Geochronology Discuss., referee comment RC1
https://doi.org/10.5194/gchron-2022-20-RC1, 2022

Comment on gchron-2022-20

Anonymous Referee #1

Referee comment on "Examination of the accuracy of SHRIMP U–Pb geochronology based on samples dated by both SHRIMP and CA-TIMS" by Charles W. Magee Jr. et al., Geochronology Discuss., https://doi.org/10.5194/gchron-2022-20-RC1, 2022

This study reviews published and to lesser extent new U-Pb zircon geochronology data generated by two methods: secondary ionization mass spectrometry (SIMS) and thermal ionization mass spectrometry (TIMS) without and with pre-treatment of zircon by chemical abrasion (CA), respectively. The SIMS analyses were carried out over approximately the past 15 years using the SHRIMP II instrument at Geoscience Australia, and they targeted felsic plutonic and volcanic rocks with ages mostly falling into the Permo-Triassic to Cambrian age range. The exceptions are one Cretaceous “calcilutite” and Archean secondary reference zircon OG1, which was analyzed in most of the sessions to monitor Pb-isotopic fractionation. The goal of the comparison is to re-assess the reproducibility of SHRIMP U-Pb geochronology considering updated and improved methodology implemented over the past two decades. This is done in reference to CA-TIMS ages obtained for the same samples with an uncertainty that is considered negligible compared to that of SHRIMP.

This sample-based comparison is for an age range and geological provenance of zircon that is typical for the Geoscience Australia lab: the samples fall into an age range, where uncertainties for Pb-Pb ages are typically larger than those of U-Pb dating. Therefore, uncertainties are dominated by how well the U-Pb calibration curve can be defined and reproduced. The materials are mostly igneous rocks with seemingly “simple” crystallization histories (e.g., compared to metamorphic zircon). In comparing both methods, this includes zircon where the penalty of TIMS with its indiscriminate averaging over multiple growth domains in individual zircon is minor, and seemingly negligible compared to analytical uncertainties. The CA pre-treatment selectively removes zircon, but this is commonly regarded as non-detrimental in obtaining the original crystallization age of zircon.

The approach of comparing zircon ages generated by different methods from such materials, however, necessarily has limitations due to a lack of knowledge and temporal resolution regarding the age homogeneity of the samples. Although some aspect of zircon longevity in magma systems are probably indeed negligible, there are other concerns for zircon age heterogeneity even at the limit of TIMS resolution, that are less clear-cut to dismiss. Only with sufficient sampling, which is however rarely achieved in CA-TIMS studies, might this become adequately constrained. Another critical aspect is that CA-untreated and CA-treated zircon is compared; this – at least for zircon with younger SHRIMP relative to CA-TIMS ages – leaves some doubts whether indeed the same materials are compared in case SHRIMP ages would overlap on zircon areas affected by Pb-loss that would have otherwise been removed by CA. What the manuscript is also missing is a presentation and discussion of U abundances in the analyzed zircons, as this would control potential metamictization, and there are also documented matrix effects for high-U zircon in SIMS analysis.

Considering these problems and omissions, three critical aspects stand out, where I find that the interpretation overstepped what can be extracted from the data:

- There is speculation that the distribution of the age differences between SHRIMP and TIMS ages is bimodal. I doubt this, and would argue that this apparent bimodality is an artifact of the small sampling size. As a demonstration, a simple test was made using the Excel Rand function 35-times and normalizing the results to a mean of 0 and a standard deviation (using the NORM-INV function). It is very easy to generate an apparent bimodal distribution with a difference in modes at approximately the same value stated for the difference between volcanic and plutonic samples (0.7%):

  Moreover, a clear identification of a bimodal distribution requires differences between the modes of 3–5 times the standard error (Keller et al. 2018). This is not the case in a distribution where the two modes only differ by 0.7% if the SHRIMP uncertainty was 1%; the standard error of each subpopulation would then be between 0.2 and 0.3% (because of division by square-root of n). Applying the conservative criterion of 4-times the standard error, the difference between the two modes should be at least 0.8 to 1.2%, which is larger than the postulated difference. I therefore do not believe that it is statistically justified to discriminate between plutonic and volcanic samples regarding the apparent deviations between ages determined by different methods; if the authors think that the bimodality in the data is robust, it is their onus to provide an adequate statistical analysis to demonstrate the validity of their assessment.

- There is a risk of circular argument in the data analysis in that the authors aim to realistically constrain the age uncertainty of SHRIMP (which is a priori unknown), but then they use the apparent overdispersion of the data to identify non-analytical causes for the overdispersion, whereas in fact, an underestimation of the uncertainty could produce the same overdispersion. Therefore, one cannot make the unequivocal statement that the distribution of the age differences is non-Gaussian, as this may be biased by a potential underestimation of the actual uncertainties. An important aspect here is scrutiny about the data in this comparison. Although some of the double-dated samples in the database are based on a comparatively large number of analyzed zircons (both, for CA-TIMS and SHRIMP results), there are also samples with severe limitations in one or both data sets (e.g., 1528025, 1594761, 1954030 with only three zircon analyses in the TIMS-group, and often 2 to 3 excluded grains, as well as 1978295 and 1978296 with only 6 or 7 zircons analyzed by SHRIMP, again with a comparatively large number of data excluded). The selection of data to be included/excluded may play an important role (i.e., the sample with the largest deviation has the least SHRIMP data points, and many excluded analyses). It would be reasonable to restrict the comparison to those samples with a robust number of data in each population. In this same context, I think the regression in Fig. 3 is dubious; it is not a valid fit (see P value), and it may be biased by a few data points of questionable quality. Especially as the authors make a point that there is no linear bias in the age difference vs. age, I find this diagram unnecessary and distractive.

- The discussion on potential "geological" causes for the age difference is oversimplified, and it does adequately reflect the current research status on how and to what extent "real" age heterogeneity in zircon is generated (e.g., Burgess et al., 2019, for a well-documented case study for tephras). Studies of young (e.g., Quaternary) systems show age variability within a single volcanic or plutonic sample of several 100's of ka. This is, however, much less than the percent-level uncertainties of SHRIMP (e.g., 1% for a 400 Ma zircon = 4 Ma) and even less than that of CA-TIMS (e.g., 0.1% uncertainty = 400 ka). Hence, calling for "real age differences in crystallization age" due to "…eruption-aged zircons grains with zircon crystallized earlier in the history of the volcanic edifice" needs to be better justified, and may not advance the understanding on how zircon age variability for comparatively ancient samples arises. There are, however, other ways of generating age heterogeneity in volcanoclastic rocks, for example when ashes are mixed during or after the eruption with materials from other eruptions (please note that this is different from zircon crystallization in the same magma system). In this case, age differences in a single sample may reflect the longevity of a volcanic region, which can span multi-million years. This, however, raises question on the number of crystals that needs to be analyzed for representative sampling (see the discussion on maximum depositional ages from detrital zircon studies). Lastly, I do not understand why plutonic zircon would intrinsically be more vulnerable to Pb-loss than volcanic zircon ("plutonic zircons, which are more likely to have accumulated enough radiation damage to have undergone minor Pb loss")? Due to comparatively slow cooling of plutonic rocks relative to volcanic deposits, one may argue that volcanic zircons spend more time at low temperature where they are vulnerable to accumulate radiation damage. This assumption is admittedly also simplistic, and accumulation of radiation damage will dependent on the exact thermal history of the samples, but it serves to demonstrate that the authors' preference is not at all straightforward. Because of this ambiguity, more explanation is required on what the authors insinuate, especially as U-abundances and structural state of the investigated zircon crystals (e.g., through EBSD, or Raman spectroscopy) are not mentioned or discussed.

The manuscript refers to another study in preparation, where a direct comparison of CA-treated zircon analyzed with both methods, and involving reference zircons will be made. I actually expect more insight on the analytical comparability and the assessment of realistic uncertainties from this future study. Due to the significant uncertainties regarding the homogeneity of the natural samples studied here, especially those of pyroclastic origin, there is little new insight, and in fact, the main conclusion stated in this study is that an uncertainty <0.7 % for SHRIMP data is over-optimistic. Hence, in essence, the 20 year old estimate for SHRIMP reproducibility at ~1% appears to be still valid.

Additional comments:

Line 23: This apparent bimodality needs to be statistically verified.

Line 25: "better single-grain age-resolution of TIMS" = this is a bit awkward to read, as the integration of multiple age domains is the main drawback of TIMS. I also doubt if CA-TIMS can resolve genuine pre-eruptive zircon crystallization in the same magma system (= antecrysts) in the age range presented. Even if it did, what would this mean for dating a geological event such as deposition of a tephra (see Keller et al., 2018)?

Line 298: is not included

Line 310: This section is repetitive and tedious to read; this can be condensed to summarizing the main points in a table. In fact, I think section 4.1 which is presently in the discussion, should be presented as the main result.

Line 402: Pb-loss after 3-5 million years seems highly speculative, and not supported by any experimental data on zircon interaction with fluids. As U abundances are not discussed, there is no way to gauge timescales for metamictization, but this is something that the authors should look into and add to the presentation.

Line 405: Yes, I totally agree that this is not statistically robust.

Line 408: I disagree that the shape of the distribution has been assessed in a statistically robust way.

Line 416. Not sure if "p-hacking" is an adequate term; in any case, it has a negative connotation.

Line 456: Why would this be more likely? There seem to be some underlying assumptions here that should be explicitly stated.

Lien 457: This sentence is awkward: What are natural ages? What are chemically abraded ages?

Line 465: I am not convinced that this is a valid interpretation.

Line 469: "SIMS geochronology is not the best method in geologic settings where grains may have real differences in crystallization age that are smaller than the precision of a single spot, but larger than the precision of the final age of the pooled spot values." I don't agree with this statement, as a bulk method will create artificially small uncertainties for an age that may not have any geological significance (see discussion in Keller et al., 2018, and elsewhere).

Line 478: "improvements in SHRIMP manufacturing and installation may have reduced the fundamental uncertainty associated with the calibration equation" "May have" reads awkward; the data and interpretation in this paper at least do not support this.

Fig. 2: The PDF is based on assigned uncertainties that may or may not be adequate (see comment to Fig. 4)

Fig. 3: I would omit this plot; the fit has a probability of only 0.003, the slope generated is probably an artifact of the data selection, and the results for OG1 show that this relation is invalid (including younger reference zircon would probably also confirm this). "Cherry picking" and "p-hacking": why even go there?

Fig. 4: MSWD and probability of fit suggest that there is overdispersion/underestimation of uncertainties for the SHRIMP results.

Fig. 5: Statistical testing of the difference/equivalence of both distributions would be required to demonstrate that this distinction is significant (e.g., using a Kolmogorov—Smirnov comparison).

Additional references:

Burgess, S. D., Coble, M. A., Vazquez, J. A., Coombs, M. L., & Wallace, K. L. (2019). On the eruption age and provenance of the Old Crow tephra. Quaternary Science Reviews, 207, 64-79.

Keller, C. B., Schoene, B., & Samperton, K. M. (2018). A stochastic sampling approach to zircon eruption age interpretation. Geochemical Perspectives Letters (Online), 8(LLNL-JRNL-738859).