

# Welcome to PyREnArA!

## Table of Contents

Authors.....	1
About the project .....	1
Workflow .....	2
Required software .....	3
Running PyREnArA.....	3
Literature.....	4

## Authors

Name	Affiliation	Email
Robin John	Institute of Prehistoric Archeology, University of Cologne, Bernhard- Feilchenfeld-Str. 11, 50969 Köln, Germany.	robin_john@web.de
Florian Linsel	Institute of Computer Science, Martin- Luther-Universität Halle-Wittenberg, Von- Seckendorff-Platz 1, 06120 Halle, Germany.	linsel.florian@gmail.com
Dr. Georg Roth	Institut für Prähistorische Archäologie, Freie Universität Berlin, Fabeckstr. 23-25, 14195 Berlin, Germany	g.roth@fu-berlin.de
Dr. Andreas Maier	Institute of Prehistoric Archeology, University of Cologne, Bernhard- Feilchenfeld-Str. 11, 50969 Köln, Germany.	a.maier@uni-koeln.de

## About the project

**PyREnArA (Python-R-Environment for Artifact Analysis)** is a code for half-automated, trait-based recording of morphological properties of (stone-)artifacts and the statistical evaluation of the recorded properties. PyREnArA, as the name says, is written both in Python and R. It is specialized for the analysis of paleolithic projectile points depicted in drawings or photographs.

It provides traditional and new information on artifact morphology, customized for the implemented statistical methods, which allow for quantitative analyses of morphological change and to statistically determine the amount of variation that correlates with the progress of time or with the geographical position. In doing so, it provides insight into material culture evolution far beyond traditional typology. However, as completely independent analytical system, it is not intended to substitute typological research, but rather to complement it and to provide new perspectives on the archaeological record, which are otherwise invisible. The recording system, on which PyREnArA is based on, has been developed over several semesters in seminars with students from the Friedrich-Alexander-University Erlangen-Nürnberg and the University of Cologne. The coding was done by Robin John, Florian Linsel and Dr. Georg Roth.

## **Workflow**

### **Image Preparation**

The user needs to prepare the image(s) to be analyzed by adding a black filled rectangle with edge length of one by one centimeter obtained from the original scale as reference object and by removing objects that should not be analyzed.

### **Code intern image preparation**

Artifact drawings and photos are automatically scanned and oriented according to a standard protocol to minimize measurement errors

### **Metrical data**

Traditional as well as new traits of object morphology and size, such as dimensions, the geometric center, the outline to area ratio, angles, symmetry, the position of the longest extension (e.g., longitudinal or transversal), are obtained automatically. While most recorded traits are geometric and metric properties of the objects, PyREnArA also computes the fractal dimension (FD; Seidel, 2018) of the artifact outline using the box-counting method. Those obtained properties are saved within automatically created folders as images and .csv-datafiles.

### **Data merging**

All data is merged into one .csv-datafile. If the users wants so, the metrical data can be merged with a .csv-datafile provided by the user himself, which contains metadata like chronological information or geographical position. Those -csv-datafiles can be used within the build in statistical functions of PyREnArA.

### **Geometric Morphometrics**

The recorded contours are analyzed with regard to their shape. PyREnArA uses an elliptic Fourier analysis (EFA), which is a well-established procedure in geometric morphometrics (Claude, 2008, Bonhomme et al., 2014) and has proven its potential for lithic analysis (e.g.,

Leplongeon et al., 2020; Matzig et al., 2020). The resulting outline EFA data is then submitted to a between group PCA (Cardini & Polly, 2020). The PCA scores are used to distinguish an optimal number of clusters, determined by a post hoc analysis using the silhouette criterion – i.e., an internal cluster validity criterion (Schmidt et al., 2022) – in a hierarchical agglomerative clustering algorithm (Murtagh & Legendre, 2014). Besides classical applications of geometric morphometrics, the classes separable by their shape are used in order to assess the diversity present within the dataset.

#### Explorative statistics

The obtained artefact properties can be used to perform some explorative statistics. The metrical data can be compared and displayed in a pairsplot showing Pearson correlation coefficients (McKinney, 2010). The metrical data can be plotted against time to observe for chronological trends. This is also possible for the standard deviation of each trait, which gives insights into the level of standardization. Also a PCA and a LDA can be performed individually.

#### Redundancy Analysis

In order to identify possible evolutionary trends, PyREnArA provides a function for redundancy analysis (RDA; Legendre & Legendre, 2012; Roth 2022), which uses the chronological position of the tools as canonical variable. Of course, also other criteria (e.g., the geographic location) can be used as canonical variables to assess broader spatial trends.

#### Retouch analysis

The standard-oriented images are given as output-files with lateral bars for each artifact. These bars are used by the researcher to color-code information on retouches. Subsequently, the thus-prepared images can be uploaded again in order to obtain information on length, position, and kind of retouch are extracted.

### **Required software**

PyREnArA requires Python and R. In order to run functions in Python and R, we recommend using Anaconda and Jupyter Notebook. PyREnArA makes use of the rpy2-library, that allows for a combination of both programming languages. It is highly recommended to install all software in the default directory and settings.

### **Running PyREnArA**

PyREnArA consists of two files where R and Python Scripts are stored separately. Within the Python file a main.py script is provided. There, the user has to insert the folder directory that stores the images and the directory to R. We recommend using .png or .tif format for the images.

The naming of the images should be strictly following the scheme “SitenamePosition-Layer\_number” (e.g. AggsbachB-3\_04) as the code develops the naming of the folder directory

and data from the image name. Do not use spaces within the naming. The data from artefacts with identical site name and layers will later be merged together. When Paint is used to create the reference object or to color for retouch analysis, the image needs to be saved again by another program (e.g. Adobe Photoshop). As mentioned before, the code creates a folder directory for every image, where all output is stored in. It also creates folders for the results of the statistical functions, that are by default named “R\_statistics” and “Python\_statistics”. The names of the statistical folders can be customized easily. The default set of metrical data, that is used in the statistical functions can be customized by changing the “feature” argument of the functions, which is a list of strings ([‘...’, ‘...’]).

## Literature

Bonhomme, V., Picq, S., Gaucherel, C. & Claude, J. (2014). Momocs: Outline analysis using R. *Journal of Statistical Software*, 56, 1–24.

Cardini, A. & Polly, P.D. (2020). Cross-validated Between Group PCA Scatterplots: A Solution to Spurious Group Separation? *Evolutionary Biology*, 47, 85–95. <https://doi.org/10.1007/s11692-020-09494-x>

Claude, J. (2008). Rfunctions1 (R functions for geometric morphometrics). In: Claude, J. (Ed.) *Morphometrics with R*. <https://doi.org/10.13140/RG.2.2.24525.36324>

Legendre, P., & Legendre, L. (2012). *Numerical ecology*. 3<sup>rd</sup> edition.

Leplongeon, A., Ménard, C., Bonhomme, V. & Bortolini, E. (2020). Backed Pieces and Their Variability in the Later Stone Age of the Horn of Africa. *African Archaeological Review*, 37, 437–468. <https://doi.org/10.1007/s10437-020-09401-x>

Matzig, D.N., Hussain, S.T. & Riede, F. (2021). Design Space Constraints and the Cultural Taxonomy of European Final Palaeolithic Large Tanged Points: A Comparison of Typological, Landmark-Based and Whole-Outline Geometric Morphometric Approaches. *Journal of Paleolithic Archaeology*, 4: 27. <https://doi.org/10.1007/s41982-021-00097-2>

McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 51–56.

Murtagh, F. & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, 31, 274–295. <https://doi.org/10.1007/s00357-014-9161-z>

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

Roth, G. (2022). Überflüssige Information? Zum Verständnis moderner Deutung archäologischer Zusammensetzungsdaten mit transformationsbasierter Redundanzanalyse (tb-RDA). In: Kaiser, E. Meyer, M., Scharl, S. & Suhrbier, St. (Eds.), *Wissensschichten. Festschrift für Wolfram Schier zu seinem 65. Geburtstag*. Studia honoraria 41 (pp. 28–42).

Schmidt, S. C., Martini, S., Staniuk, R., Quatrelivre, C., Hinz, M., Nakoinz, O., Bilger, M., Roth, G., Laabs, J., & Plath, R. V. (2022). Tutorial on Classification in Archaeology: Distance Matrices, Clustering Methods and Validation. *Zenodo*. <https://doi.org/10.5281/zenodo.6325372>

Seidel, D. (2018). A holistic approach to determine tree structural complexity based on laser scanning data and fractal analysis. *Ecology and Evolution*, 8, 128–134. <https://doi.org/10.1002/ece3.3661>

Thioulouse J, Dray S, Dufour A, Siberchicot A, Jombart T, Pavoine S (2018). Multivariate Analysis of Ecological Data with ade4. Springer. <https://doi.org/10.1007/978-1-4939-8850-1>

Viengkham, C., Isherwood, Z. & Spehar, B. (2019). Fractal-Scaling Properties as Aesthetic Primitives in Vision and Touch. *Axiomathes*, 2019. <https://doi.org/10.1007/s10516-019-09444-z>