Ordination analyses in sedimentology, geochemistry and paleoenvironment - current trends and recommendations

Or Mordechay Bialik[1], **Emilia Jarochowska**[2], and Michal Grossowicz[3]

[1]Marine Geology and Seafloor Surveying, Department of Geosciences, University of Malta, Msida MSD 2080, Malta.

[2]GeoZentrum Nordbayern, Friedrich Alexander University of Erlangen-Nürnberg, 91054 Erlangen, Germany.

[3]GEOMAR Helmholtz Centre for Ocean Research Kiel, 24105 Kiel, Germany.

Ordination is a family of multivariate exploratory data analysis methods. With the advent of high-throughput data acquisition protocols, community databases, and multiproxy studies, the use of ordination in Earth sciences has snowballed. As data management and analytical tools expand, this growing body of knowledge opens new possibilities of meta-analyses and data-mining across studies. This requires the analyses to be chosen adequately to the character of Earth science data, including pre-treatment consistent with the precision and accuracy of the variables, as well as appropriate documentation. To investigate the current situation in Earth sciences, we surveyed 174 ordination analyses in 163 publications in the fields of geochemistry, sedimentology and palaeoenvironmental reconstruction and monitoring. We focussed on studies using Principal Component Analysis (PCA), Non-Metric Multidimensional Scaling (NMDS) and Detrended Correspondence Analysis (DCA).

PCA was the most ubiquitous type of analysis (84%), with the other two accounting for ca. 12% each. Of 128 uses of PCA, only 5 included a test for normality, and most of these cases were not applied or documented correctly. Common problems include: (1) not providing information on the dimensions of the analysed matrix (16% cases); (2) using a larger number of variables than observations (24 cases); (3) not documenting the distance metric used in NMDS (55% cases); and (4) lack of information on the software used (38% cases). The majority (53%) of surveyed studies did not provide the data used for analysis at all and a further 35% provided data sets in a format that does not allow immediate, error-free reuse, e.g. as data table directly in the article text or in PDF appendix. The "golden standard" of placing a curated data set in an open access repository was followed only by 6 (3%) of the analyses. Among analyses which reported using code-based statistical environments such as R Software, SAS or SPSS, none provided the code that would allow reproducing the analyses.

Geochemical and Earth science data sets require expert knowledge which should support analytical decisions and interpretations. Data analysis skills attract students to Earth sciences study programmes and offer a viable research alternative when field- or lab-based work is limited. However, many study curricula and publishing process have not yet endorsed this methodological progress, leading to situations where mentors, reviewers and editors cannot offer quality

assurance for the use of ordination methods. We provide a review of solutions and annotated R Software code for PCA, NMDA and DCA of geochemical data sets in the freeware R Software environment, encouraging the community to reuse and further develop a reproducible ordination workflow.