Outline of the new palaeo plotting program P4

Antony Blundell

School of Geography, University of Leeds, Leeds, UK

SUMMARY

This short communication introduces P4, a new open source palaeo plotting program that offers a quick and intuitive method for producing publication-quality plots of data from palaeo archives. Capabilities of the program are described together with some example plots of macrofossil, testate amoebae and pollen data. Strengths of the program and areas for future development are highlighted.

KEY WORDS: General Public License v3, palaeo archive, Python, visual presentation, Zenodo

INTRODUCTION

Palaeo-environmental and palaeo-climate studies prominently feature plots of palaeo analyses versus depth to visualise data for readers. Over the past 40 years, researchers have successfully employed specialised packages to create these plots. Examples include Tilia (https://www.tiliait.com/), Psimpol (https://chrono.qub.ac.uk/psimpoll/psimpoll.html), POLPAL (Nalepka & Walanus 2003), C2 (https://www.staff.ncl.ac.uk/stephen.juggins/software/C2Home.htm) and Rioja (Juggins 2022). These packages have varying degrees of usability, functionality, complexity and accessibility; and encompass command line and graphical user interface (GUI) formats.

This short communication introduces a new plotting focused program, namely P4 (Python, Palaeo, Plotting, Program) (Blundell 2023), which is a free and open-source program written in Python 3 (3.7.6). It uses a number of Python data science packages such as pandas 1.3.0 (The pandas development team 2021), matplotlib 3.2.2 (Caswell *et al*. 2020) and NumPy 1.18.1 (Harris *et al*. 2020). Unlike many of the other applications mentioned here P4 is focused solely on plotting of palaeo data in the traditional taxa versus depth or age style, be it for records from peat, lakes or potentially any palaeo archive. P4 does not include statistical analyses. The primary objective is to produce publication-quality plots in a quick and intuitive way. P4 is free to download, use and modify (GPL v3.0), from zenodo (Blundell 2023).

GENERAL OUTLINE OF P4

Full details of all aspects of P4 are not provided in this article as the program code and comprehensive user manual, together with example data files and resultant outputs for peat palaeo-archives from Keighley Moor (Blundell & Holden 2015) and Ardkill Moss (Blundell *et al*. 2008), are provided within the package (Blundell 2023) and are referred to here. The user manual covers program installation, setup and all aspects of data input and options for plot styles, aesthetics and functionality covering all of P4's present capabilities. The program has been tested primarily in Windows 7, 10 and 11 operating systems but successful installations and operation on Linux and MacOS systems have been carried out and the required steps are documented in the user manual. P4 is in continual development and, therefore, the user should download the latest version listed. Here the potential of the P4 program is illustrated using previously published macrofossil data from Keighley Moor, testate amoebae data from Ardkill Moss and pollen data from Carlshof forest hollow (Bradley *et al*. 2013). The plots contain supplementary data outside of the main proxies, such as charcoal, and transfer function estimations to demonstrate how P4 deals with data that may have different plotting requirements. A 'test' plot of completely fictional data is also displayed to demonstrate much of the aesthetic variation and functionality that is possible within a single diagram.

P4 employs two or three csv files, an input file (and potentially an extra input file) and a parameter file (the user can modify the names of these files). The files incorporate (a) palaeo data and (b) all the available options for plot type and aesthetics (Figure 1). Template input and parameter files are provided (Blundell 2023). The input file is laid out so that it contains the depth or age data and the palaeo data, but it also contains coding in the uppermost 39 rows for taxa-specific plotting options. The parameter file largely covers non-taxa-specific plotting options and is laid out in 15 logical groupings. Instructions and notes for potential entries are repeated in both the input and parameter files and the user manual.

After palaeo data have been added and plot options edited within the required csv files, the program is run from the command line (see user manual) although experienced Python users can run it within any integrated development environment (IDE). Required user participation within the command line environment is minimal and should not be off-putting for users. Having navigated to the directory where the P4 program has been saved, a statement is required containing four pieces of text that provide (a) the word 'python', (b) the full name of the P4 program, (c) the location of the parameter file (preceded by '--input') and (d) the parameter file name. Full details of the command line statement are provided in the manual, which should be consulted.

Upon running the program, a figure is produced in svg, pdf or png format in the location designated by the user via the parameter file. Figure rotation (figures are produced in portrait so should be rotated) and inspection is followed by further adjustment of options within the parameter/input files until the desired figure is produced. Relatively limited experience is required to substantially reduce the number of iterations needed before figures are completed to the user's satisfaction (Figure 1). A single iteration of P4 on the Keighley Moor example takes between 5 and 20 seconds from program start to figure production, depending on computer specification. Individual iteration time is a few seconds longer if a high dpi (dots per inch) count is selected for png production.

The program offers some further functionality such as: (a) drawing and labelling of zones; the addition of (b) radiocarbon date labels; (c) timelines; (d) taxa grouping labels; (e) taxa exaggerations; and (f) individual plots with multiple or stacked series. P4 may be described as 'simple' because it intentionally does not perform any statistical analyses. Statistical analyses can be achieved through other Python or R-based programming packages. Palaeo data plotting programs also often offer the ability to add extras such as stratigraphy logs, but this is something the author has not found particularly useful, especially when Troels Smith style symbols have been employed. Upon publication the figure produced by the journal is often smaller than ideal and the symbology used can be hard to differentiate, especially in greyscale. Therefore, employing a dedicated open access drawing package either to amend a P4 program output or to display stratigraphic logs independently of the palaeo-archive figure is often quicker and more satisfactory (e.g., Blundell & Holden 2015). A graphical user interface (GUI) has purposely been omitted to avoid a myriad of mouse clicks and options hidden in a hierarchy of menus. In P4 all the options available are easily accessed in the parameter and input files.

Substantial effort has been made to provide useful error messages and guiding instructions so users can rectify common input or parameter mistakes. Invalid entries to the parameter or input files are common errors and may include, for example, offering a string entry when numerical data are required. For the user to take advantage of the available features, it is strongly recommended that the program manual is read thoroughly before the program is used.

Despite extensive testing, bugs may become apparent (as with any program). However, as the program code is freely available, Python users can, under the terms of the licence, edit the code to fix bugs and improve the current version, tailoring it to their needs. The author envisages further development to aid usability and to add useful features as and when required, providing they do not detract from the aim of producing a 'relatively simple' program and a relatively quick program execution time.

EXAMPLE FIGURES

The manual explains all aspects of the P4 program. Here three figures representing a mix of proxy data types (macrofossil, testate amoebae and pollen) from previously published work are used to demonstrate some of the program's functionality. The aim of these examples is to centre on the figures themselves and not the significance of the palaeo data, although site information is included in the Appendix and can be gained from the relevant cited papers. A fourth figure containing fictional data has been constructed with the aim of showing as much of the functionality and aesthetic options available as is feasible in one figure.

Keighley Moor macrofossil record

Data plotted in Figure 2 originate from an area of upland blanket bog in the English Pennine Hills

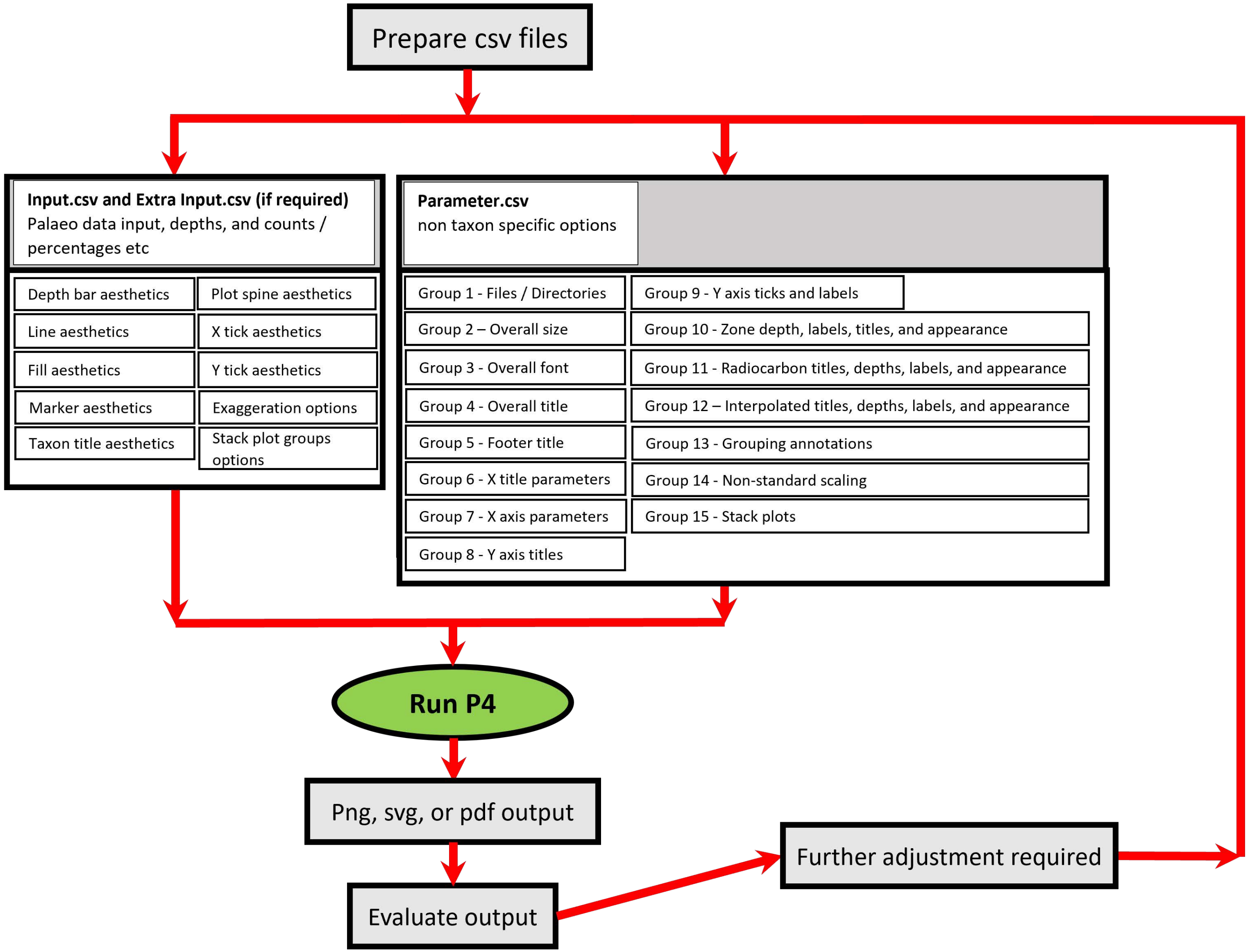


Figure 1. Flow diagram displaying the process of creating a plot in P4. The palaeo data, depths and taxa-specific plot options are provided in the input file (and extra input file if multiple series are required on a single plot), and preferences are specified via the 15 groups of non-taxa-specific options in the parameter file. This is followed by running the program and figure inspection. Adjustment of options in the input and parameter files is continued until the figure is to user satisfaction.

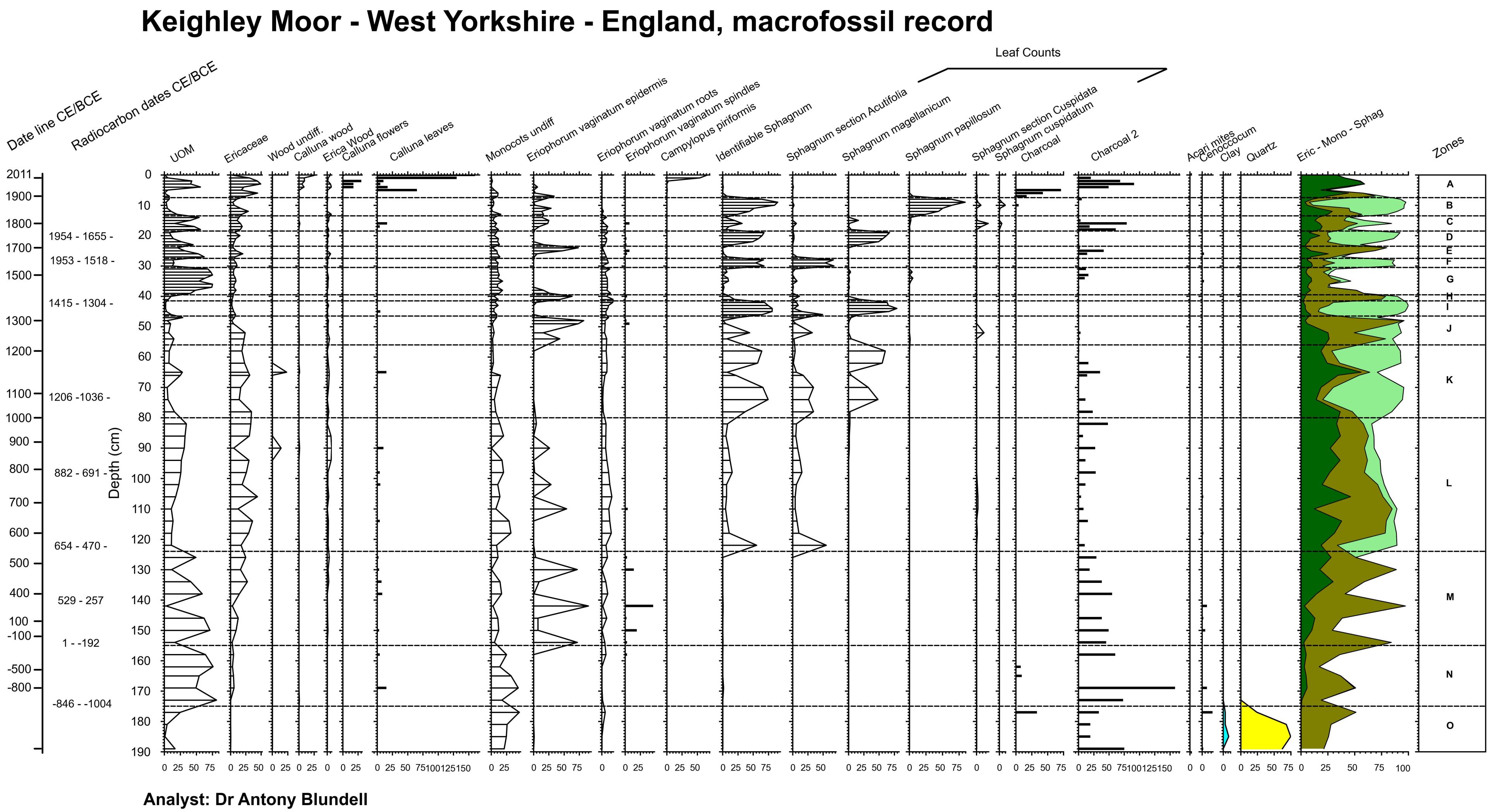


Figure 2. This figure illustrates the drawing of line with depth bars, bar, shadow and stack graph types. The zone, timeline, RC date display and grouping label functions are demonstrated. This is a macrofossil record from Keighley Moor, England; see Blundell & Holden (2015) for full information regarding the actual data presented. Compared to the presentation of these data in Blundell & Holden (2015), some colour has been added to highlight aesthetic capabilities in P4.

(Blundell & Holden 2015). Figure 2 demonstrates the program's ability to provide a range of plot types, utilising three of the seven (graph types 1–3 in the manual) available including line with depth bars, bar and shadow plots. Although Figure 2 is largely black and white the clay, quartz and stack plots make use of some of the 23 colours currently available for component parts, a number that can be expanded extensively with adaptation of the program code to include any available colours within the matplotlib module. Addition of radiocarbon dates, production of a timeline, a grouping label ('Leaf counts' here; it is possible to add up to ten grouping labels at present) and zone display functionality are also shown. The final plot in the figure is an example of the stack plot function which allows the user to amalgamate data into groups. In this case, taxa within the *Ericaceae* family, Monocots, and *Sphagnum* genus have been grouped. These can be displayed simply as abundance values added together, as shown, or as a percentage of all the groups nominated. Groupings are defined in the input file and other remaining options selected within the parameter file. The full range of possible figure titles is exhibited including main, x and y axis titles and a footer. Figure 2 also demonstrates the ability to have major and minor ticks at both ends of the plot and on each plot's vertical axis.

Ardkill Moss testate amoebae record

The data plotted in Figure 3 originate from Ardkill Moss, a raised bog in County Kildare, Ireland (Blundell *et al*. 2008). Figure 3 demonstrates the program's ability to display a different proxy and any specific additions that proxy may require. The testate plots are percentage based but the reconstructed water table element (Mean Water Table Depth (mm)) derived via a transfer function (Charman *et al*. 2007) requires different scaling as it has a range of -250 to 50 mm. Here the user can employ the 'non-standard scaling' (NSC) function of the program (options in the parameter file, see Figure 1) to resize plots with unusual scales compared to the main palaeo data displayed. At present, P4 allows five such plots and the width can be altered using a factor based on the initial size produced after the first iteration. The Mean Water Table Depth (mm) plot also exhibits the functionality of adding more than one series on a single plot by using a third optional csv file, here called the 'extra input.csv' file (Figure 1). This option is available only with the plot styles line, line and marker and scatter (graph types 3, 5 and 6 in the manual) because they provide good visual clarity. The dotted line used demonstrates one of the four line styles presently available.

The NSC function has also been applied to the 'No. of testates per 1cm*1000' plot displayed as a brown shadow plot which has a degree of transparency to the shading (this feature can be adjusted in the input csv file). A further function applied compared to those illustrated in Figure 2 is taxa exaggeration. In this case, selected taxa with low percentage values have an exaggerated extra series placed on the same plot using green shading. Exaggeration (multiplication of 5 here) can be applied to all, none, or any of the taxa plotted. Exaggerations can only be applied to plots using the graph types line with depth bars, line plot or shadow plot (graph types 2, 3 and 4 in the manual), and at present the exaggeration component can only be plotted using line plot or shadow plot (graph types 3 and 4 in the manual). These limitations are purely to aid clarity as other plot types create poor visualisations.

Carlshof forest hollow pollen record

The pollen percentage and charcoal data plotted in Figure 4 are from Carlshof small forest hollow in the Peutscher Forest, Germany (Bradley *et al*. 2013). This example is employed to show another type of proxy and the associated plotting requirements that P4 can accommodate. Bradley *et al*. (2013) plot taxa against age and, although that is easily achieved in P4, here the data are plotted against depth and a timeline is included so age can be tracked. Although the pollen record comprises 40 taxa in total it is dominated by eight taxa with the rest rarely contributing more than 10 % at any depth. This makes the exaggeration function particularly pertinent to making the low abundance taxa visible. Exaggeration can be applied to any number of taxa or to none, and here 20 times exaggeration has been applied to all taxa. Unexaggerated data are depicted in solid black and those with exaggeration applied in light grey shading with a degree of transparency (0.2 or 20 %) and a black boundary line. All taxa except *Juniperus*, which has a maximum value of 0.15 % at 648 cm, are clearly visible. The total pollen sum plot again demonstrates the use of the NSC function in P4, as noted for previous plots, and is displayed here as a blue dashed line which is another of the four line styles presently available. Charcoal was measured only below 101 cm and at 1 cm intervals, which conflicts with the sampling interval of the main pollen plots. As P4 uses the same depth or age master column for all plots within the figure, this can be worked around by employing the extra input file. Here a dummy column with an exceptionally low value can be applied in the main input file to effectively provide an empty plot. The data with different depth sampling frequencies can then be

Figure 3. Testate amoebae diagram for Ardkill Moss, Ireland. Further to Figure 1, this figure demonstrates the non-standard scaling function used for plotting Mean Water Table Depth and No. of testates per 1cm^3 * 1000. The former plot displays multiple series by employing the extra input file and uses three colours and two line styles. Exaggerations (multiplication of 5) are employed for selected taxa with low abundance (shaded green). See Blundell *et al*. (2008) for full figure captions regarding the actual data presented. Some colours have been altered compared to those in Blundell *et al*. (2008), to highlight aesthetic capabilities in P4.

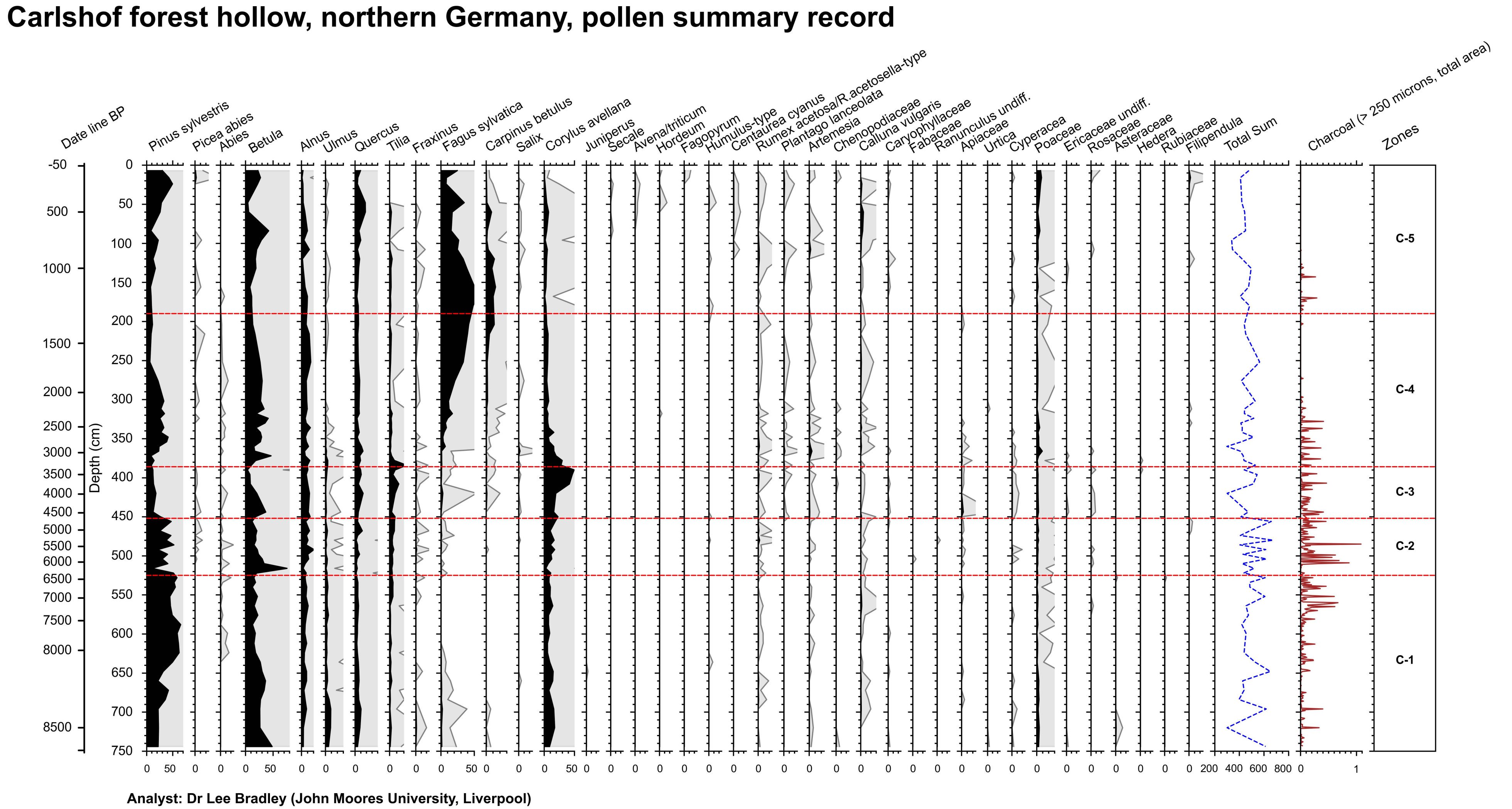


Figure 4. Summary percentage pollen diagram for Carlshof forest hollow, Germany (Bradley *et al*. 2013). For the > 250 µm charcoal record, no samples were analysed between 900 cal. yr BP and present. All taxa have had the P4 exaggeration function (multiplication of 20) applied and are shown as grey shading with black outlines. Non-exaggerated percentages have solid black shading. The charcoal record, sampled at a different frequency to the main pollen data, has been displayed by using the extra input file (see manual for full details of use).

added to the extra input file with the same column name as that in the main input file. This achieves the output in Figure 4 and at present data plotted using the extra input file are limited to the plot styles line, line and marker and scatter (graph types 3, 5 and 6 in the manual). The requirement to plot data with different sampling frequencies is relatively rare in these types of palaeo archive plots, but two different sampling frequencies can be accommodated here by using the extra input file. This could be expanded if required, but such a scenario has not presented itself to the author or program testers so far. These instances will normally apply to supplementary data rather than the main pollen, testate or macrofossil plot, for example charcoal (in this instance) or perhaps magnetic susceptibility or loss on ignition, for which it is less time consuming to achieve high sampling frequencies.

Fictional data plot

An attempt has been made in Figure 5 to display as much as possible of the program's current aesthetic and plotting potential using a fictional dataset. The figure produced has a strange appearance but allows the user to see a range of the plots and features that can be produced. All seven available graph types have been used, one in each of the labelled groups (group numbers reflect the graph type numbers in the user manual), the final one being the stack plot. Further additions include a multiple series plot, exaggeration to the first plot, six of the possible ten grouping labels, fictitious radiocarbon dates, a timeline, and a range of zones, zone lines and zone labels. Features are displayed in various indiscriminate aesthetic combinations with changes in colour, line styles, line widths, transparencies, font sizes and font weights.

STRENGTHS, LIMITATIONS AND AREAS FOR DEVELOPMENT

Strengths

- P4 is available to download for free from zenodo (Blundell 2023). Program development will continue and any existing bugs will be fixed when possible. The program code is available to be improved or added to by users via GitHub within the terms of the General Public Licence v3.0.
- A user manual and example data from two published papers are supplied with accompanying files to make it easy for the user to plot either example and alter aesthetics and plot options, and thus quickly get accustomed to P4.
- With experience, the lack of a GUI makes options easy to find and quick to change and transfer from one project to another simply by using copy-and-paste.
- The rate of figure production. Each of the figures displayed here took the author less than an hour to produce. This includes adding data to the input and parameter files, successive iterations of running the program, option adjustments and figure inspection as required.
- The outputs are of high visual quality and are available at present as pdf, svg or png formats with the option to specify dpi for pngs.

Hard and soft limitations

There are some limitations and some areas for development in the program, the latter of which can be considered in future program versions.

- The figures produced are limited to 60 plots per figure as the program employs the GridSpec module in Matplotlib. Although a hard limitation, this will rarely be an issue. If more than 60 plots are required, it is advantageous visually to split the plot into two or more parts.

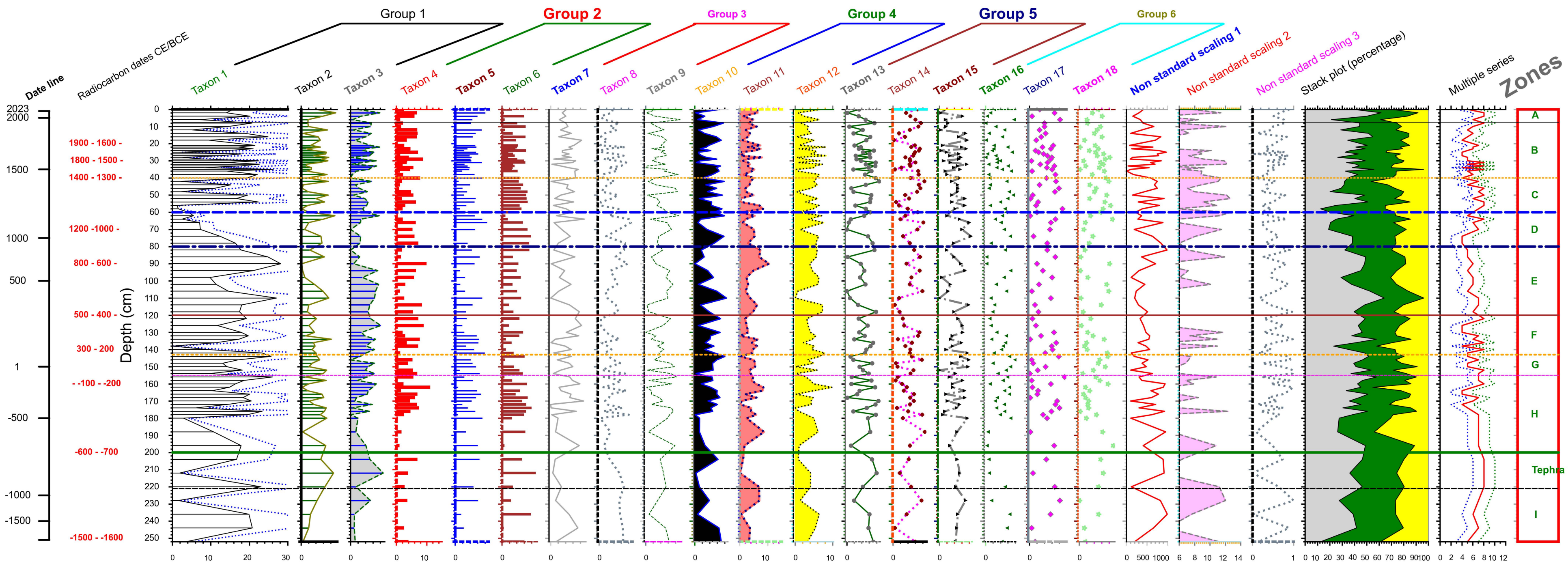
At present there are soft limits on some of the functionality of P4, but they are sensible limits in the author's experience. They include the following:

- Four available typefaces, four line styles, 23 colours and seven plot styles.
- A total of ten grouping labels can be added and the total number of stack plots per figure is two at present.
- Data are expected to be primarily of the same sampling frequency, although additional data at different frequency can be accommodated.
- These soft limits can be expanded relatively easily with additional code but at present are those that the author's experience, together with trials by other users testing the program with their data, have found acceptable.

Areas for development and perceived weaknesses

- At present, the procedure for placement of grouping labels takes a bit of practice and several iterations. As with many parts of the program, placements are carried out in data units and require an initial attempt and some subsequent fine tuning, as for plotting in many programs.

Illustrative figure showing functionality of P4.



Analyst: Dr Antony Blundell

Figure 5. Attempt, using a fictional dataset, to demonstrate much of the aesthetic and plotting potential the program has at present.

- Without a GUI the program is not immediately intuitive so the user **must** invest a small amount of time in reading the manual and trying at least one of the two examples provided. Command line usage is limited to a single statement with four pieces of text. Users should not be deterred as this is a simple procedure which, when combined with use of the input and parameter csv files, results in rapid figure production in the author's experience.
- Conflict has arisen in using the program for countries using commas as opposed to points as decimal separators. This can be partly addressed in the preparation of the input and parameter files before P4 is run, but future development will aim to address this within the program code itself. The freely available code means experienced Python users other than the author can navigate such issues and contribute to the solutions.
- The program was not written by a professional programmer. Comments have been used extensively and the PEP8 Python coding style (https://peps.python.org/pep-0008/) has largely been adhered to.
- The program is written in an imperative not procedural or object-oriented paradigm with tasks carried out in order as the program runs. The code could be reduced in size using a more procedural or object-oriented approach, but the run time is rapid (seconds rather than minutes) nonetheless. However, over time some of the repeated code may be assimilated into functions to aid future development.
- Options for drawing informative annotations including labels or symbols such as arrows anywhere on the plot is the next function being considered.

ACKNOWLEDGEMENTS

For program testing and proofreading, thanks go to Dr Dylan Young (University of Leeds) and Dr Lee Bradley (John Moores University, Liverpool). Thanks also to Dr Lee Bradley for providing pollen data used in Figure 4, Richard Rigby (Software Development Scientist, University of Leeds) for helping to test the program in Linux and MacOS systems, Professor Joseph Holden (University of Leeds) for help with proofreading, the editor Dr Olivia Bragg who has been very helpful regarding the niche nature of this submission, and Professor Andy Baird (University of Leeds) for the pivotal encouragement to pursue this endeavour.

REFERENCES

Blundell, A. (2002) *Late-Holocene Multi-Proxy Climate Records from Northern Britain and Ireland Derived from Raised Peat Stratigraphy*. PhD thesis, University of Southampton, Southampton, UK, 493 pp.

Blundell, A. (2023) ANTB25/P4-PYTHON-PALAEO-PLOTTING-PROGRAM: P4 release 1.0.3 (v1.0.2). Zenodo. https://doi.org/10.5281/zenodo.7544255

Blundell, A., Holden, J. (2015) Using palaeoecology to support blanket peatland management. *Ecological Indicators*, 49, 110–120.

Blundell, A., Charman, D.J., Barber, K. (2008) Multiproxy late Holocene peat records from Ireland: towards a regional palaeoclimate curve. *Journal of Quaternary Science*, 23, 59–71.

Bradley, L.R., Gieseke, T., Halsall, K., Bradshaw, R.H.W. (2013) Exploring the requirement for anthropogenic disturbance to assist the stand-scale expansion of *Fagus sylvatica* L. outside southern Scandinavia. *The Holocene*, 23, 579–586.

Caswell, T., Droettboom, M., Lee, A., Hunter, J. and 24 others (2020) matplotlib/matplotlib: REL: v3.2.2 (v3.2.2). Zenodo. https://doi.org/10.5281/zenodo.3898017

Charman, D.J., Blundell, A., ACCROTELM members (2007) A new European testate amoebae transfer function for palaeohydrological reconstruction on ombrotrophic peatlands. *Journal of Quaternary Science*, 22, 209–221.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R. and 22 others. (2020) Array programming with NumPy. *Nature*, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Juggins, S. (2022) *rioja: Analysis of Quaternary Science Data*. R package version 1.0-5, https://cran.r-project.org/package=rioja.

Nalepka, D., Walanus, A. (2003) Data processing in pollen analysis. *Acta Palaeobotanica*, 43, 125–134.

The pandas development team (2021) pandas-dev/pandas: Pandas 1.3.0 (v1.3.0). Zenodo. https://doi.org/10.5281/zenodo.5060318

Submitted 17 Jan 2023, revision 03 May 2023
Editor: Olivia Bragg

Author for correspondence: Dr Antony Blundell, School of Geography, University of Leeds, Leeds, Yorkshire LS2 9JT, UK. Tel: +44(0)113 343 1593; Email: a.blundell@leeds.ac.uk

Appendix: Brief descriptions of sites

KEIGHLEY MOOR

Keighley Moor Reservoir catchment (KMRC) is situated 3.5 km west of Oakworth in northern England (53º 85' 31" N, -02º 02' 13" E) and has an area of 1.48 km^2. The condition of blanket bog within this catchment influences the quality of water abstracted from the reservoir, and thus the cost of treatment by the responsible water authority (Yorkshire Water) to achieve the standard required for drinking water. Changes in particulate and dissolved organic carbon (POC and DOC) are of particular concern, as are heavy metals. The site has been managed for grouse shooting since the 1870s, has had some minor artificial drainage, and is grazed by sheep. It exhibits a typical surface patchwork due to managed burning and, since 2018, mowing. Wildfires, including a substantial event in 1918 and another during the 1940s, have resulted in some localised 'whaleback' and pedestal erosion formations. A comprehensive field survey of peat depth and stratigraphy was combined with construction of a detailed palaeoecological record to demonstrate the value of these tools in informing the development of improved land management practice and restoration targets. The detailed study of stratigraphy and palaeoecology (Blundell & Holden 2015) highlighted the atypical present-day state of the peatland in terms of vegetation, burning regime and peat composition, in the context of the last 6000 years. Figure 2 provides evidence of the changing composition of vegetation throughout the peatland's history, the role that *Sphagnum* has played, and the effect of burning during the last 200 years in creating the present-day vegetation which is relatively *Sphagnum* free and dominated by ericaceous plants.

ARDKILL MOSS

Ardkill Moss (53º 21' 52" N, 6º 57' 23" W) is a 48-hectare remnant of the Carbury raised bog system adjacent to the River Boyne in Ireland. Like most Irish raised bogs, the site has been subjected to peat extraction, although the core represented in Figure 3 was taken from its near-intact southern part. This is one of five cores taken from raised bogs in north–south and east–west transects across the British Isles as part of the author's PhD research to reconstruct climatic and environmental changes over the last 2000 years (Blundell 2002). Ardkill Moss was paired with another Irish site (Cloonoolish Bog, Galway) to form a Southern Ireland bog surface wetness record using multiple proxies, which evidenced clear changes to a wetter climate from ca. cal. CE 30 (1920 BP) through CE 310 (1640 BP) after a phase of drier and/or warmer climate in the 'Roman Period', CE 805 (1145 BP), CE 1040 (910 BP) and ca. CE 1300 (650 BP) after a drier/warmer period in early Mediaeval times (Blundell *et al*. 2008). In the last few centuries, Ardkill Moss has experienced substantial anthropogenic disturbance with accidental fires, reflected in the testate record by elevated *Hyalosphenia subflava* within zone ARD-g (Figure 3).

CARLSHOF FOREST HOLLOW

Carlshof small forest hollow (53º 25′ N, 13º 04′ E) is located in the Peutscher Forest, north of the town of Neustrelitz in Mecklenburg-Vorpommern, Germany. The work of Bradley *et al*. (2013) focuses on stand-scale expansion of *Fagus sylvatica* L. during the mid-Holocene epoch. The pollen record (Figure 4) shows that disturbance is key to expansion of this species in the forest surrounding the hollow. The record suggests that climate and fire frequency are also important factors in *F. sylvatica* expansion at stand scale.