Multivariate statistical protocols to illuminate provenance studies of ceramic materials: showcasing the cerUB package in R

Andreas Angourakis, Verònica Martínez Ferreras, and Josep M. Gurt Esparraguera

ERAAUB, Department of History and Archaeology, University of Barcelona, Barcelona, Spain.

contact: Andreas Angourakis andros.spica@gmail.com

1. Overview

- Statistics and statistical software are a **challenge** for most archaeologists
- Different **archaeometric** analysis on ceramics are often **never integrated through statistics**
- Need for **concrete**, **well-described**, and **ready-to-use** statistical protocols to approach common archaeological questions/data, while retaining the advantages of **open-source** software
- A challenge for multivariate statistics: cope with different data formats (continuous and ordinal variables)
- We propose **four protocols** to integrate the most common archaeometric data available for ceramics: geochemical, mineralogical, and petrographic data
- The **cerUB** package includes the functions needed to perform these protocols [1]
- We demonstrate the four protocols with two datasets: wine Roman **amphorae** from NE Spain and **tableware** of more diverse chronology from SE Uzbekistan

2. Data sources

- **Geochemical composition**: X-ray fluorescence readings (XRF-WD) using a Philips PW 2400 spectrometer
- **Range of firing temperature**: estimated with mineralogical phase by X-ray diffraction readings (XRD) using Siemens D-500 and Panalytical X'Pert PRO alpha 1 diffractometers
- **Petrographic observations**: thin-section analysis using polarising optical microscope Olympus BX41, digital camera Olympus DP70, and Analysis Five software (following Whitbread 1995[*])

3. Methods

- **Compositional data** require a log-ratio transformation (Aitchison 1982[*], Filzmoser et al. 2009[*], Pawlowsky-Glahn & Bucciantti 2011[*])
- **Ordinal variables** require a specific treatment (Podani 1999[*])
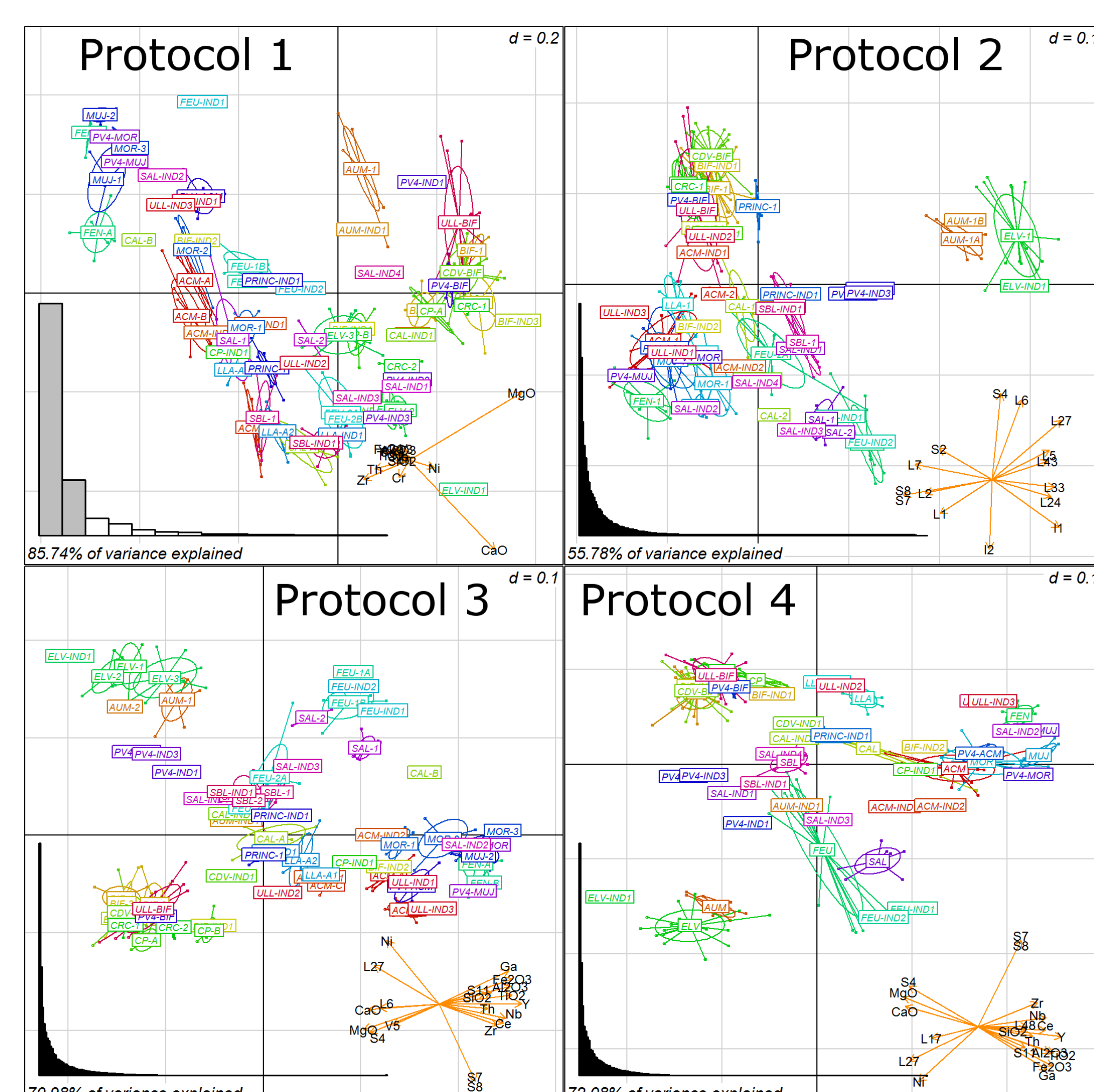- **'Mixed-mode'** approach using Gower coefficient of (dis)similarity (Pavoine et al. 2009[*])

4. Case study 1: amphorae

- Wine Roman amphorae from *Hispania Citerior-Tarraconensis* (NE Spain)
- 175 individuals from 15 workshops (e.g., BIF; Martínez Ferreras 2014[*])
- 61 individuals from 3 shipwrecks in the region (CDV, PV4, ULL; Martínez Ferreras et al. 2015, Martínez Ferreras et al. 2013[*])
- Dated from c. 75 BC to c. 50 AD

5. Case study 2: tableware

- Tableware from NW ancient *Bactria* (Surkhan Darya region, Uzbekistan)
- 45 individuals from 4 sites: Kampyr Tepe (KPT) [2] and three locations associated to ancient Termez (KT, TT, ISL) [4]
- KPT dates from the Hellenistic Period (c. 327-140 BC), TT the Yuezhi Period (c. 148 BC to 78 AD), KT the Kushan-Sassanian Period (c. 280-400 AD), and ISL the Late Pre-Mongol Islamic Period (c. 1000-1200)

Protocols	1	2	3	4
Data	CHEM	PETRO	CHEM and PETRO	CHEM and PETRO_PROV
Transformation	ilr	ranking	clr & ranking	clr & ranking
Distance function	Euclidean	RRD/NI	Extended Gower	Extended Gower
Ordination method	robust PCA	PCoA/NMDS	PCoA	PCoA

CHEM: Geochemical composition; **PETRO**: Range of firing temperature and petrographic observations; **PETRO_PROV**: A selection of PETRO variables aimed at provenance identification; **ilr**: isometric log-ratio; **clr**: centred log-ratio; **RRD**: relative ranking difference (Podani 1999[*]); **NI**: neighbour interchange (Podani 1999[*]); **Extended Gower**: see Pavoine et al. 2009[*]; **PCA**: Principal Components Analysis; NMDS: Non-metric Multidimensional Scaling; **PCoA**: Principal Coordinates Analysis

6. Conclusion

- **Data integration** is possible, even with different types of variables
- The **'mixed-mode'** approach facilitates the detection of groups in space and time (also agrees with Baxter et al. 2008[*]
- Tools such as R [3] promote the **comparability** and **reproducibility** of research
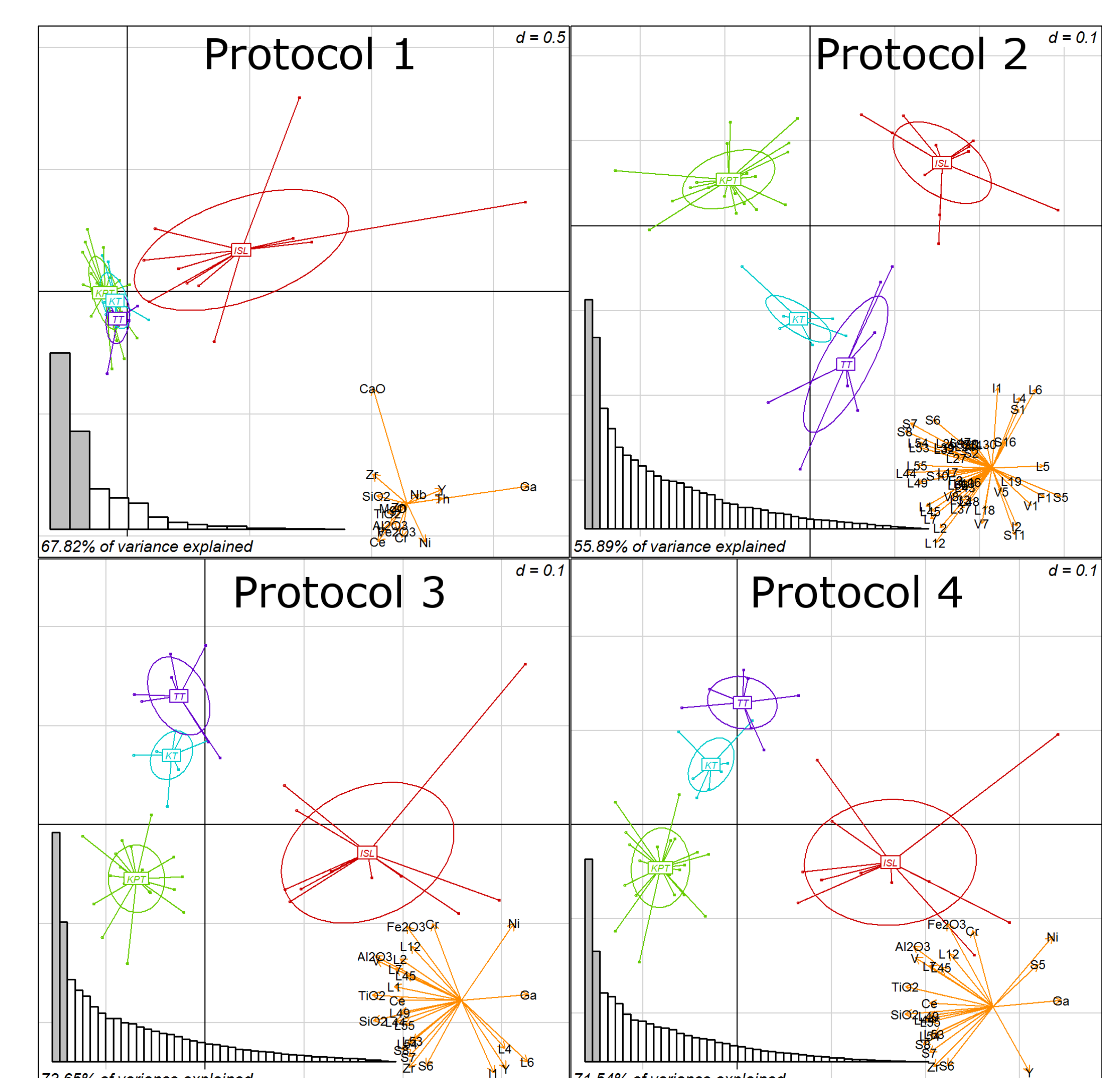
Further information

The **cerUB** package has a tutorial online, showing how to install and use the package following the analyses in the reference article [1]

https://andros-spica.github.io/cerUB_tutorial

References

[*] Marked references can be found in [1]

[1] A. Angourakis, V. Martínez Ferreras, A. Torrano, and J. M. Gurt Esparraguera. Presenting multivariate statistical protocols in R using Roman wine amphorae productions in Catalonia, Spain. *Journal of Archaeological Science*, 93:150–165, may 2018.

[2] V. Martínez Ferreras, J. M. Gurt Esparraguera, A. Hein, S. Pidaev, E. V. Rtveladze, and S. B. Bolelov. Tableware in the Hellenistic Tradition from the City of Kampyr Tepe in Ancient Bactria (Uzbekistan). *Archaeometry*, pages 736–764, nov 2015.

[3] R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2015.

[4] E. Tsantini, V. Martínez Ferreras, E. Ariño Gil, J. M. Gurt I Esparraguera, and S. Pidaev. Pottery Production in the Buddhist Communities in Central Asia: The Kushan-Sassanian Pottery Workshop of Kara Tepe (Termez, Uzbekistan). *Archaeometry*, 58(1):35–56, feb 2016.