

Screening Method for Rare Earth Element Data - Automated Geochemical Data Assessment

DAVID M. ERNST¹, KERSTIN A. LEHNERT², GERHARD
WÖRNER³, MARIE K. TRAUN³ AND MALTE MUES⁴

¹CritMET - Critical Metals for Enabling Technologies, School of
Science, Constructor University Bremen

²Lamont-Doherty Earth Observatory, Columbia University

³Georg-August-University Göttingen

⁴Chair for Software Engineering, TU Dortmund University

The amount of published geochemical data is continuously increasing due to the availability of advanced and highly efficient analytical tools. These data are made accessible in curated, domain-specific, and interconnected synthesis databases. These databases present great potential for leading-edge research in geochemistry and the application of compositional data of rocks, minerals, and other natural materials in neighbouring fields. However, using data compiled from the literature over many years also introduces pitfalls and uncertainties for data analysis. While the authors of the original data usually know their significance and limitations, “external” data users often find it challenging to assess the reliability of data, inherent uncertainties, and possibly, analytical artefacts.

The group of Rare Earth Elements (REE) is widely used in geochemical research as tracers of chemical transport and differentiation, as well as geochemical fingerprinting of petrological and Earth system processes in general. Therefore, we developed a method to automatically screen geochemical data for noisy, conspicuous, and potentially erroneous REE patterns and anomalies. Our approach is based on the knowledge of characteristic REE behaviour in geochemical processes. Specifically, we address the “smoothness” of normalized REE patterns, identify single (and multiple) elemental outliers in REE plots, and various types of scatter.

We extensively tested our screening method on large data sets extracted from the geochemical databases of GEOROC and PetDB using Monte Carlo simulations and defined deviation scores from expected REE behaviour. Distinct types of anomalous REE patterns are related to potential artefacts and data quality issues. Additionally, we test the application of our REE screening method on individual published data sets from various sources.