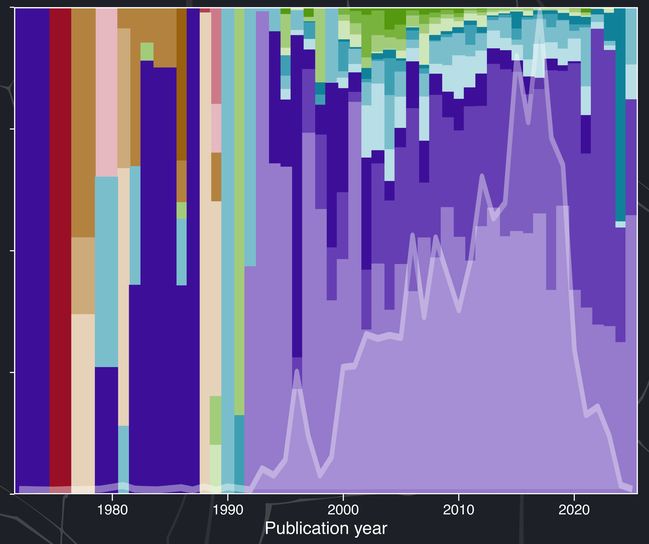
How comparable are geochemical datasets really? and why it matters!



The relative proportion of the most widely used methods to measure trace element concentrations (n = 268447) by publication year. Data source: GeoReM.

Marie Katrine Traun*, A. Renno, L. Kallas, M. Willbold, T. Waight, D. Garbe-Schönberg & G. Wörner

*marie.traun@uni-goettingen.de

Abstract:

Datasets must be comparable to be compilable

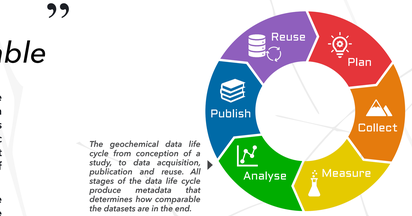
Over the past decades, data volumes and big data applications in the natural sciences have grown at exponential rates. Research based on analysis of large geochemical datasets to understand Earth processes has increased with data volume growth, driving several scientific breakthroughs and paradigm shifts in geochemistry in recent decades. Such studies are in general based on large compilations of smaller geochemical datasets from local or regional studies.

To truly carry out meaningful big data studies in geochemistry, the smaller datasets must be fundamentally comparable in order to be compilable. Assessing the comparability of two or more datasets in order to compile them is no simple task. A single sample typically goes through a myriad of ever evolving processing as well as analytical procedures and methodological steps before eventually reaching a numeric value for an element concentration, isotope ratio et cetera. Each geochemical value produced should ideally be accompanied by comprehensive metadata describing its acquisition from field to file, in other words the long-tail of metadata it acquired through its data life cycle journey.

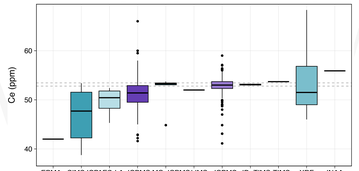
Examples of relevant metadata include: Citation and bibliographical information; location information; sample type and acquisition; geological classification including lithology and mineralogy; stratigraphy; geochronology; sample processing methods; analysis techniques; analytes and units; uncertainties; reproducibility; and calibration.

The metadata diversity associated with a single data point is remarkable and in particular illustrates why external reproducibility is key to ensure comparability between studies. Reproducibility is strongly associated with the reporting of reference materials. Reference materials are used in the calibration procedure of an instrument and normalization of a result to a shared baseline.

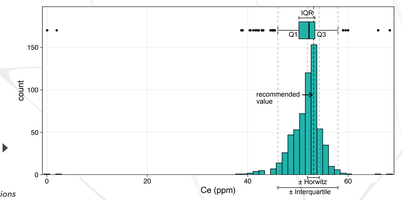
Here we assess reference material offsets between geochemical datasets (ie. The inter-study bias) using the Geological and Environmental Reference Material database – GeoReM.



The geochemical data life cycle from conception of a study, to data acquisition, publication and reuse. All stages of the data life cycle produce metadata that determines how comparable the datasets are in the end.



Method bias of Ce concentrations (n = 621) measured in reference material BCR-2 for different methods. The dashed grey lines indicate the 95% confidence level interval of the reference concentration value for BCR-2. Data source: GeoReM.



Histogram and boxplot of the distribution of BCR-2 reference material values for Ce concentration. The quantiles Q1 (25%) and Q3 (75%) and the interquartile range (IQR) are marked above. The gray dashed lines indicate the accepted range limits for interquartile outlier criteria and the limits of the Horwitz outlier criteria. Data source: GeoReM.

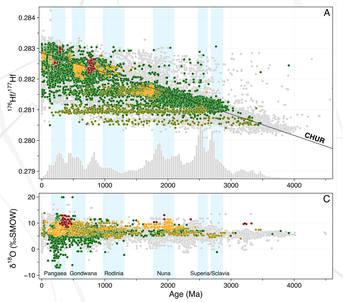
Unidentified bias of unknown impact

The use of the GeoReM based data quality tools is showcased by assessing the presence of inter-study bias and its impact on geological interpretations for two famous geochemical big data research topics: 1) The Mantle Zoo end-members of oceanic basalts and 2) signatures of continental crust evolution in zircons. The inter-study bias offset of each data point in each publication can be visually quantified by tallying up the true/false flags from the outlier tools into an overall offset score. Based on the offset scores, we get a top-down overview of the metadata and inter-study bias offsets and how it impacts the data sample trends.

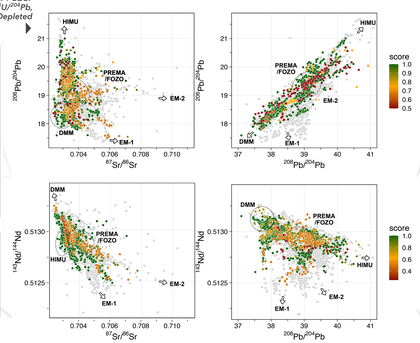
For the Mantle Zoo for example, there is a noticeable lack of metadata quality information available in the database for the EM-1. We also observe that the FOZO/PREMA end-member zone has a larger proportion of data that did not pass the reference material outlier checks for Pb-isotopes. Much emphasis is placed on the interpretation of the end-member compositions and how they reflect different recycled mantle components and thus the comparability of underlying compiled data, on which these interpretations rely on, is critical for the Mantle Zoo paradigm.

For the zircons through time example, the GEOROC zircon age peaks align well with intervals of supercontinent formation. In the older range > 3 Ga, there is notably less metadata coverage in the database for the reference material values for both $^{176}Hf/^{177}Hf$ and $\delta^{18}O$. Comprehensive quality and metadata information is critical for Early Earth studies, where access to ancient materials is limited, geographically biased towards few surviving Archean provinces and the timing of geological events is especially challenging to constrain. Low scores coincide with high $\delta^{18}O$ values at around 400 Ma, 1.9 Ga, 2.4 Ga and 3.3 Ga. This would potentially have significant implications for interpretations determining the onset of crustal reworking and recycling during the Precambrian.

Compositions of zircon spots through time (U-Pb ages) coloured by offset score. The histogram shows the distribution of zircon ages weighted linearly by age. The time intervals of the last supercontinent phases are highlighted. Data source: GEOROC.



"Classic" Mantle Zoo plots for oceanic basalts coloured by the offset score, defined as the proportion that passed our outlier checks. FOZO/PREMA: Prevalent mantle, HIMU: High $^{238}U/^{204}Pb$, EM-1 and EM-2: enriched mantle, DMM: Depleted MORB mantle. Data source: GEOROC.



GeoReM-based quality tools

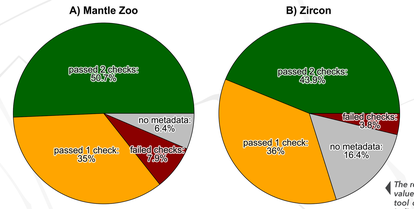
To assess the comparability, compilability and inter-study biases in geochemical compilations, we present three tools integrated into the GeoReM database.

The first tool allows users to assess if a measured value for a reference material is within the expected distribution of values from GeoReM, according to the interquartile outlier criteria. This tool is equivalent of the classical boxplot graph, where the outliers are represented as points beyond the "boxplot whiskers". The interquartile outlier check returns a boolean (true/false) flag: true if the value lies within the range and false if the value falls beyond this range. If the tool cannot find comparison values in GeoReM, either because a parameter name is spelled wrong or data matching the input does not exist, it returns a null. The interquartile range-based tool has the advantage of being more stable to uneven and non-normal data distributions.

The second tool allows users to check if a measured value for a reference material is within range of the recommended value available in GeoReM, based the Horwitz outlier criteria [1]. It determines if a value falls within the estimated uncertainty range of a recommended reference value. It functions by estimating the theoretical analytical standard deviation of a reference value based on a κ-factor ranging from 0.01 to 0.02. The best and most recent recommendation level value available is looked up dynamically in GeoReM. If there are multiple high-accuracy recommended values from the same most recent publication year, the average of these is used.

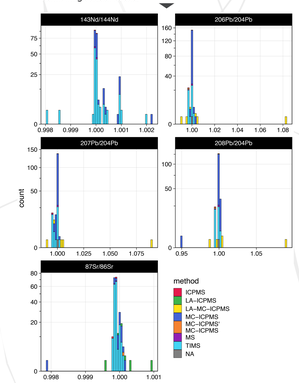
The third tool is a data quality report tool, which dynamically calculates and returns the inter-study bias as an absolute and relative offset of reference material values to the best recommended value in GeoReM for user-provided paper identifiers like DOIs. It returns a table with the relative and absolute offsets from the certified or recommended reference material values for each analyte, unit, and reference material in each publication, including associated method metadata. If the paper DOI does not exist in GeoReM, the tool returns a blank row for that publication id.

[1] Horwitz, W., & Albert, R. (1995). Precision in analytical measurements: Expected values and consequences in geochemical analyses. Fresenius' Journal of Analytical Chemistry, 351(6), 507–513.

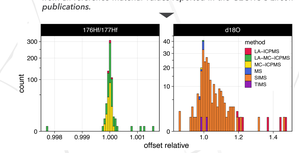


The result of the quality outlier checks for two use cases. Red indicates reference material values that are certain outliers. Green indicates the fraction of values that pass both outlier tool check, and orange indicates the fraction that only passed one outlier check. Gray indicates where reference materials were either not reported or not included in GeoReM.

Histograms showing the relative offsets from our quality report tool (measured value/recommended value) for the radiogenic isotope analytes from all reference materials reported in the Mantle Zoo publications. The y-axis is square root transformed to make the low count range easier to see.



Histograms showing the relative offsets from our quality report tool (measured value/recommended value) for $^{176}Hf/^{177}Hf$ and $\delta^{18}O$ from all reference material values reported in the GEOROC zircon publications.

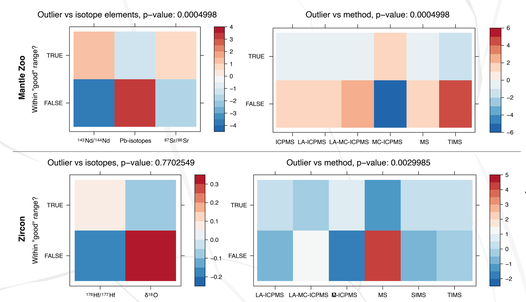


Test your own compilations

Our tools are made available to anyone for free through the newly developed GeoReM API. How to use the tools through the API is demonstrated in Python Jupyter Notebooks and R Notebooks provided in the accompanying GitLab repository of this work (QR code).

The key advantage of these tools is the direct integration within GeoReM, allowing dynamic data fetching and calculation of outlier criteria for the reference material values and best available recommended values. These tools provide a robust knowledge foundation for users who wish to conduct statistical analyses of inter-study bias. Statistical evaluation on the on our use-cases here indicate that inter-study bias does not arise from a single cause in our use case examples. Inter-study bias arises from a complex interplay of factors, including analytical method bias, laboratory bias and the used reference materials, and is strongly dependent on the measured analyte. The inter-study bias must therefore be assessed at the level of individual studies, compilations and analytes.

Assessing inter-study bias is fundamentally constrained by the availability and publication of geoanalytical metadata and reference material. Metadata and reference materials are often not reported, only partially disclosed, or when included, not provided in a machine-readable format in the publications. In these cases, the comparability and compilability of datasets cannot reliably be evaluated, representing a major limitation for the advancement of 'big data' applications in geochemistry.



Contingency tables showing the correlation between interquartile outliers, analytes and methods. The colours represent the Pearson residuals, showing whether co-occurrence of an outlier and analyte/method is proportionally higher (red) or lower (blue) than expected.

We perform a Fisher Exact Test in the R programming language to get a p-value indicating whether outliers and analytes/methods are significantly dependent (p-value < 0.05) or not.

Try our new GeoReM API quality tool endpoints and demo notebook (R/python) on our NFDI4Earth's project GitLab!

Get an API key from: digis-info@uni-goettingen.de

GEOROC
Geochemistry of Rocks of the Oceans and Continents

GeoReM
Geological and Environmental Reference Materials

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
IN PUBLICA COMMODA SEIT 1737