settings in menu L); string; 'Age';

any text up to 75 characters.

`head.ps1[C14AD]`: first line of heading for age axis (if present) for ages in radiocarbon years AD/BC; Ig (access depends upon settings in menu L); string; 'Age'; any text up to 75 characters.

`head.ps1[CALAD]`: first line of heading for age axis (if present) for ages in calendar years Ad/BC; Ig (access depends upon settings in menu L); string; 'Age'; any text up to 75 characters.

`head.ps2`: second line of heading for depth axis; Ih; string; 'cm' or 'm', as appropriate, if data are presented by depth, none if data is presented as surface samples; any text up to 75 characters.

`head.ps1[C14BP]`: second line of heading for age axis (if present) for ages in radiocarbon years BP; Ih (access depends upon settings in menu L); string; 'yr BP'; any text up to 75 characters.

`head.ps1[CALBP]`: second line of heading for age axis (if present) for ages in calendar years BP; Ih (access depends upon settings in menu L); string; 'yr BP'; any text up to 75 characters.

`head.ps1[C14AD]`: second line of heading for age axis (if present) for ages in radiocarbon years AD/BC; Ih (access depends upon settings in menu L); string; 'yr AD/BP'; any text up to 75 characters.

`head.ps1[CALAD]`: second line of heading for age axis (if present) for ages in calendar years Ad/BC; Ih (access depends upon settings in menu L); string; 'yr BP'; any text up to 75 characters.

`head.sed`: heading for sediment column; Ie; string; none; any text up to 75 characters.

`head.time[C14BP]`: heading for timescale column in radiocarbon years BP; Ii (but access depends upon settings in menu L); string; 'Age ({14}C yr BP)'; any text up to 75 characters.

`head.time[CALBP]`: heading for timescale column in calendar years BP; Ii (but access depends upon settings in menu L); string; 'Age (Cal. yr BP)'; any text up to 75 characters.

`head.time[C14AD]`: heading for timescale column in radiocarbon years AD/BC; Ii (but access depends upon settings in menu L); string; 'Age ({14}C yr AD/BC)'; any text up to 75 characters.

`head.time[CALAD]`: heading for timescale column in calendar years AD/BC; Ii (but access depends upon settings in menu L); string; 'Age (Cal. yr AD/BC)'; any text up to 75 characters.

`head.subtitle`: subtitle text, printed on diagram immediately after main title, but with smaller lettering; Ib; string; none; any text up to 75 characters.

`head.zone`: heading for zonation column; Ic; string; none; any text up to 75 characters.

`inage.fname`: input file with radiocarbon (or other) age details; Ee; string; `cC14`, where `c` is the drive and path (if any) plus first four characters of the input filename; any text [must be legal on the operating system being used, but that is not checked by the menu].

`incal.fname`: input file with calibratedage details; Ef; string; `cCAL`, where `c` is the drive and path (if any) plus first four characters of the input filename; any text [must be legal on the operating system being used, but that is not checked by the menu].

`intfg`: flag for interactive plotting; K; int; 0; 0 or -1.

`ints.fname`: input file with sediment description details; Ef; string; `cTS`, where `c` is the drive and path (if any) plus first four characters of the input filename; any text [must be legal on the operating system being used, but that is not checked by the menu].

`ipform`: flag for format of dataset output file; Jb; int; 0; 0 or 1.

`ipdat`: flag for printing dataset to file; Ja; int; 0; 0 or -1, entered as 1 or 0, respectively.

`lev.adsc`: flag for turning off the default interval between major and minor marks on the age-depth scale; Bc; int; 0; 0 or -1.

`lev.interp`: interval between interpolated samples; Bf; real; 1; 1-10^7.

`lev.intstart`: level at which to start interpolating samples (highest level); Bf; real; -2000; 1-10^7.

`lev.intstop`: level at which to stop interpolating samples (lowest level); Bf; real; 0; -2000-10^7.

`lev.inttype`: method for interpolating samples; Bf; int; 0; 0 or 1.

`lev.majmark`: interval between major marks on age-depth axis; Bc; real; 0; 0-10^7.

`lev.minmark`: interval between minor marks on age-depth axis; Bc; real; 0; 0-10^7.

`lev.udata`: flag for whether maximum and minimum values for age or depth values have been given; H; int; 0; 0 or -1.

`lev.umax`: maximum value for age or depth to be plotted; H; real; -2000-10^7.

`lev.umin`: minimum value for age or depth to be plotted; H; real; -2000-10^7.

`num.stat`: flag for statistical description; Mg; int; 0; 0 or -1.

`num.thresh`: minimum value for the abundance of a taxon that must be reached for the inclusion of that taxon in most data analyses (negative to disable); Mi; real; 0.05% for data presented as proportions, or 0 for other data types; -1-1 for data presented as proportions, or -10,000,000-10,000,000 for

other data types.

`outmain.fname`: main PostScript output file; Ea; string; `c.PS`, where `c` is the input filename; any text [must be legal on the operating system being used, but that is not checked by the menu].

`pca.flag`: flag for principal components analysis; Md.0; int; 0; 0 or -1, entered as 1 or 0, respectively.

`pca.matrix`: selection of matrix for principal components analysis; Md.1; int; 3; 1-3 (see menu).

`pca.model.ci`: flag for modelling principal components analysis; Md.4.0; int; 0; 0 or -1, entered as 1 or 0, respectively.

`pca.model.n`: number of models to carry out for principal components analysis; Md.4.1; int; 10; 1-1000.

`pca.naxes`: number of principal axes for output from PCA; Md.2; int; 6; 1-(number of taxa).

`pca.tran`: method of transformation of dataset before PCA; Md.3; int; 0; 0-4 (see menu).

`rarer.flag`: flag for rarefaction analysis; Mb; int; 0; 0 or -1, entered as 1 or 0, respectively.

`rarer.n`: standardization count for rarefaction analysis; Mb; int; 0-1000.

`roc.diss`: selection of dissimilarity measure for rate-of-change calculations; Ma.1; int; 1; 1-3.

`roc.flag`: flag for rate-of-change calculations; Ma; int; 0 or -1, entered as 1 or 0, respectively.

`roc.model.ci`: flag for modelling rates of change; not yet implemented; int; 0 or -1, entered as 1 or 0, respectively.

`roc.model.n`: number of models to carry out for rates of change; not yet implemented; int; 10; 1-1000.

`roc.smooth`: flag for smoothing data before rate-of-change analyses; Ma.2; int; 0; 0, 3, 5, 7 or 9.

`stat.fname`: output file for data analyses; Ec; string; none; any text [must be legal on the operating system being used, but that is not checked by the menu].

`sunit[0]`: text for labelling pollen concentration (volume) data; Il; string; ‘ grains cm{-3}’; any text up to 39 characters.

`sunit[1]`: text for labelling pollen concentration (weight) data; Il; string; ‘ grains g{-1}’; any text up to 39 characters.

`sunit[2]`: text for labelling pollen accumulation rate data; Il; string; ‘ grains cm{-2} year{-1}’; any text up to 39 characters.

sunit[3]: text for labelling charcoal concentration (volume) data; Il; string; ‘ cm{2} cm{-3}’; any text up to 39 characters.

sunit[4]: text for labelling charcoal accumulation rate data; Il; string; ‘ cm{2} cm{-2} year{-1}’; any text up to 39 characters.

sunit[5]: text for labelling charcoal concentration (weight) data; Il; string; ‘ cm{2} g{-1}’; any text up to 39 characters.

sunit[6]: text for labelling charcoal:pollen quotient values; Il; string; ‘ cm{2} grain{-1}’; any text up to 39 characters.

sunit[7]: text for labelling rate of change data; Il; string; ‘ cent.{-1}’; any text up to 39 characters.

sunit[8]: text for labelling ratio data; Il; string; ‘ ratio’; any text up to 39 characters.

sunit[9]: text for labelling Fourier analysis output values; Il; string; ‘ power’; any text up to 39 characters.

sunit[10]: text for labelling dendrogram; Il; string; ‘ Total dispersion’; any text up to 39 characters.

zone.fname: input file with zone details; Ed and Mc.3; string; cZONE, where c is the drive and path (if any) plus first four characters of the input filename; any text [must be legal on the operating system being used, but that is not checked by the menu].

zoner.flag: flag for zonation; Mc; int; 0 or -1, entered as 1 or 0, respectively.

zoner.meth: method for carrying out zonation; Mc.2; int; 1-6 (see menu).

zoner.model.ci: flag for modelling zonation; Mc.7.0; int; 0 or -1, entered as 1 or 0, respectively.

zoner.model.n: number of models to carry out for zonation; Mc.7.1; int; 10; 1-1000.

zoner.n: number of zones to be obtained by zonation; Mc.1; int; 0-(number of samples - 2).

zoner.pre: prefix for zone labels; Mc.4; string; first character of dataset title followed by a hyphen; any characters, up to a maximum of 24.

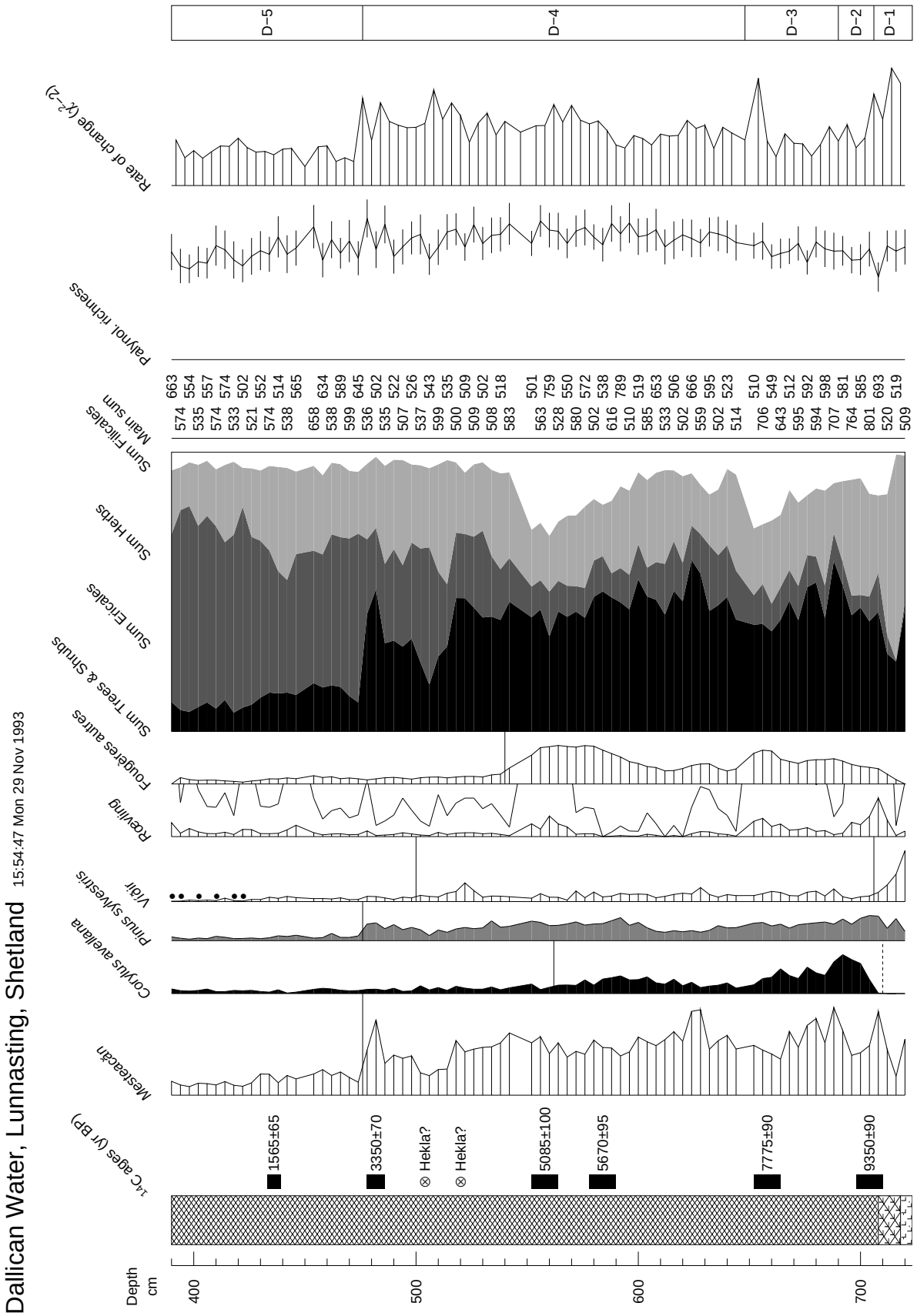
zoner.ransam: flag for shuffling samples before zonation; Mc.6; int; 0; 0 or -1.

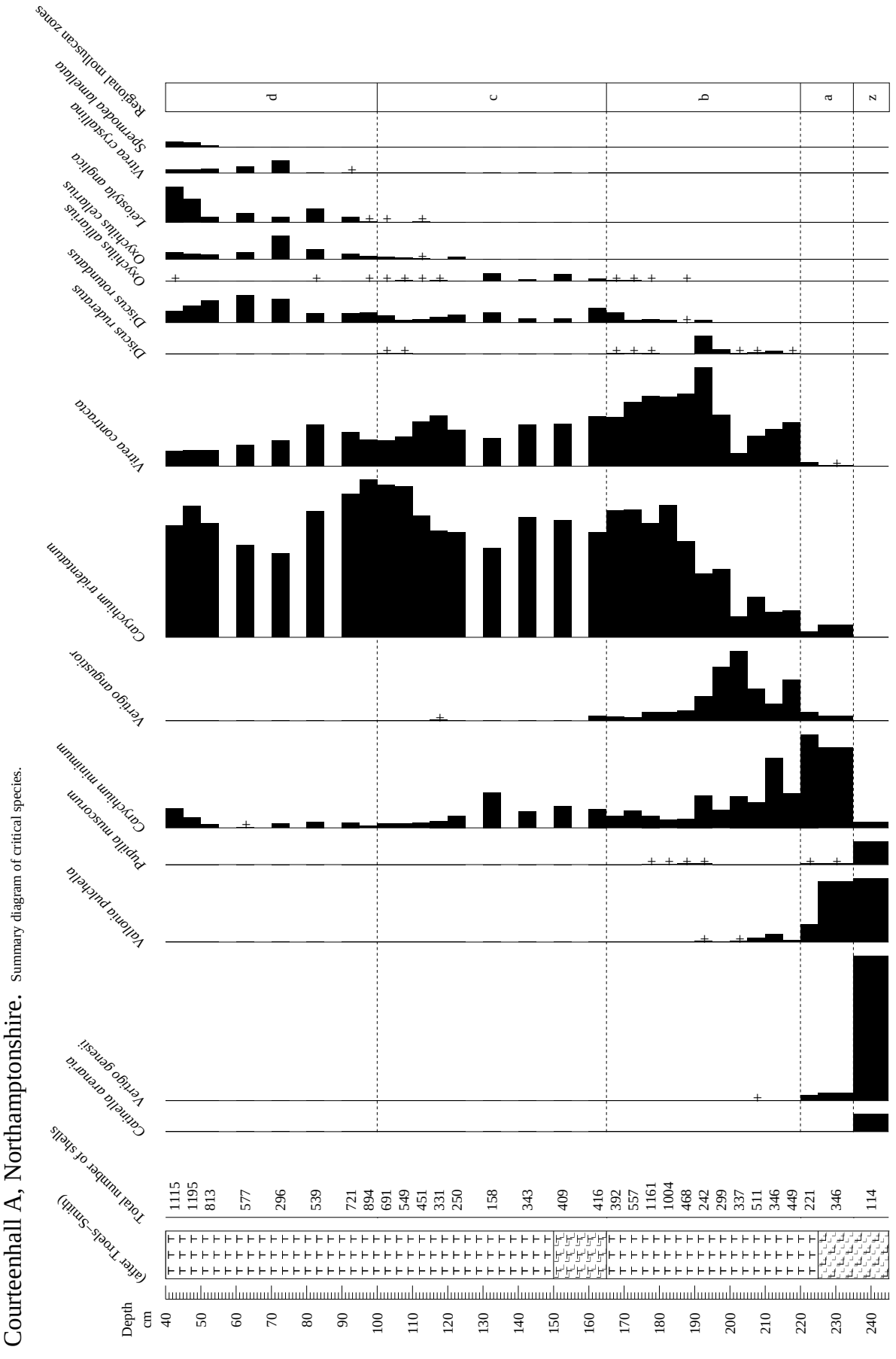
zoner.tran: method of transformation of dataset before zonation; Mc.5; int; 0; 0-4 (see menu).

14 Examples of psimpoll output

Two diagrams generated by **psimpoll** are included to show something of the range of output available. The program output is PostScript images. These have been converted to EPS with gsview, and are shown on the next two pages.

1. Pollen diagram from Dallican Water, Shetland
2. Molluscan diagram from Courteenhall A, Northamptonshire, produced by Rich Meyrick (e-mail: ram23@cus.cam.ac.uk).


Dallican Water, Lunnasting, Shetland 15:54:47 Mon 29 Nov 1993
Depth cm
14C ages (yr BP)
1565±65
3350±70
Hekla?
Hekla?
5085±100
5670±95
7775±90
9350±90
Mesleacán
Corylus avellana
Pinus sylvestris
Betula
Rosaceae undiff.
Fagus sylvatica
Quercus robur type
Ulmus
Tilia cordata
Alnus glutinosa type
Fraxinus excelsior
Sum Trees & Shrubs
Sum Ericales
Sum Herbs
Sum Filicales
Main sum
Palynol. richness
Rate of change (χ²-2)
D-5
D-4
D-3
D-2
D-1
Percentages of total terrestrial pollen and spores
E(T250)
cent.-1


Courteenhall A, Northamptonshire. Summary diagram of critical species.
Depth cm
(after Troels-Smith)
Total number of shells
Cantinella arenaria
Vertigo genesii
Valloniа pulchella
Pupilla muscorum
Carychium minimum
Vertigo angustior
Carychium tridentatum
Vitrea contracta
Discus rudderatus
Discus rotundatus
Oxychilus diliartus
Oxychilus cellarius
Leiosyra anglica
Vitrea crystallina
Spermodea lamellata
Regional molluscan zones
a
b
c
d
z
%
Values expressed as percentages of total land snails. (+ represents < 0.5%)

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

Bennett, K.D. (1990). Pollen counting on a pocket computer. *New Phytologist* **114**, 275–280.

Bennett, K.D. (1994). Confidence intervals for age estimates and deposition times in late-Quaternary sediment sequences. *The Holocene* **4**, 337–348.

Bennett, K.D. (1996). Determination of the number of zones in a biostratigraphical sequence. *New Phytologist* **132**, 155–170.

Bennett, K.D., Boreham, S., Sharp, M.J. & Switsur, V.R. (1992). Holocene history of environment, vegetation and human settlement on Catta Ness, Lunnasting, Shetland. *Journal of Ecology* **80**, 241–273.

Bennett, K.D. & Humphry, R.W. (1995). Analysis of late-glacial and Holocene rates of vegetational change at two sites in the British Isles. *Review of Palaeobotany and Palynology* **85**, 263–287.

Birks, H.J.B. & Berglund, B.E. (1979). Holocene pollen stratigraphy of southern Sweden: a reappraisal using numerical methods. *Boreas* **8**, 257–279.

Birks, H.J.B. & Gordon, A.D. (1985). *Numerical Methods in Quaternary Pollen Analysis*. London: Academic Press.

Birks, H.J.B. & Line, J.M. (1992). The use of rarefaction analysis for estimating palynological richness from Quaternary pollen-analytical data. *The Holocene* **2**, 1–10.

Birks, H.J.B., Line, J.M., Juggins, S., Stevenson, A.C. & Ter Braak, C.J.F. (1990). Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society of London Series B* **327**, 263–278.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829–836.

Cleveland, W.S. & Devlin, S.J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* **83**, 596–610.

Davis, O.K. (1993). Program sources. *INQUA Commission for the Study of the Holocene: Working Group on Data-handling Methods, Newsletter* **10**, 30–33.

Gauthier, T.D. (2001). Detecting Trends Using Spearman's Rank Correlation Coefficient. *Environmental Forensics* **2**, 359–362.

Grimm, E.C. (1987). CONISS: a FORTRAN 77 program for stratigraphically constrained cluster analysis by the method of incremental sum of squares. *Computers & Geosciences* **13**, 13–25.

Grimm, E.C. (1992). *TILIA*. Springfield, Illinois, USA: Illinois State Museum. Software.

Hill, M.O. (1979). DECORANA – a FORTRAN program for detrended correspondence analysis and reciprocal averaging. Technical report, Ecology and Systematics, Cornell University.

Jackson, D.A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* **74**, 2204–2214.

Jacobson, Jr, G.L. & Grimm, E.C. (1988). Synchrony of rapid change in late-glacial vegetation south of the Laurentide ice sheet. *Bulletin of the Buffalo Society of Natural Sciences* **33**, 31–38.

Jacobson, Jr, G.L., Webb, III, T. & Grimm, E.C. (1987). Patterns and rates of vegetation change during the deglaciation of eastern North America. In W.F. Ruddiman & H.E. Wright, Jr (Eds.), *North America and Adjacent Oceans During the Last Deglaciation*, The Geology of North America Volume K-3, pp. 277–288. Boulder, CO: Geological Society of America.

Legendre, L. & Legendre, P. (1983). *Numerical Ecology*. Amsterdam: Elsevier.

Leva, J.L. (1992). A fast normal random number generator. *ACM Transactions on Mathematical Software* **18**, 449–453.

Line, J.M., Ter Braak, C.J.F. & Birks, H.J.B. (1994). WACALIB version 3.3 — a computer program to reconstruct environmental variables from fossil assemblages by weighted averaging and to derive sample-specific errors of prediction. *Journal of Paleolimnology* **10**, 147–152.

Maher, Jr, L.J. (1972). Nomograms for computing 95% limits of pollen data. *Review of Palaeobotany and Palynology* **13**, 85–93.

Maher, Jr, L.J. (1981). Statistics for microfossil concentration measurements employing samples spiked with marker grains. *Review of Palaeobotany and Palynology* **32**, 153–191.

Marsaglia, G. & Zaman, A. (1994). Some portable very-long-period random number generators. *Computers in Physics* **8**, 117–121.

Newman, W.M. & Sproull, R.F. (1981). *Principles of Interactive Computer Graphics* (Second ed.). Singapore: McGraw-Hill.

Oksanen, J. & Minchin, P.R. (1997). Instability of ordination results under changes in input data order: explanations and remedies. *Journal of Vegetation Science* **8**, 447–454.

Parratt, L.G. (1961). *Probability and Experimental Errors in Science: an Elementary Survey*. New York: Wiley.

Prentice, I.C. (1980). Multidimensional scaling as a research tool in Quaternary palynology: a review of theory and methods. *Review of Palaeobotany and Palynology* **31**, 71–104.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (1992). *Numerical Recipes in C: the Art of Scientific Computing* (Second ed.). Cambridge: Cambridge University Press.

Squires, G.L. (1985). *Practical Physics* (Third ed.). Cambridge: Cambridge University Press.

Troels-Smith, J. (1955). Karakterisering af lose jordater. Characterization of unconsolidated sediments. *Danmarks Geologiske Undersøgelse Række IV* **3(10)**, 73 pp.

Walker, D. & Pittelkow, Y. (1981). Some applications of the independent treatment of taxa in pollen analysis. *Journal of Biogeography* **8**, 37–51.

Walker, D. & Wilson, S.R. (1978). A statistical alternative to the zoning of pollen diagrams. *Journal of Biogeography* **5**, 1–21.

Zaman, A. & Marsaglia, G. (1991). A new class of random number generators. *Annals of Applied Probability* **1**, 462–480.