

pyrolite: Python for geochemistry

Morgan J. Williams¹, Louise Schoneveld¹, Yajing Mao², Jens Klump¹, Justin Gosses³, Hayden Dalton⁴, Adam Bath¹, and Steve Barnes¹

¹ CSIRO Mineral Resources ² Institute of Geology and Geophysics, Chinese Academy of Geosciences ³ NASA Johnson Space Center ⁴ School of Earth Science, University of Melbourne

DOI: [10.21105/joss.02314](https://doi.org/10.21105/joss.02314)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Arfon Smith](#) ↗

Reviewers:

- [@arfon](#)

Submitted: 08 June 2020

Published: 09 June 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

`pyrolite` is a Python package for working with multivariate geochemical data, with a particular focus on rock and mineral chemistry. The project aims to contribute to more robust, efficient and reproducible data-driven geochemical research.

Features

`pyrolite` provides tools for processing, transforming and visualising geochemical data from common tabular formats. The package includes methods to recalculate and rescale whole-rock and mineral compositions, perform compositional statistics and create appropriate visualisations and also includes numerous auxiliary utilities (e.g. a geological timescale). In addition, these tools provide a foundation for preparing data for subsequent machine learning applications using `scikit-learn` (Pedregosa et al., 2011).

Geochemical data are compositional (i.e. sum to 100%), and as such require non-standard statistical treatment (Aitchison, 1984). While challenges of compositional data have long been acknowledged (e.g. Pearson, 1897), appropriate measures to account for this have thus far seen limited uptake by the geochemistry community. The submodule `pyrolite.comp` provides access to methods for transforming compositional data, facilitating more robust statistical practices.

A variety of standard diagram methods (e.g. ternary, spider, and data-density diagrams; see Figs. 1, 2), templated diagrams (e.g. the Total-Alkali Silica diagram, Le Bas, Le Maitre, & Woolley, 1992; and Pearce diagrams, Pearce, 2008) and novel geochemical visualisation methods are available. The need to visualise geochemical data (typically graphically represented as bivariate and ternary diagrams) has historically limited the use of multivariate measures in geochemical research. Together with the methods for compositional data and utilities for dimensional reduction via `scikit-learn`, `pyrolite` eases some of these difficulties and encourages users to make the most of their data dimensionality. Further, the data-density and histogram-based methods are particularly useful for working with steadily growing volumes of geochemical data, as they reduce the impact of 'overplotting'.

Reference datasets of compositional reservoirs (e.g. CI-Chondrite, Bulk Silicate Earth, Mid-Ocean Ridge Basalt) and a number of rock-forming mineral endmembers are installed with `pyrolite`. The first of these enables normalisation of composition to investigate relative geochemical patterns, and the second facilitates mineral endmember recalculation and normative calculations.

`pyrolite` also includes some specific methods to model geochemical patterns, such as the lattice strain model for trace element partitioning of Blundy & Wood (2003), the Sulfur Content at Sulfur Saturation (SCSS) model of Li & Ripley (2009), and orthogonal polynomial decomposition for parameterising Rare Earth Element patterns of O'Neill (2016).

Extensions beyond the core functionality are also being developed, including `pyrolite-melts` `util` which provides utilities for working with `alphaMELTS` and its outputs (Smith & Asimow,

2005), and is targeted towards performing large numbers of related melting and fractionation experiments.

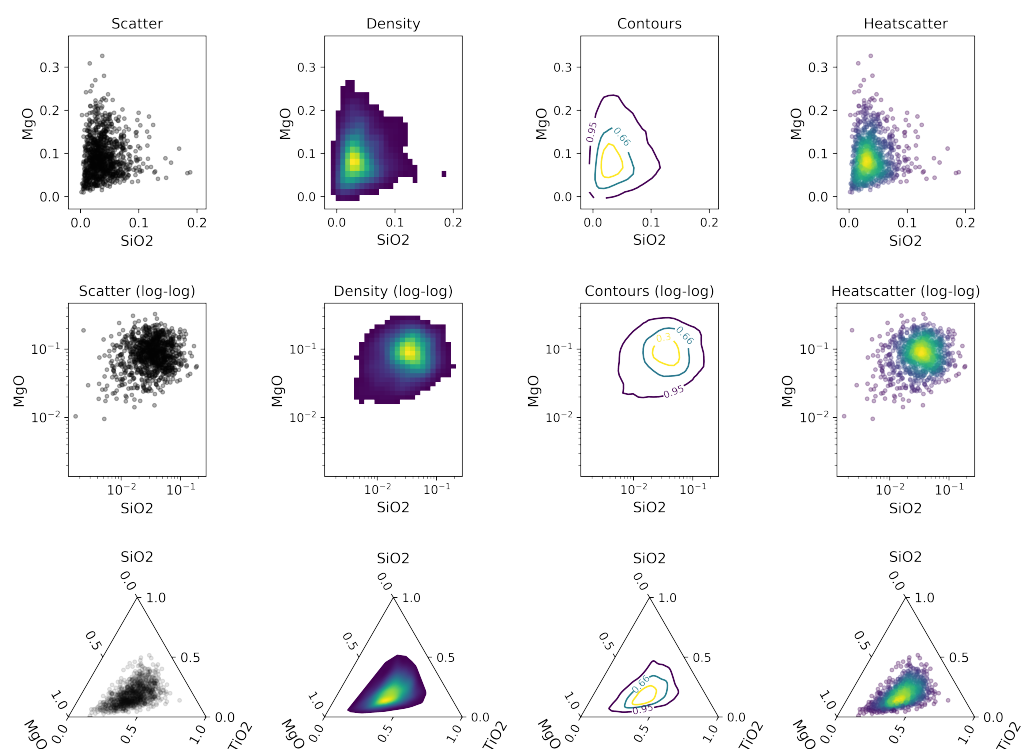


Figure 1: Example of different bivariate and ternary diagrams, highlighting the ability to visualise data distribution.

API

The `pyrolite` API follows and builds upon a number of existing packages, and where relevant exposes their API, particularly for `matplotlib` (Hunter, 2007) and `pandas` (McKinney, 2010). In particular, the API makes use of dataframe accessor classes provided by `pandas` to add additional dataframe 'namespaces' (e.g. accessing the `pyrolite` `spiderplot` method via `df.pyroplot.spider()`). This approach allows `pyrolite` to use more familiar syntax, helping geochemists new to Python to hit the ground running, and encouraging development of transferable knowledge and skills.

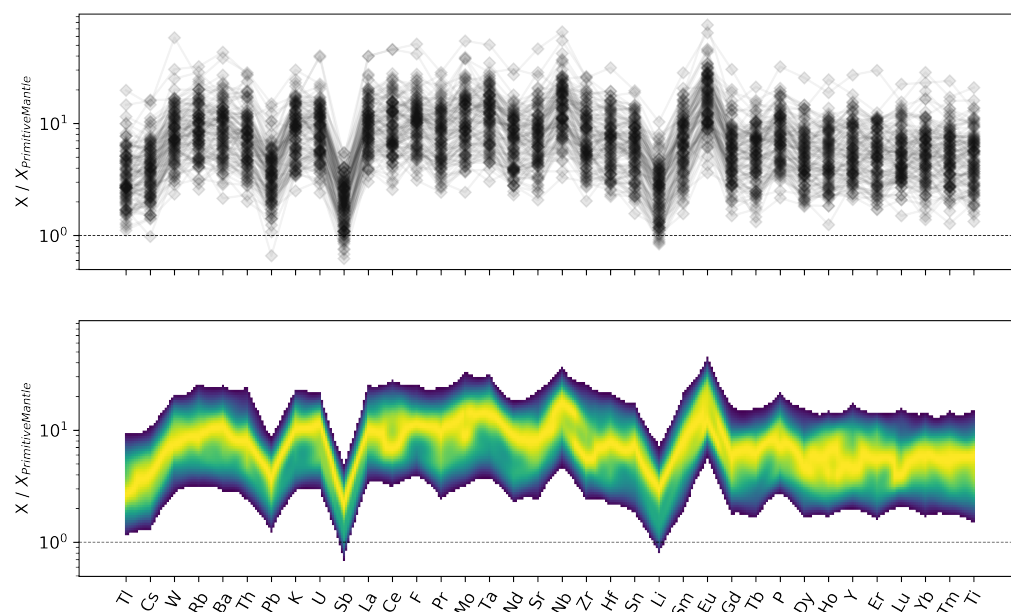


Figure 2: Standard and density-mode spider diagrams generated from a synthetic dataset centred around an Enriched- Mid-Ocean Ridge Basalt composition (Sun & McDonough, 1989), normalised to Primitive Mantle (Palme & O'Neill, 2014). Elements are ordered based on a proxy for trace element 'incompatibility' during mantle melting (e.g. as used by Hofmann, 2014).

Conventions

Tidy Geochemical Tables

Being based on `pandas`, `pyrolite` operations are based on tabular structured data in dataframes, where each geochemical variable or component is a column, and each observation is a row (consistent with “tidy data” principles, Wickham, 2014). `pyrolite` additionally assumes that geochemical components are identifiable with either element- or oxide-based column names (which contain only one element excluding oxygen, e.g. *Ca*, *MgO*, *Al₂O₃*, but not *Ca₃Al₃(SiO₄)₃* or *Ti_ppm*).

Open to Oxygen

Geochemical calculations in `pyrolite` conserve mass for all elements excluding oxygen (which for most geological scenarios is typically in abundance). This convention is equivalent to assuming that the system is open to oxygen, and saves accounting for a ‘free oxygen’ phase (which would not appear in a typical subsurface environment).

Community

`pyrolite` aims to be designed, developed and supported by the geochemistry community. Community contributions are encouraged, and will help make `pyrolite` a broadly useful toolkit and resource (for both research and education purposes). In addition to developing a library of commonly used methods and diagram templates, these contributions will contribute to enabling better research practices, and potentially even establishing standards for geochemical data processing and analysis within the user community.

Acknowledgements

The authors of this publication (listed in reverse alphabetical order) have all contributed to the development of `pyrolite` at an early stage.

References

- Aitchison, J. (1984). The statistical analysis of geochemical compositions. *Journal of the International Association for Mathematical Geology*, 16(6), 531–564. doi:[10.1007/BF01029316](https://doi.org/10.1007/BF01029316)
- Blundy, J., & Wood, B. (2003). Partitioning of trace elements between crystals and melts. *Earth and Planetary Science Letters*, 210(3), 383–397. doi:[10.1016/S0012-821X\(03\)00129-8](https://doi.org/10.1016/S0012-821X(03)00129-8)
- Hofmann, A. W. (2014). 3.3 - Sampling Mantle Heterogeneity through Oceanic Basalts: Isotopes and Trace Elements. In H. D. Holland & K. K. Turekian (Eds.), *Treatise on Geochemistry (Second Edition)* (pp. 67–101). Oxford: Elsevier. doi:[10.1016/B978-0-08-095975-7.00203-5](https://doi.org/10.1016/B978-0-08-095975-7.00203-5)
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95. doi:[10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Le Bas, M. J., Le Maitre, R. W., & Woolley, A. R. (1992). The construction of the Total Alkali-Silica chemical classification of volcanic rocks. *Mineralogy and Petrology*, 46(1), 1–22. doi:[10.1007/BF01160698](https://doi.org/10.1007/BF01160698)
- Li, C., & Ripley, E. M. (2009). Sulfur Contents at Sulfide-Liquid or Anhydrite Saturation in Silicate Melts: Empirical Equations and Example Applications. *Economic Geology*, 104(3), 405–412. doi:[10.2113/gsecongeo.104.3.405](https://doi.org/10.2113/gsecongeo.104.3.405)
- McKinney, W. (2010). Data structures for statistical computing in python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (pp. 51–56).
- O'Neill, H. S. C. (2016). The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts. *Journal of Petrology*, 57(8), 1463–1508. doi:[10.1093/petrology/egw047](https://doi.org/10.1093/petrology/egw047)
- Palme, H., & O'Neill, H. S. C. (2014). Cosmochemical Estimates of Mantle Composition. (H. D. Holland & K. K. Turekian, Eds.) *Treatise on Geochemistry (Second Edition)*, 3, 1–39. doi:[10.1016/B978-0-08-095975-7.00201-1](https://doi.org/10.1016/B978-0-08-095975-7.00201-1)
- Pearce, J. A. (2008). Geochemical fingerprinting of oceanic basalts with applications to ophiolite classification and the search for Archean oceanic crust. *Lithos*, 100(1), 14–48. doi:[10.1016/j.lithos.2007.06.016](https://doi.org/10.1016/j.lithos.2007.06.016)
- Pearson, K. (1897). Mathematical contributions to the theory of evolution. On A form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, 60(359-367), 489–498. doi:[10.1098/rspl.1896.0076](https://doi.org/10.1098/rspl.1896.0076)
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Smith, P. M., & Asimow, P. D. (2005). `Adiabat_1ph`: A new public front-end to the MELTS, pMELTS, and pHMELTS models. *Geochemistry, Geophysics, Geosystems*, 6(2). doi:[10.1029/2004GC000816](https://doi.org/10.1029/2004GC000816)

- Sun, S.-s., & McDonough, W. F. (1989). Chemical and isotopic systematics of oceanic basalts: Implications for mantle composition and processes. *Geological Society, London, Special Publications*, 42(1), 313–345. doi:[10.1144/GSL.SP.1989.042.01.19](https://doi.org/10.1144/GSL.SP.1989.042.01.19)
- Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(1), 1–23. doi:[10.18637/jss.v059.i10](https://doi.org/10.18637/jss.v059.i10)