An extensible, open, and web-based research environment to understand geochemical data

T. Tamanna[1], D.C. Hezel[1]

Institute of Geosciences, Goethe University, Frankfurt, Altenhöferallee 1, 60438 Frankfurt am Main, Germany
e-mail: tamanna@uni-frankfurt.de

Introduction: The application of Data Science in Geochemistry and Cosmochemistry is becoming increasingly important and relevant to handle and work with the exponentially growing number of geo- and cosmochemical data over the last few decades. Exploring large databases provides the opportunity to research huge datasets to better understand the Earth's dynamics (Jiao et al. 2018). In the last 20-30 years, geochemical data and standards have been collected in databases such as GeoROC, EarthChem, MetBase, or AusGeochem.

The objective of this project is to create methods to work with databases such as, organising and cleaning of data, visualising data on different kinds of plots, statistical analyses of the geochemical data, detection and checking the influence of outliers, integrate these with existing tools such as pyrolite, IsoplotR, GeoPyTool and implement these tools in a modular, open, and web-based research environment. The overall goal is to combine geochemical data and software, which will enable new routes of research and possibly a new area of original scientific discoveries (Ma 2023).

Methods: The web application was developed using python and deployed through streamlit. The python source codes are publicly available on GitHub.

Results and Discussion: We developed a web page with which geochemical database data can be visualized on a map or various types of plots such as scatter plots (bivariate plots), category plots which show the normalised concentrations with respect to e.g., chondrite, or primitive mantle versus a list of user-defined elements and ternary plots. For this we used GEOROC data as an example. The entire database was initially organized by choosing a subset of required elements needed to complete the various web application components. The nan values were replaced by 0 and the columns with no information were dropped. Boxplots were used to identify the outliers. Those outliers, which might be a result of typing errors, were dropped.

Jiao et al. (2018): Progress and challenges of big data research on petrology and geochemistry. - Solid Earth Sci 3, 105-114

Ma X (2023): Data science for geoscience: Recent progress and future trends from the perspective of a data life cycle. - Geol Soc Amer, Special Paper 558