# straditize: Digitizing stratigraphic diagrams

**Philipp S. Sommer**[1]**, Dilan Rech**[1]**, Manuel Chevalier**[1]**, and Basil A. S. Davis**[1]

**1** Institute of Earth Surface Dynamics, University of Lausanne, Géopolis, 1015 Lausanne, Switzerland

## Summary

In an age of digital data analysis, gaining access to data from the pre-digital era – or any data that is only available as a figure on a page – remains a problem and an underutilized scientific resource. Whilst there are numerous programs available that allow the digitization of scientific data in a simple x-y graph format, we know of no semi-automated program that can deal with data plotted with multiple horizontal axes that share the same vertical axis, such as pollen diagrams (see image below) and other stratigraphic figures that are common in the Earth sciences. Straditize (Stratigraphic Diagram Digitizer) (Sommer, 2019) fills this gap. It is an open-source program that allows stratigraphic figures to be digitized in a single semi-automated operation. It is designed to detect multiple plots of variables analyzed along the same vertical axis, whether this is a sediment core or any similar depth/time series.

The program supports mixtures of many different diagram types, such as bar plots, line plots, as well as shaded, stacked, and filled area plots. Other features of straditize include text recognition to interpret the names of the different plotted variables, the automatic and semi-automatic recognition of picture artifacts, as well an automatic measurement finder to exactly reproduce the data that has been used to create the diagram.

Straditize is written in the programming language Python and is available for Windows, Linux and MacOS. Being implemented in the visualization framework psyplot (Sommer,
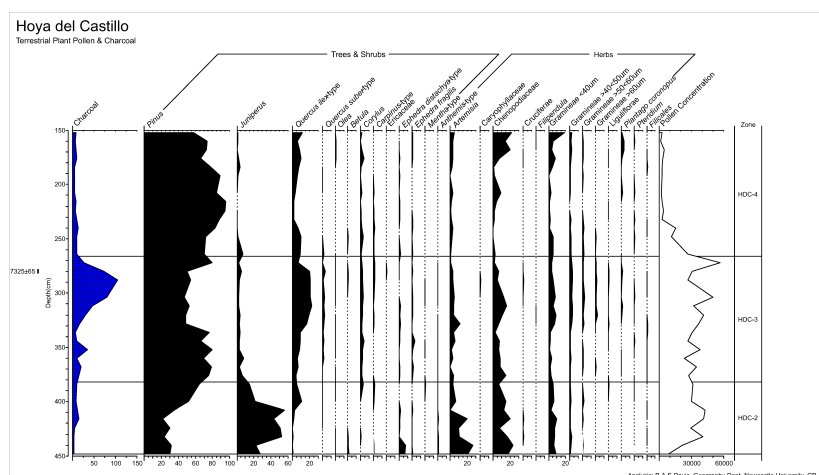
**Figure 1:** A standard pollen diagram from Hoya del Castillo with data from Davis & Stevenson (2007)

2017) it provides an extensively documented graphical user interface for a point-and-click handling of the semi-automatic process, but can also be scripted or used from the command line. The visualization is based on matplotlib (Hunter, 2007) and most of the detection algorithms use image recognition functions from the scikit-image package (Walt et al., 2014) and numeric routines from scipy (Jones, Oliphant, Peterson, & others, 2001) and numpy (T. E. Oliphant, 2006).

# Acknowledgements

# References

Davis, B. A. S., & Stevenson, A. C. (2007). The 8.2ka event and early–mid holocene forests, fires and flooding in the central ebro desert, ne spain. *Quat. Sci. Rev.*, *26*(13-14), 1695–1712. Journal Article. doi:10.1016/j.quascirev.2007.04.007

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science Engineering*, *9*(3), 90–95. doi:10.1109/MCSE.2007.55

Jones, E., Oliphant, T., Peterson, P., & others. (2001). SciPy: Open source scientific tools for Python. Retrieved from http://www.scipy.org/

Oliphant, T. E. (2006). *A guide to numpy* (Vol. 1). Trelgol Publishing USA. Retrieved from http://www.numpy.org/

Sommer, P. S. (2017). The psyplot interactive visualization framework. *The Journal of Open Source Software*, *2*(16). doi:10.21105/joss.00363

Sommer, P. S. (2019). Straditize: Python package for digitizing pollen diagrams. Retrieved January 15, 2019, from https://github.com/Chilipp/straditize

Walt, S. van der, Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., et al. (2014). Scikit-image: Image processing in Python. *PeerJ*, *2*, e453. doi:10.7717/peerj.453