Geochronology Discuss., referee comment RC1
https://doi.org/10.5194/gchron-2022-4-RC1, 2022

A Solid Foundation for Accessible Best Practice in SIMS U-Pb Geochronology Data Reduction

Morgan Williams (Referee)

Referee comment on "An algorithm for U–Pb geochronology by secondary ion mass spectrometry" by Pieter Vermeesch, Geochronology Discuss., https://doi.org/10.5194/gchron-2022-4-RC1, 2022

This manuscript provides a well-written and logically-organised description of some of the key challenges in addressing data reduction of SIMS U-Pb geochronology, and succinctly describes best-practice approaches to the required processing steps. The author clearly identifies the main limitations and inaccuracies of commonly used approaches to dealing with mass spectrometric data in compositional systems, which as noted by the author have been acknowledged in the community for some time. The documented approach builds upon other recent work in the community encouraging the use of appropriate compositional data handling, and the preservation of full covariance matrix throughout the data reduction process where possible. The software accompanying this manuscript (*simplex*) provides a flexible interface (with online and offline in-browser applications and access to a command line interface) and supports the key data formats for SIMS geochronological data from SHRIMP and Cameca instruments. The author is to be commended for providing the software in an open-source manner with a permissive-use license, and for ensuring the approach is accessible via multiple interfaces (allowing both high-level and low-level interfaces catering to the broader community). Some comments regarding the manuscript and associated software are attached below, with some specific technical suggestions for the manuscript and software noted separately.

- The implied extension of other software (e.g. isoplotr, in the first instance) to accept the fully-specified outputs of *simplex* will be a key next step to make best use of this work and bridge the full pipeline from instrument outputs to research outputs.
- The key concepts are of general interest beyond the geochronology community, with direct application to other SIMS-based isotope ratio measurements. In encompassing the major steps in data reduction of compositional mass spectrometric data, the software is flexible enough to reduce e.g. Cameca oxygen and sulfur isotope data (as is demonstrated in the application). In some cases, additional steps might be appropriate for specific scenarios (e.g. dealing with electron-induced secondary ion emission for negative-ion measurements on SHRIMP instruments where an electron gun is used for charge compensation in insulating materials, Ickert et al. 2008), but the foundational steps of the workflow are the same.

- Regarding the operation of the software, I can verify that online application interface generally works as expected and the provided installation instructions (in the GitHub Readme) were sufficient guidance for getting the package running locally (at least for those who either have or are comfortable setting up an R distribution). I've been able to verify the use of the software across most of the claimed platforms (with a few minor bugs, described in the attached document) – online, on both Windows and Linux (via Windows Subsystem for Linux) using an Anaconda R distribution. The references to *simplex* outputs in the manuscript corresponding to the demonstration datasets (e.g. Figure 6b, Figure 7) match the current outputs of the software. With appropriate adjustments, I was able to process SHRIMP U-Pb data (both .op and .pd files, measured on the Australian National University SHRIMP II in 2013) in the same manner as the demonstration datasets.
- The software provides some functionality to enable reproducibility of data reduction even in a graphical interface where this is commonly a challenge, with the ability to save and load configurations for data processing. With future development, any extension to include provenance information would be a valuable addition, potentially via metadata extracted from the input files where possible.
- The documentation for processing options within the in-browser application is simple but explains in general terms the processes and options available, and the interface-specific documentation for the R package itself indeed appears to be complete.
- Regarding propagation of matrix-based uncertainties in geochronological data, McLean et al. (2011) provide a relevant digestible overview (although with a focus on specific considerations for ID-TIMS) which may be a useful reference to cite in text or in the appendix, at the author's discretion.

Manuscript: Technical Corrections and Specific Suggestions

- L9: typo "handes" should be "handles".
- L19: The series of short sentences here are succinct but do not read well; rephrasing along the lines of 'In contrast to SIMS instrumentation, LAICPMS instruments are built by numerous manufacturers, and the widely used data reduction software packages are generally compatible with all of their data formats".
- L122, L126: Specifically refer to the additive log ratio (ALR) used.
- L142: The term 'trick' here perhaps confounds the intention and doesn't lend to confidence in the use of it. When referring to generalization, perhaps refer to Eqn. 4 (which is what is generalized, as the logratio approach is not specific to a dimensionality, even if examples are 2/3 dimensional).
- Figure 2a – I think the labels on the upper figure may be reversed. This might also be better termed a 'projection' from a four-component simplex (which would be a tetrahedron). This diagram depicts the compositional data aspect of the manuscript and software rather nicely.
- Eqn 13, Section 9: Dead-time correction – I think this exclusively applies to a non-extended deadtime correction? Generally, the difference between extending/non-extending is likely to be small except for particularly high count-rates (maybe for every high U, or analysing a minor element on an EM), but it might be worth noting this, at the author's discretion.
- L224: SHRIMP data processing – Depending on the scenario, I think Dodson interpolation may also be used to similar effect for dealing with smooth intensity changes during cycles of data collection within a single spot (Dodson, 1978).
- Figure 5: Depending on the desired manuscript formatting, the capitalization of this

figure caption may be off (e.g. "a) Blank …")?
- L321: "Figure 6.a" – the references to figures may be easier to read without the period (e.g. Figure 6a; as is used in the Figure 7 caption).
- Fig7b: I may be worth nothing that the white ellipse represents the concordia age and uncertainty.
- References: Add DOIs to references where possible.

Software and Repository: General Notes

- The license for the software is specified in the GitHub readme and R package description (the GPL-3 copyleft license), but is not in the repository; generally, a copy of this license should be included with the package repository (for visibility, and to meet the terms of the license).
- I would suggest releasing a specific software version corresponding to the publication (e.g. 1.0, or 0.1 for a 'beta' release) such that future users can readily recognise updated versions relative to that described here (particularly if this manuscript is intended to be the primary/secondary citation for *simplex*, rather than citing the software package itself).
- In the browser-based application, plotting and other functions generally work as expected, and as in the hosted online version (as far as tested, a few specific bugs/issues/suggested features are noted below).
- The selenium tests included in the repository include reference to local files which are not available; these may be able to be modified to be distributable (potentially helpful for debugging).
- Unit tests for the R package itself may be a useful addition for future development.
- To get the app to run via WSL, manual specification of the port is necessarily, as the browser can't necessarily be launched directly (I believe the specifics of this are managed by *shinylight*; this would not be a problem on a standalone Linux distribution, but may be worth documenting).
- All in-app plotting suffered from a Cairo-SVG related bug on Windows ('svg: Cairo-based devices are not available for this platform'). This may potentially be rectified with an additional dependency, or potentially a change in the plotting output format.
- During my tests in a base-R distribution (i.e. not from RStudio), remotes-based installation in from the command line requires the setting of an environment variable (related to unzipping the repository contents).

Software and Repository: Browser-based Use

- Suggested Feature: For tab 5 (samples), it would be good to exclude standards by default; and have multiple comma-separated patterns for sample matching.
- Suggested Feature: Add a CSV download option for tables (these can currently be copy-pasted into common spreadsheet programs, without headers/indexes).
- Suggested Feature: Potentially preserve source file name on JSON export for some simple provenance metadata. The source files contain more of this metadata themselves, and this could be a feature to consider for the future (e.g. with XML import for SHRIMP data, which is suggested to be planned for future development).

- Suggested Feature: For relevant stages, automatic refreshing for plots when the analysis # is changed would be useful.
- Bug (tab 5, using tabulate): adding more than one option to the table gives notation for duplicated column names (e.g. A, B, C, A.1, B.1, C.1, A.2, …). In particular, the structure of the 'cov' option with vector values indexed by integers and sparse covariance matrix across all samples is not immediately clear in the tabulated output. After a second look this table is readily understood, but would ideally it would be indexed by appropriate and unique column names.

Software and Repository: Programmatic Use/CLI

- Plotting via the programmatic interface did not exhibit any issues on Windows, in contrast to the browser-based app.
- The example CLI workflow in the GitHub Readme gives errors at the second step, after modification to use the included `./inst/SHRIMP.op` file; using the included 'SHRIMP_UPb' dataset (not from an original file) enabled the use of this workflow as below:

```
lr <- logratios(SHRIMP_UPb)
stand <- standard(preset="Temora-t")
paired <- pairing(lr,stand=stand
cal <- calibration(lr,stand=stand,pairing=paired,prefix="TEM")
result <- calibrate(cal,exterr=TRUE)
```

References

Dodson, M. H. (1978). A linear method for second-degree interpolation in cyclical data collection. Journal of Physics E: Scientific Instruments, 11(4), 296. https://doi.org/10.1088/0022-3735/11/4/004

Ickert, R. B., Hiess, J., Williams, I. S., Holden, P., Ireland, T. R., Lanc, P., Schram, N., Foster, J. J., & Clement, S. W. (2008). Determining high precision, in situ, oxygen isotope ratios with a SHRIMP II: Analyses of MPI-DING silicate-glass reference materials and zircon from contrasting granites. Chemical Geology, 257(1–2), 114–128. https://doi.org/10.1016/j.chemgeo.2008.08.024

McLean, N. M., Bowring, J. F., & Bowring, S. A. (2011). An algorithm for U-Pb isotope dilution data reduction and uncertainty propagation. Geochemistry, Geophysics, Geosystems, 12(6). https://doi.org/10.1029/2010GC003478