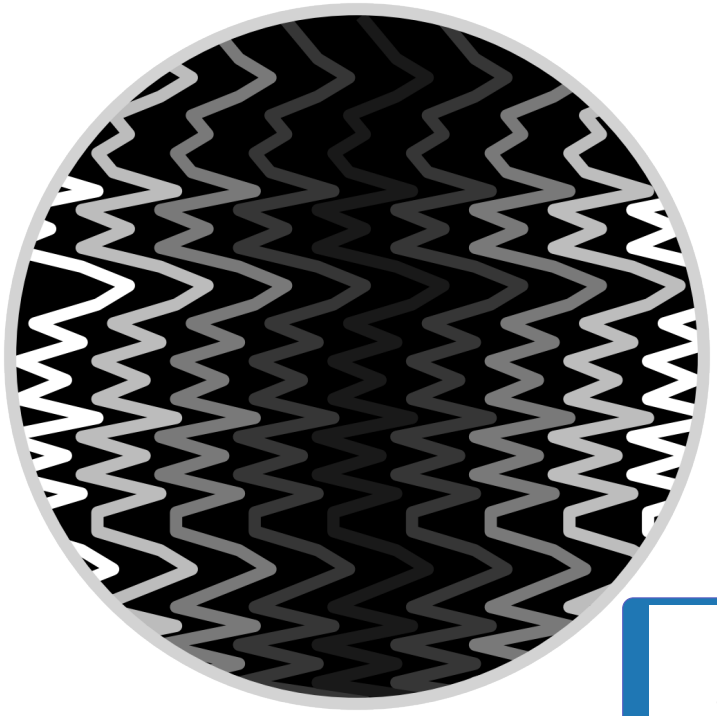
STRADITIZE

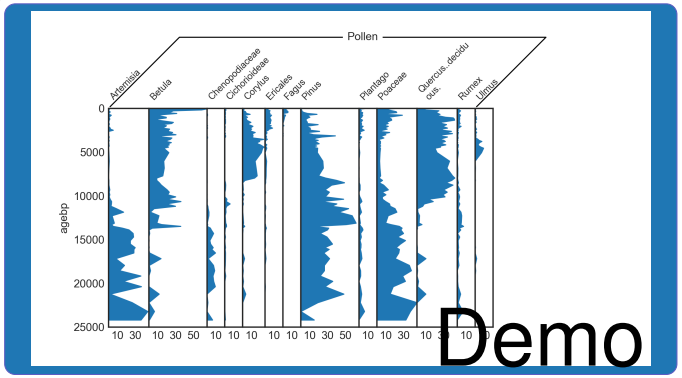
Digitizing pollen diagrams and more

Help



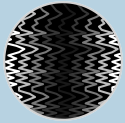
Motivation

Installation



Demo

Author

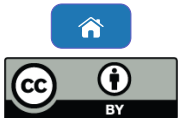


How to navigate

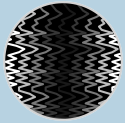
UNIL | Université de Lausanne

This presentation has been prepared for a PICO presentation at the EGU 2018 in Vienna, Austria. To facilitate the navigation, a lot of hyperlinks are used. Almost every item in this presentation is clickable:

- click buttons like this to be linked to other connected frames
- click the navigation bar above with the sections, Home, Help, About, etc. (including the dots) to navigate in the presentation
- click on navigation buttons like this < 1/2 > to show you more of the current frame.
- click on many of the images to get more information or a close-up
- click the  or the  icon to go back to the menu
- click the buttons at the lower left and lower right that bring you to the next slide



Title page⇒



How to navigate

UNIL | Université de Lausanne

This presentation has been prepared for a PICO presentation at the EGU 2018 in Vienna, Austria. To facilitate the navigation, a lot of hyperlinks are used. Almost every item in this presentation is clickable:

- click buttons like this to be linked to other connected frames
- click the navigation bar above with the sections, Home, Help, About, etc. (including the dots) to navigate in the presentation
- click on navigation buttons like this < 2/2 > to show you more of the current frame. For example this text.
- click on many of the images to get more information or a close-up
- click the  or the  icon to go back to the menu
- click the buttons at the lower left and lower right that bring you to the next slide

Title page⇒

STRADITIZE

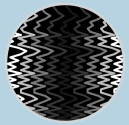
An open-source program for digitizing pollen diagrams and other types of stratigraphic data

Vienna, Austria, April 10th, 2018

Philipp Sommer, Basil A.S. Davis and Manuel Chevalier

Davis Group, Institute of Earth Surface Dynamics (IDYST)
University of Lausanne

UNIL | Université de Lausanne



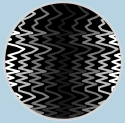
Author

UNIL | Université de Lausanne

Philipp Sommer

Institute of Surface Dynamics (IDYST)
University of Lausanne (UNIL)
Bâtiment Géopolis - Room 4413
CH - 1015 Lausanne

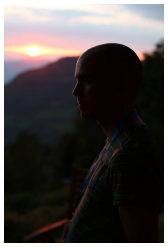
E-mail philipp.sommer@unil.ch
Telefon +41 21 692 35 40
Github `https://github.com/Chilipp/`
Webpage `https://wp.unil.ch/davisgroup/philipp-sommer/`



Acknowledgments

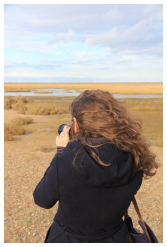
UNIL | Université de Lausanne

Dr. Basil Davis, palaeoclimatologist



Manuel Chevalier, palaeoclimatologist

Andrea Kay, archaeologist



Leanne Phelps, human ecologist

Prof. Jed Kaplan, geographer, climate modeler, ...

Shawn Koppenhöfer, informatician

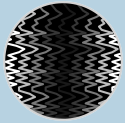
FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

UNIL | Université de Lausanne

- Olivier Cartapanis

⇐Author

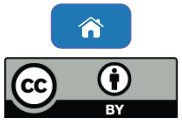
Abstract⇒



Abstract

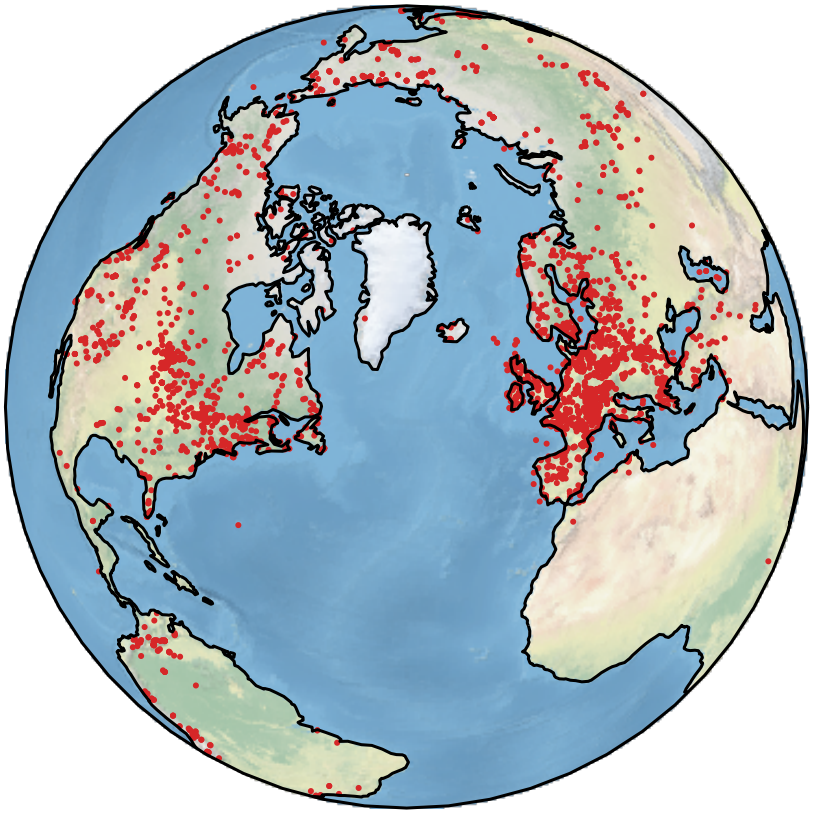
UNIL | Université de Lausanne

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 under-utilized 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 and other stratigraphic figures that are common in the Earth sciences. STRADITIZE (Stratigraphic Diagram Digitizer) is a new 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 is written in python and supports mixtures of many different diagram types, such as bar plots, line plots, as well as shaded, stacked, and filled area plots. The package 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. 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. Evaluation of the program has been undertaken comparing the digitization of published figures with the original digital data. This generally shows very good results, although this is inevitably reliant on the quality and resolution of the original figure.



Motivation

UNIL | Université de Lausanne



Large-scale climate reconstructions

- Global Pollen databases make large scale climate reconstructions possible
- But they lack of old publications from the 20th century whose data has never been published

http://www.neotomadb.org, http://www.europeanpollendatabase.net

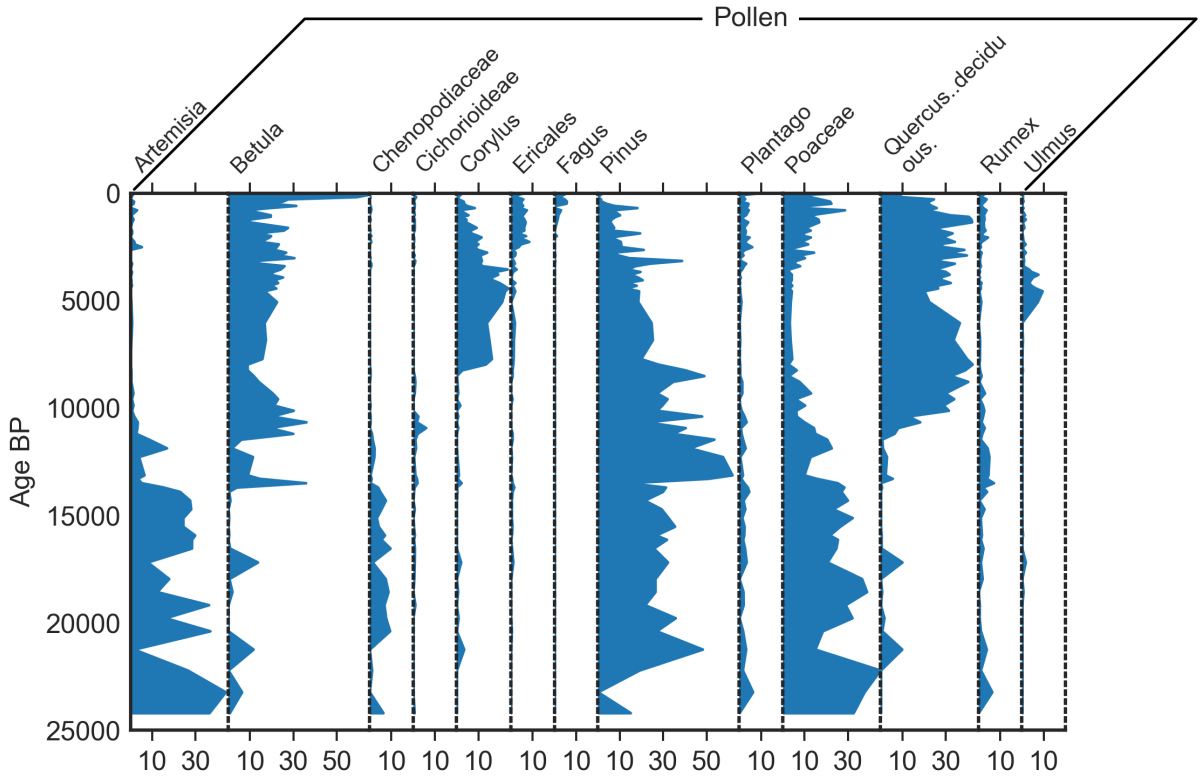
⇐Abstract

Idea behind⇒

Idea behind straditize

UNIL | Université de Lausanne

Use unpublished datasets, digitized from the published pollen diagram to fill the gaps.



Pollen Diagrams:

- One column per pollen taxa
- All share one axis, e.g. time or depth

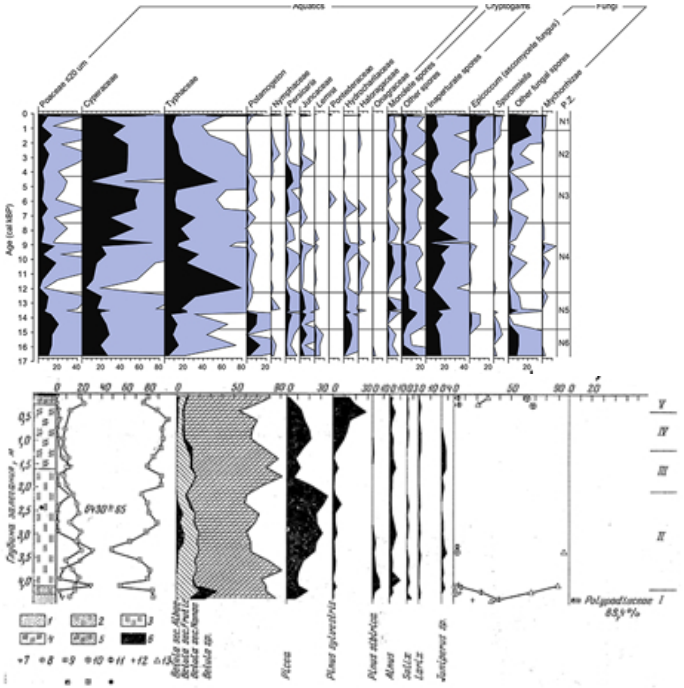
⇐Motivation

Challenges⇒

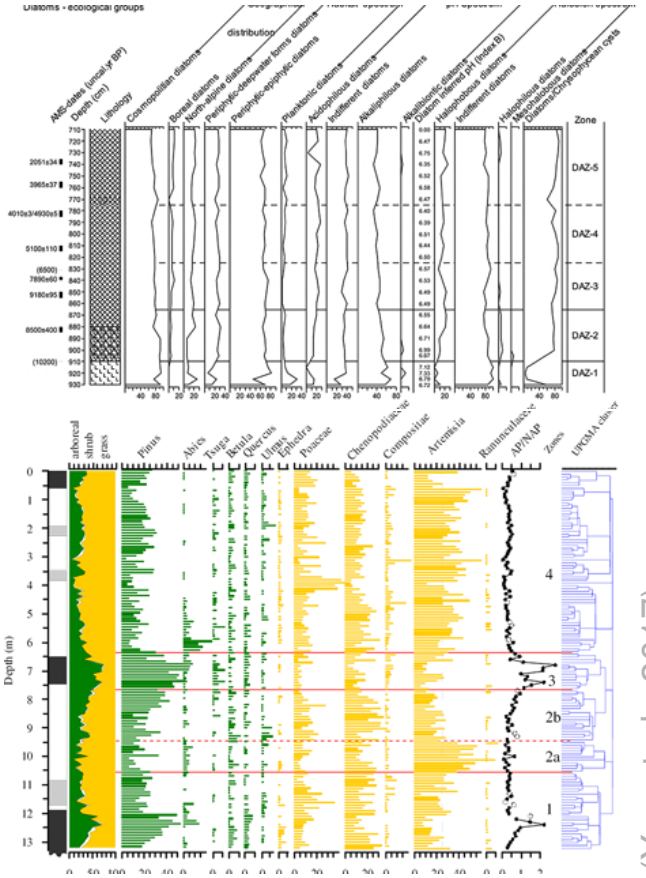
Challenges

UNIL | Université de Lausanne

- different kinds of pollen diagrams (bar plots, area plots, line plots, mixed plots, . . .)



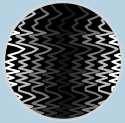
1/3



(Yao et al., 2017)

⇐Idea behind

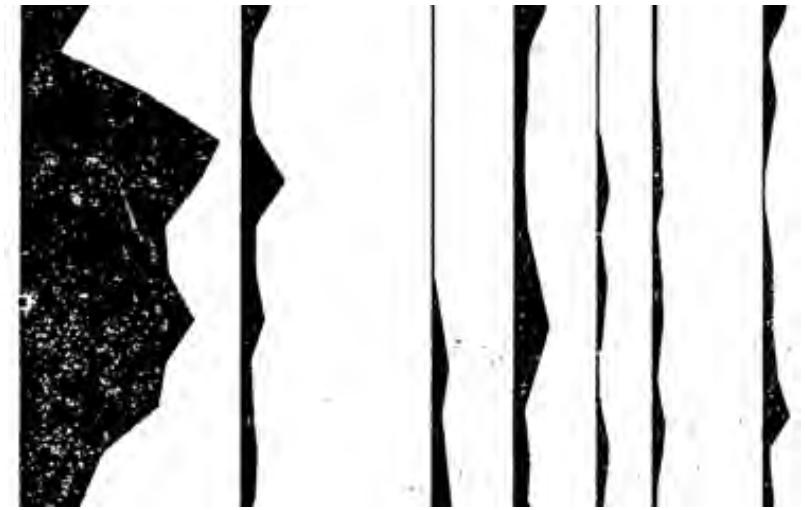
Installation⇒



Challenges

UNIL | Université de Lausanne

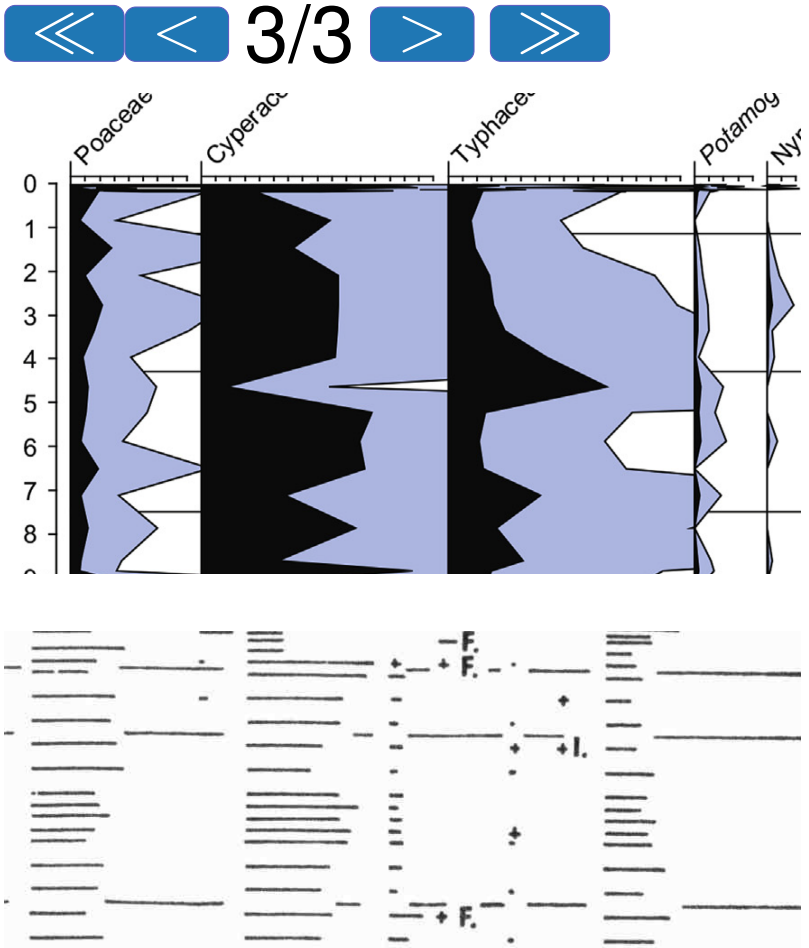
- different kinds of pollen diagrams (bar plots, area plots, line plots, mixed plots, . . .)
- bad quality of (old) diagrams

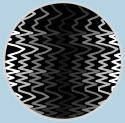


Challenges

UNIL | Université de Lausanne

- different kinds of pollen diagrams (bar plots, area plots, line plots, mixed plots, . . .)
- bad quality of (old) diagrams
- many unimportant features, such as exaggerations, letters, horizontal lines, etc.





Installation

The package is written in Python and hosted on Github:
`https://github.com/Chilipp/straditize`
and can be installed

1. from source:

```
git clone https://github.com/Chilipp/straditize.git
cd straditize
python setup.py install
```

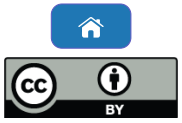
2. using pip (`https://pypi.org/`)

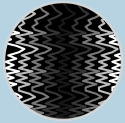
```
pip install straditize
```

3. or using anaconda (`https://anaconda.org/chilipp`)

```
conda install -c chilipp straditize
```

More installation methods will be provided soon.





Dependencies

UNIL | Université de Lausanne

The python package depends on different libraries

numpy and scipy: numeric python libraries (Jones et al., 2001)

matplotlib: python visualization package (Hunter, 2007)

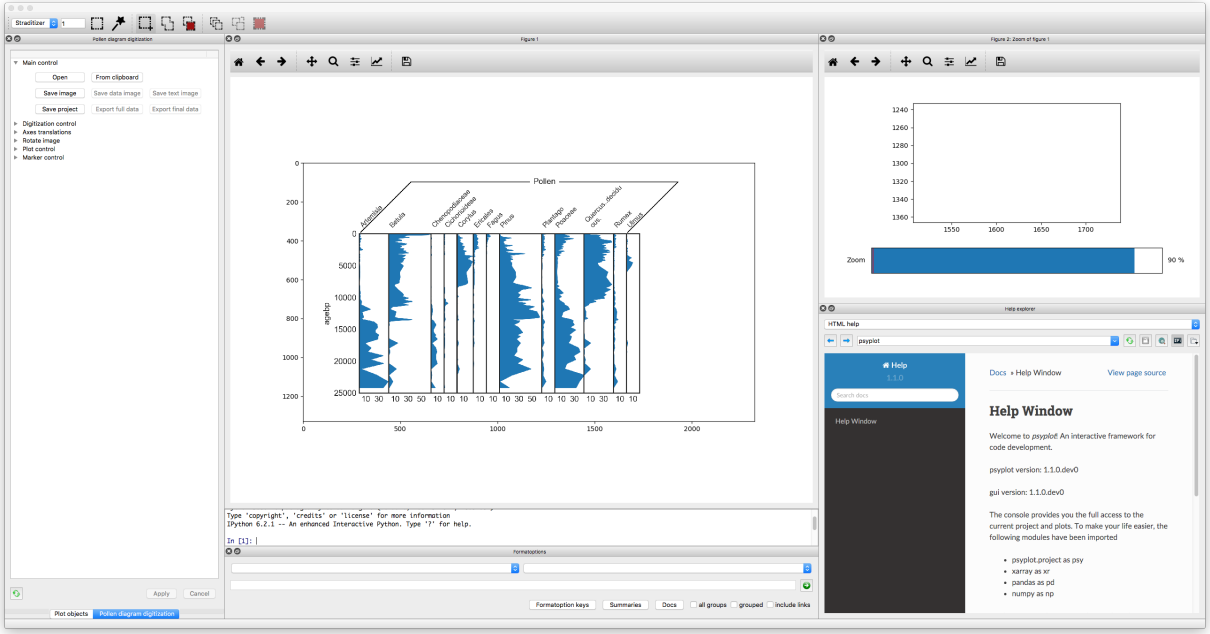
psyplot: interactive visualization framework (Sommer, 2017a)

scikit-image: image processing package for python (van der Walt et al., 2014)

Pillow: The image handling library (Clark et al., 2016)

The graphical user interface is based on the psyplot-gui package (Sommer, 2017b) which is programmed using the Qt bindings of PyQt (Summerfield, 2007).

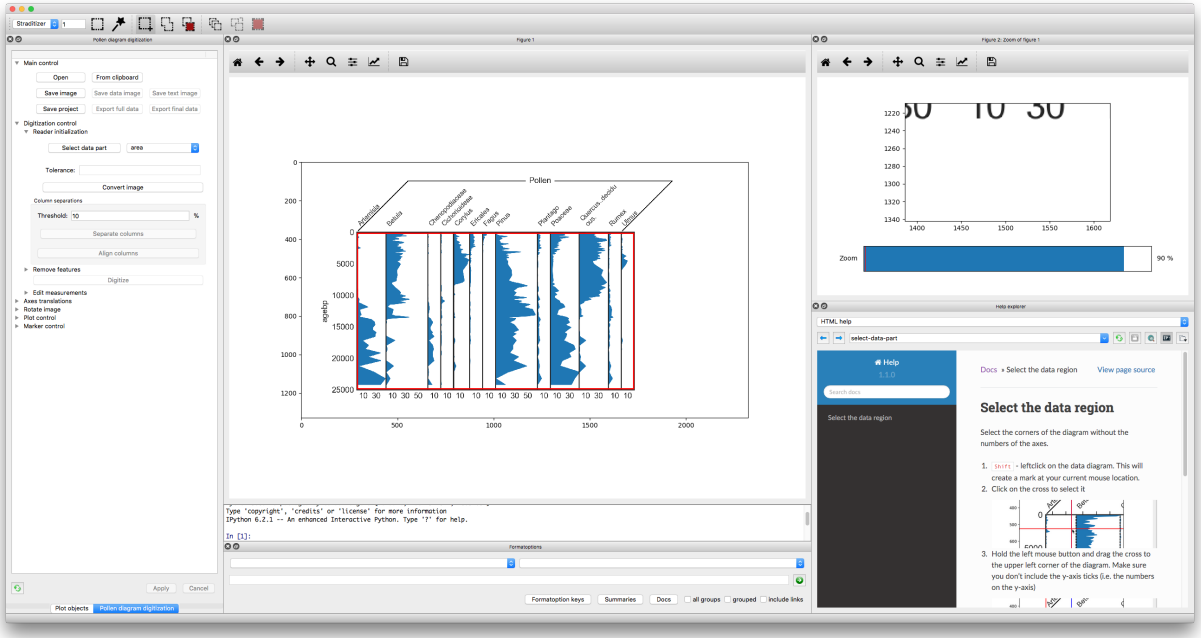
Demo



1 Load the picture. +

Demo

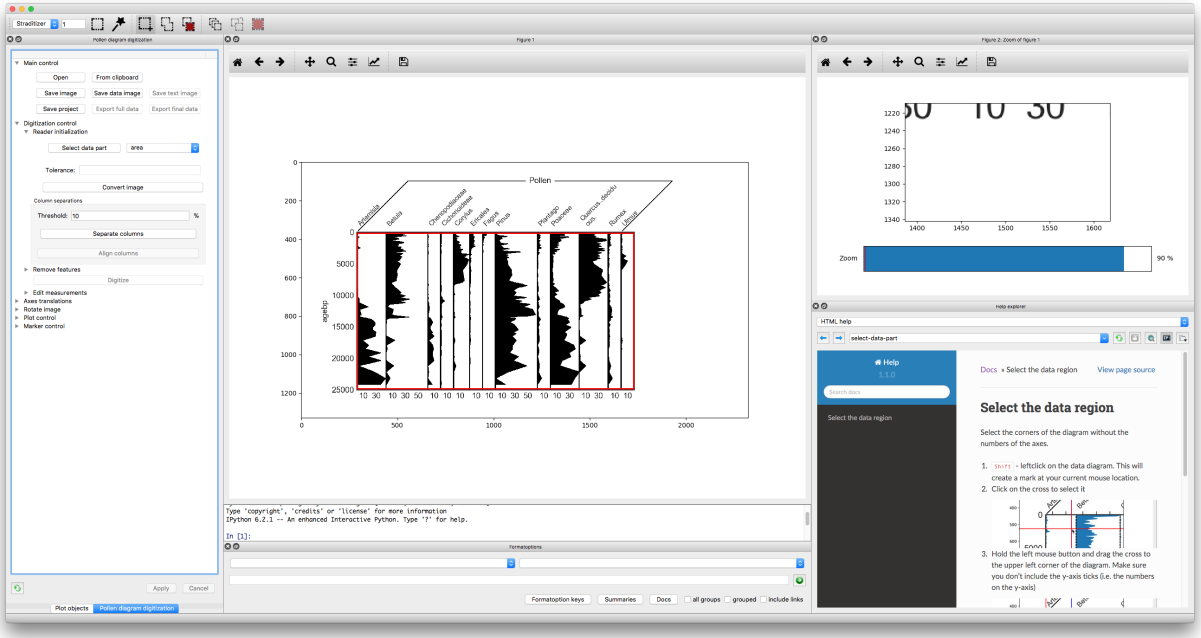
UNIL | Université de Lausanne



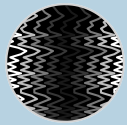
1. Load the picture. +
2. Select the diagram part. +

Demo

UNIL | Université de Lausanne

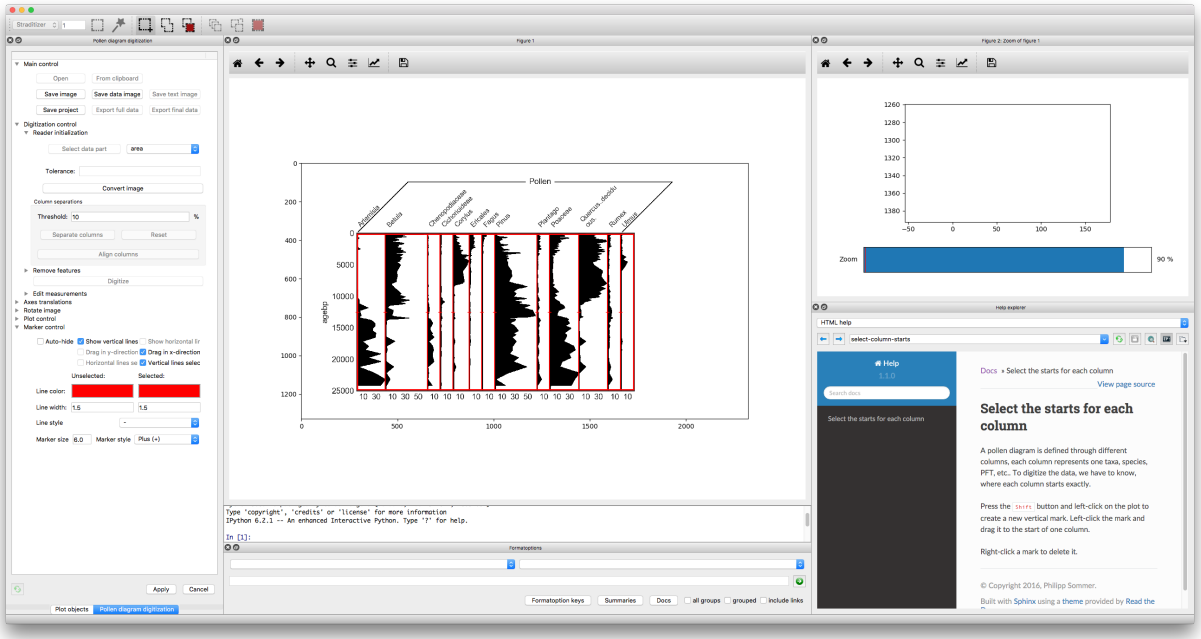


1. Load the picture. +
2. Select the diagram part. +
3. Convert to binary image. +



Demo

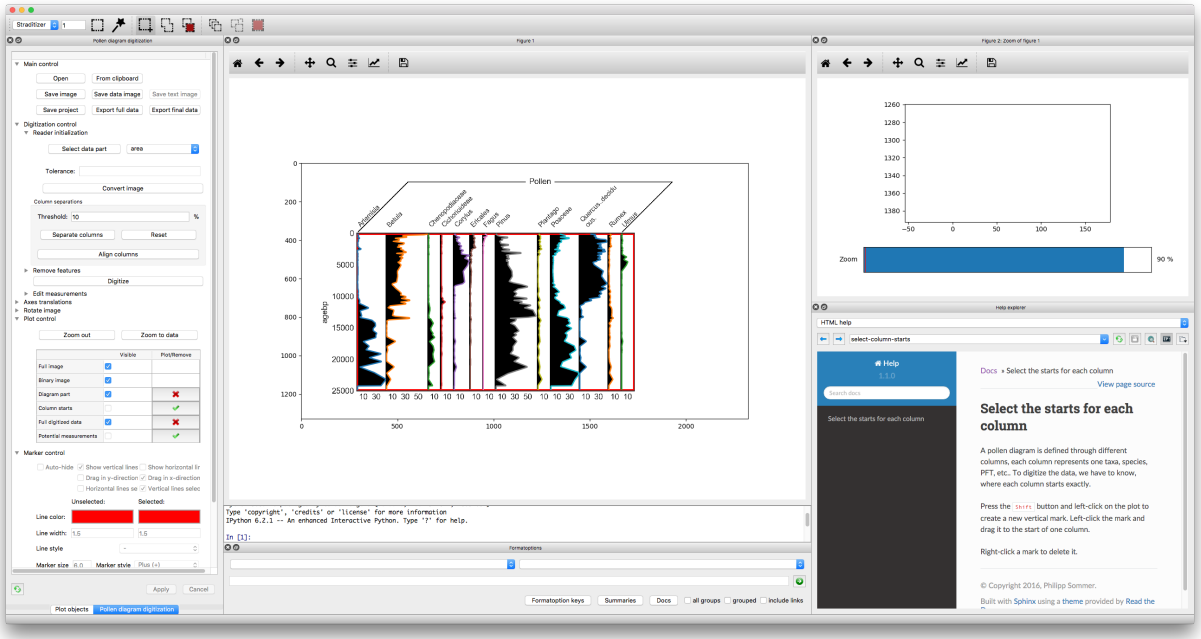
UNIL | Université de Lausanne



1. Load the picture. +
2. Select the diagram part. +
3. Convert to binary image. +
4. Automatically separate the columns. +

Demo

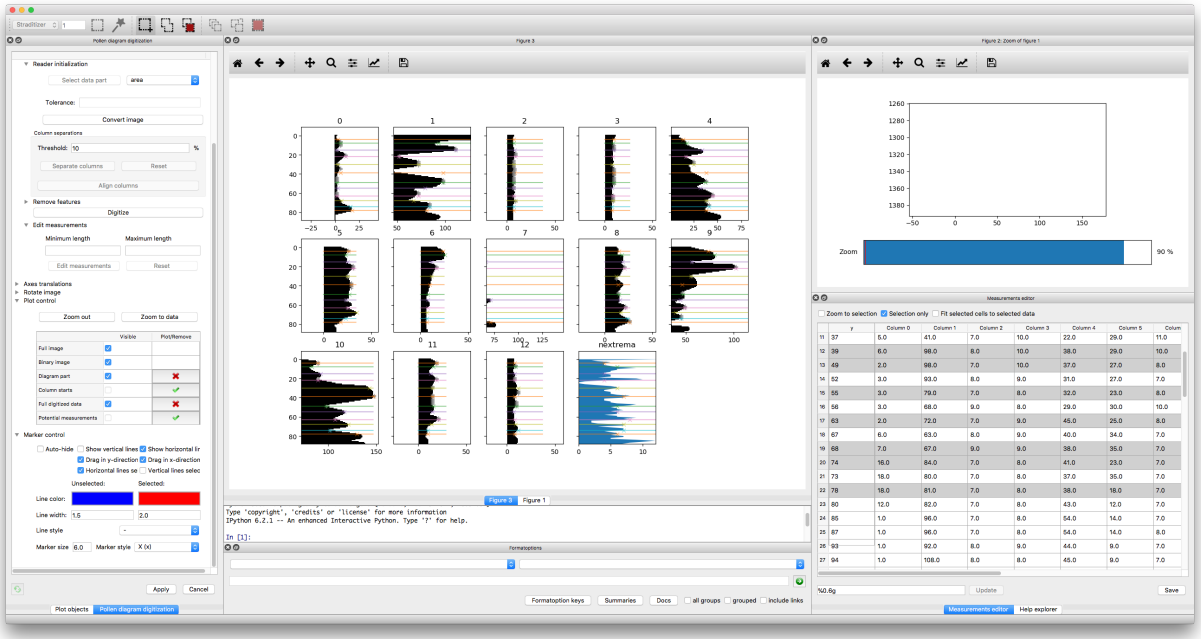
UNIL | Université de Lausanne



1. Load the picture. +
2. Select the diagram part. +
3. Convert to binary image. +
4. Automatically separate the columns. +
5. Automatically digitize the full data. +

Demo

UNIL | Université de Lausanne

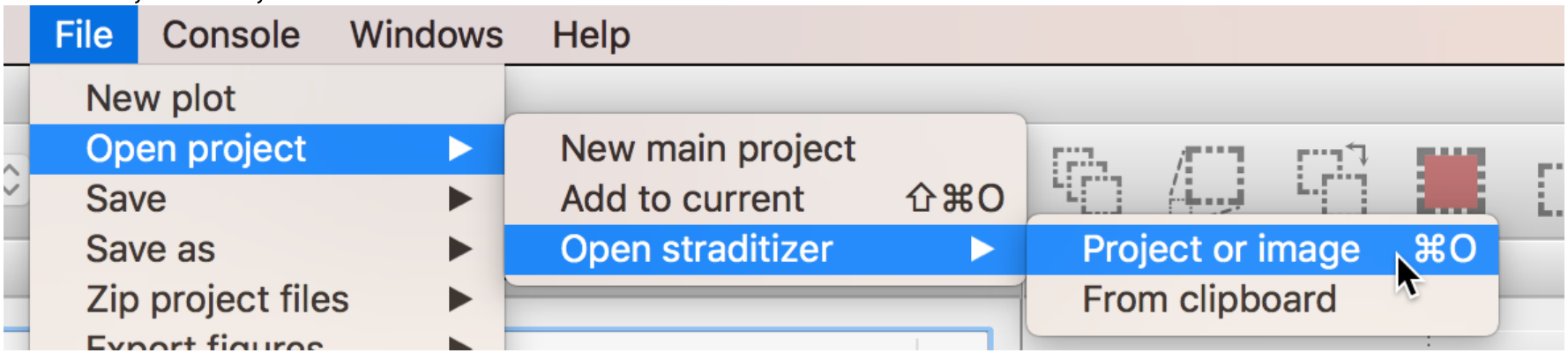


1. Load the picture. +
2. Select the diagram part. +
3. Convert to binary image. +
4. Automatically separate the columns. +
5. Automatically digitize the full data. +
6. Edit the measurements. +

Supported picture types

UNIL | Université de Lausanne

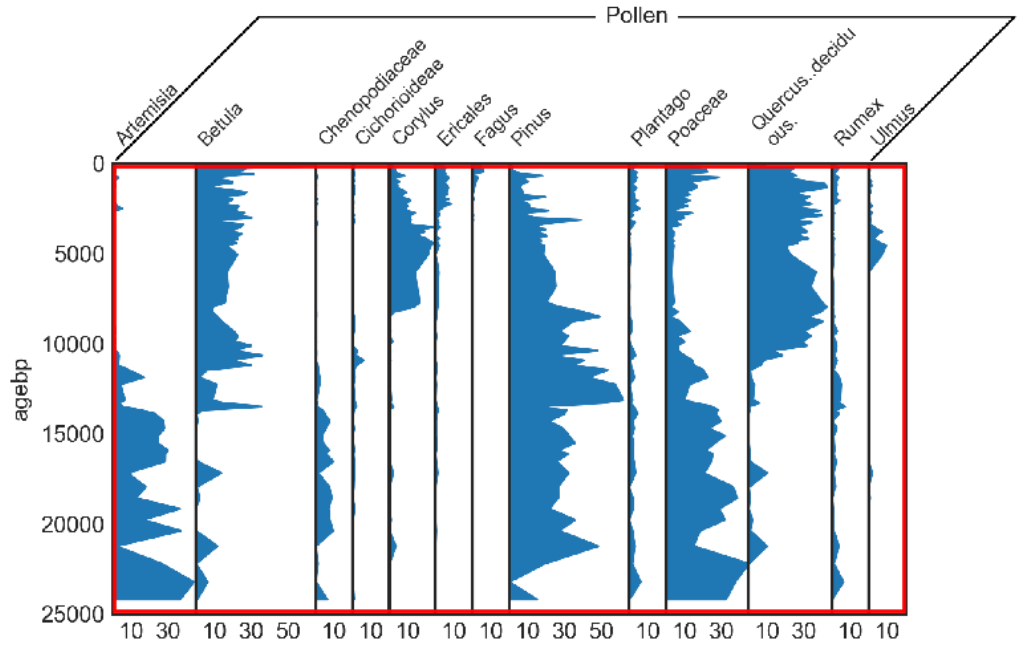
Straditize uses Pillow (Clark et al., 2016) to read and write images. Internally these images are then converted to an *RGBA* mode. Therefore all common image formats are supported, such as PNG, JPEG, TIFF, etc.



Back to demo

The diagram part

UNIL | Université de Lausanne



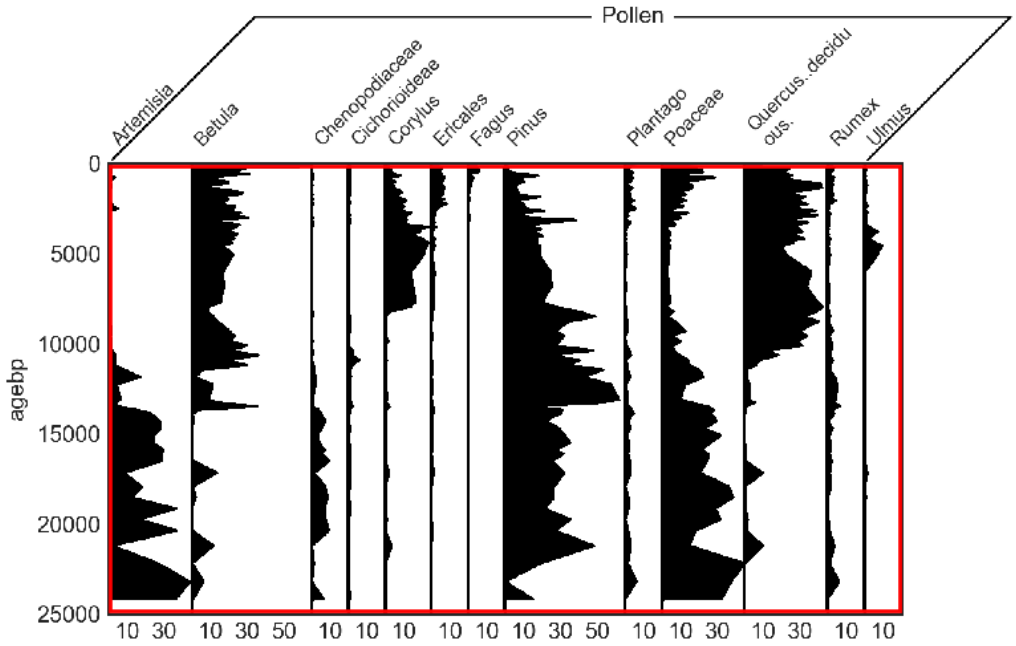
1/3

Back to demo

- The diagram part is the area where the data is shown without the axes and is selected by the user (red rectangle)

The diagram part

UNIL | Université de Lausanne



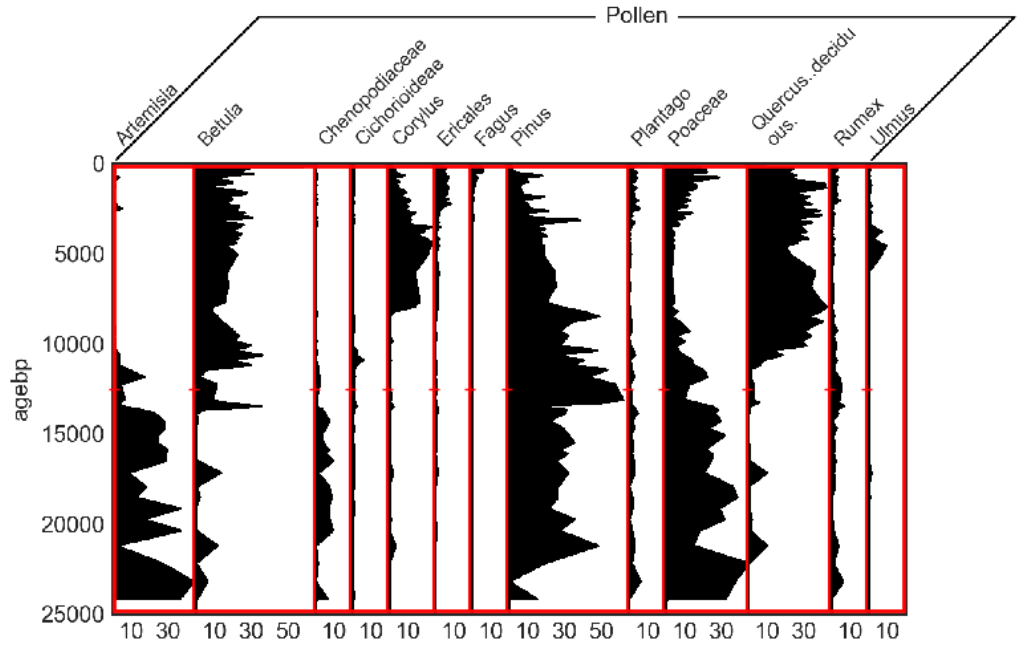
2/3

Back to demo

- The diagram part is the area where the data is shown without the axes and is selected by the user (red rectangle)
- It is converted to a binary image where everything **black** represents *data*

The diagram part

UNIL | Université de Lausanne

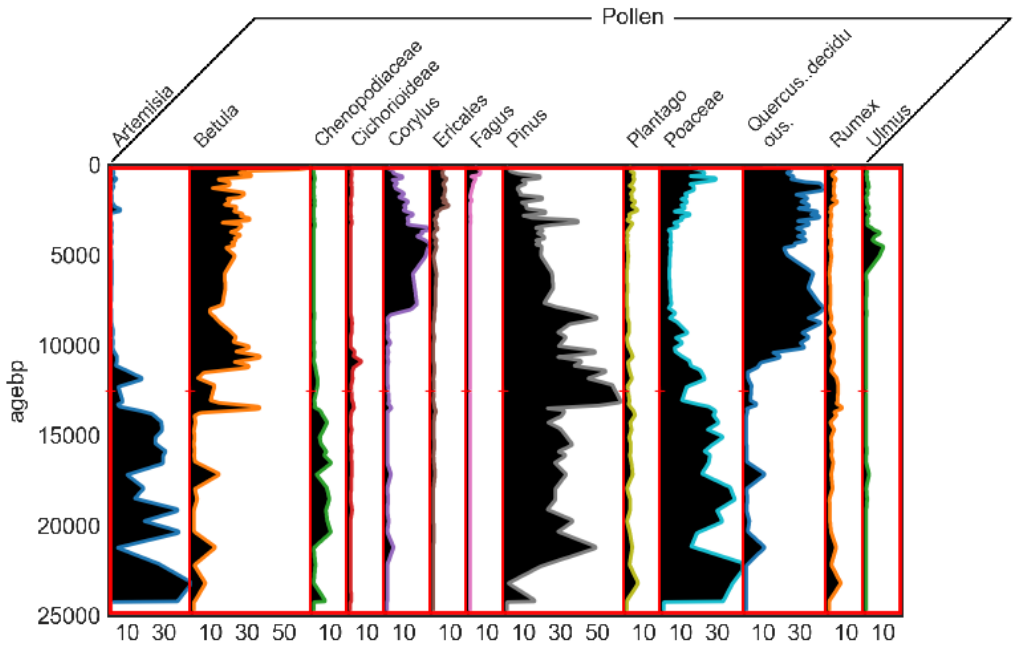


3/3

Back to demo

- The diagram part is the area where the data is shown without the axes and is selected by the user (red rectangle)
- It is converted to a binary image where everything **black** represents *data*
- Based on this *data*, the column starts are estimated (automatically)

Digitization



Back to demo

After having removed all features that do not represent data, the data is digitized based on the diagram type

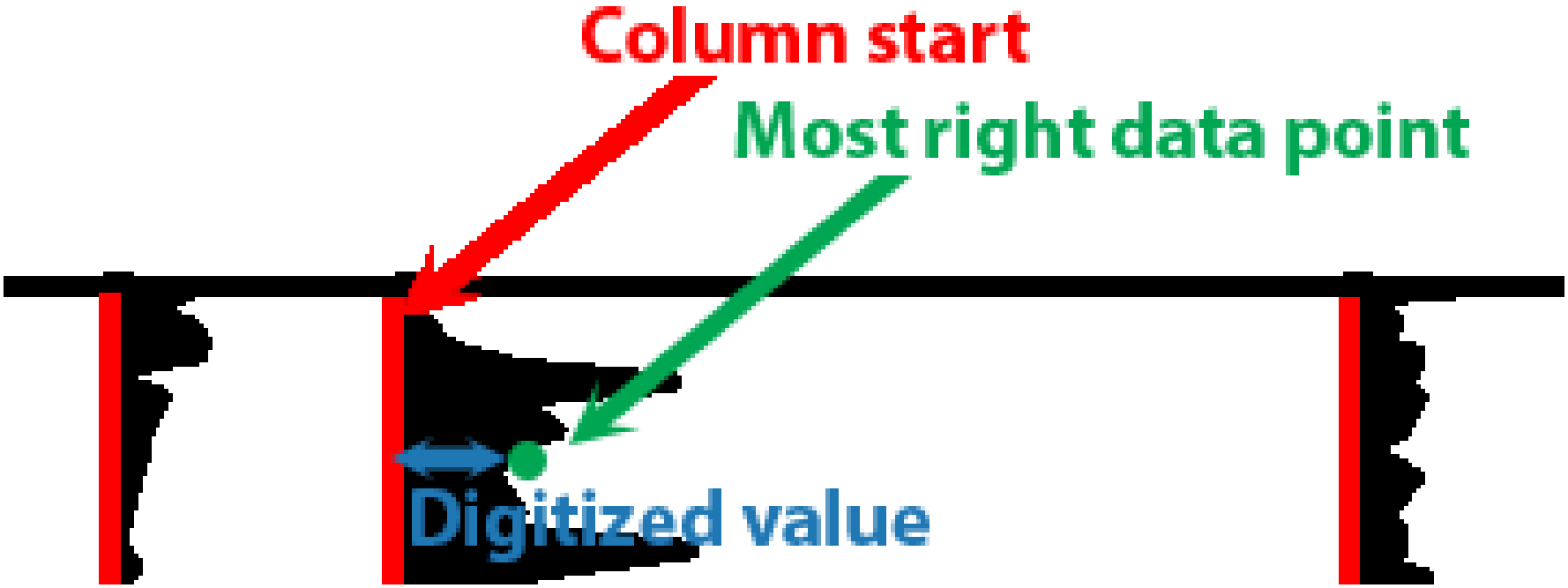
area/ line	take the last non-data point +
bars	find distinct bars and manually split over-lapping bars +
stacked area	manually select the features for each column +

⇐The diagram part

Area Digitization⇒

Digitization of an area or line plot diagram

UNIL | Université de Lausanne



This class of stratigraphic diagrams is the most popular within the pollen diagrams.
For the digitization, we use the distance of the most right *data* pixel to the column start.

Back to digitization

⇐Digitization

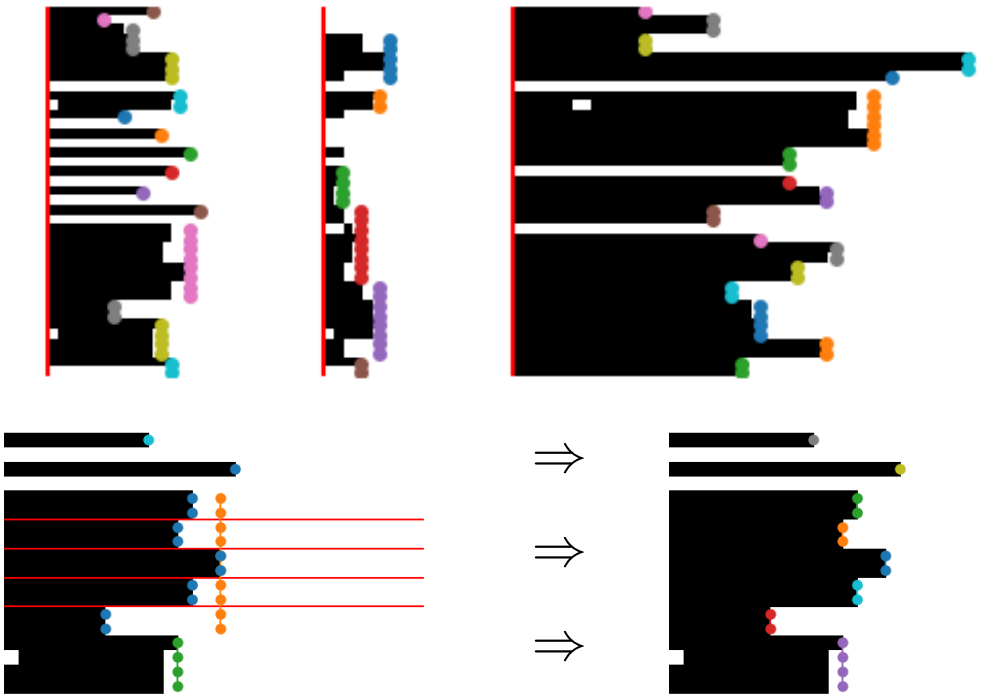
Bars Digitization⇒

Digitization of a bar diagram

- Bars are distinguished automatically based on white space between the data and/or a drastic increase

Digitization of a bar diagram

UNIL | Université de Lausanne



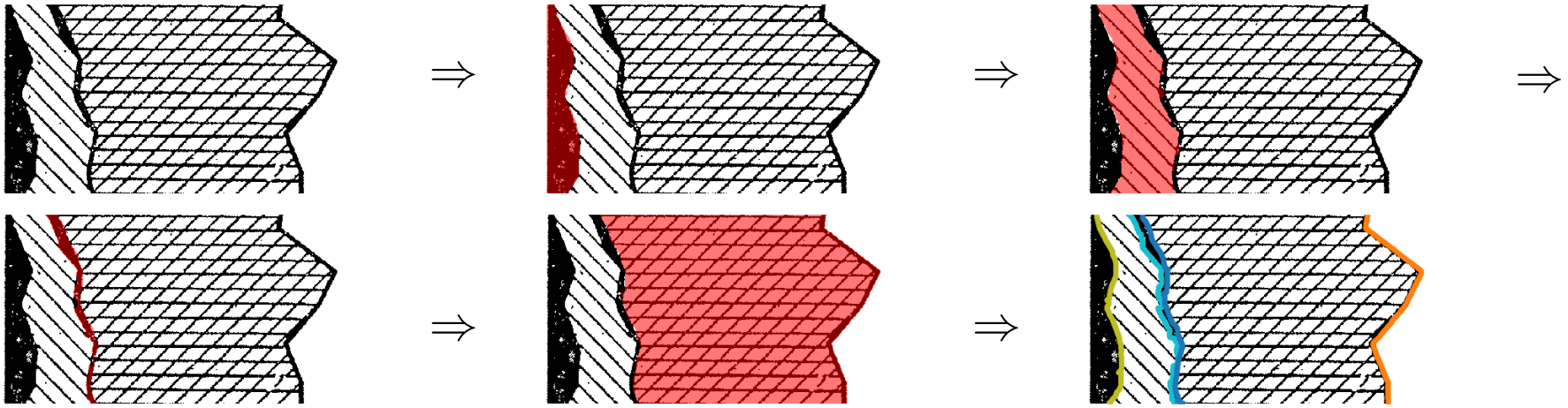
2/2

- Bars are distinguished automatically based on white space between the data and/or a drastic increase
- If the difference between overlapping bars is low, they are identified automatically and can easily be splitted manually

Back to digitization

Digitization of a stacked diagram

UNIL | Université de Lausanne

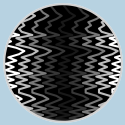


For digitizing a stacked area diagram, the user has to manually distinguish the columns. They can be selected by colors, patterns, connectivity or manually using a rectangle or polygon selector.

Back to digitization

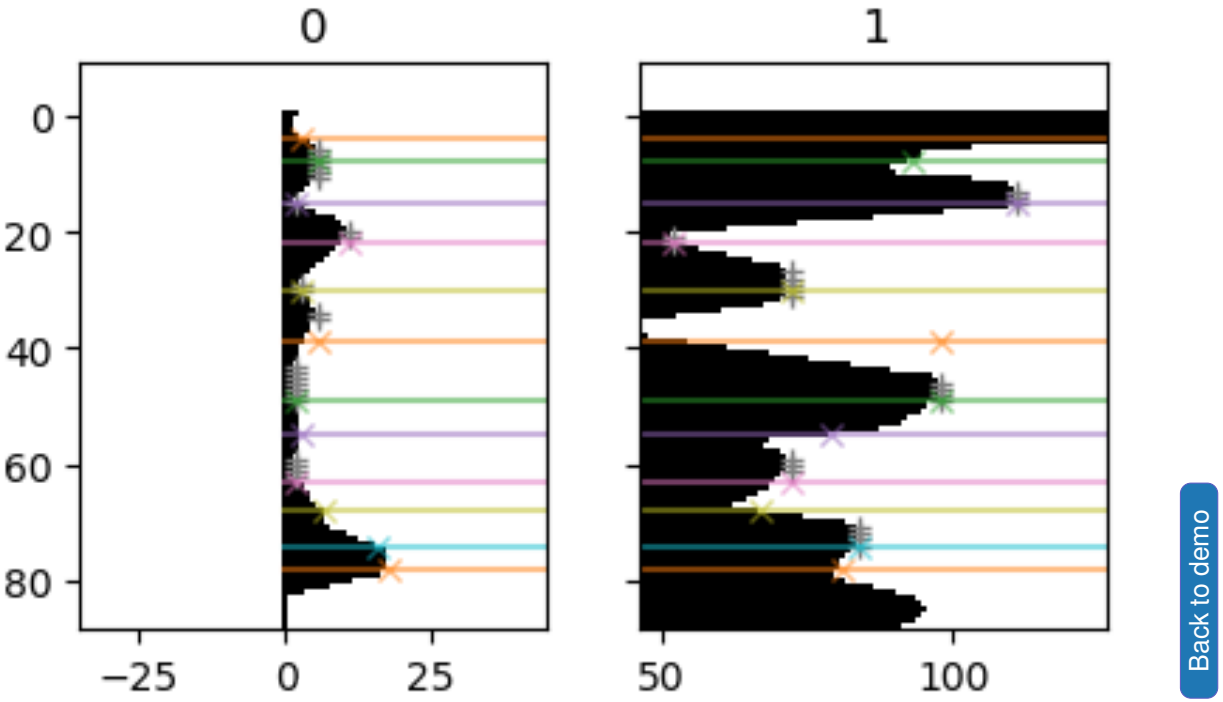
⇐Bars Digitization

Measurements⇒



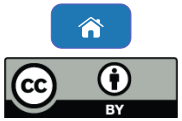
Finding of measurements

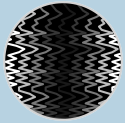
UNIL | Université de Lausanne



Back to demo

1. For each column (taxa) find the regions of potential extrema (minimum or maximum, grey + signs)
2. Find the overlaps within the columns to estimate the exact position of the measurements





References I

UNIL | Université de Lausanne

A. Clark, Hugo, A. Murray, A. Karpinsky, C. Gohlke, B. Crowell, D. Schmidt, A. Houghton, S. Johnson, and et al. Pillow: 3.1.0. Jan 2016. doi: 10.5281/zenodo.44297.

J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science Engineering*, 9(3):90–95, May 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.55.

E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001. URL `http://www.scipy.org/`. [Online; accessed 2017-02-18].

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

P. S. Sommer. Chilipp/psyplot-gui: v1.0.1: Graphical user interface for the psyplot package. Aug 2017b. doi: 10.5281/zenodo.845726. URL `https://doi.org/10.5281/zenodo.845726`.

M. Summerfield. *Rapid GUI Programming with Python and Qt (Prentice Hall Open Source Software Development)*. Prentice Hall, 2007. ISBN 978-0132354189. URL `https://www.amazon.com/Programming-Python-Prentice-Software-Development/dp/0132354187?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0132354187`.

S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014. ISSN 2167-8359. doi: 10.7717/peerj.453. URL `http://dx.doi.org/10.7717/peerj.453`.

F. Yao, C. Ma, C. Zhu, J. Li, G. Chen, L. Tang, M. Huang, T. Jia, and J. Xu. Holocene climate change in the western part of taihu lake region, east china. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 485:963 – 973, 2017. ISSN 0031-0182. doi: https://doi.org/10.1016/j.palaeo.2017.08.022. URL `http://www.sciencedirect.com/science/article/pii/S0031018217301062`.

Géopolis, Lausanne, Switzerland

UNIL | Université de Lausanne

Back to author page

Institute of Surface Dynamics (IDYST) - University of Lausanne (UNIL)
Bâtiment Géopolis - CH - 1015 Lausanne

⇐References