REVIEWER 1

The paper "Calibration methods for laser ablation Rb–Sr geochronology: comparisons and recommendation based on NIST glass and natural reference materials" by Glorie et al provides a comparison of calibration approaches for LA-ICP-MS/MS Rb-Sr dating. As Glorie et al. show in their paper, many different protocols are currently in use and there is no consensus on a best-practice approach. Importantly, they demonstrate that calibration against Mica-Mg, a very commonly used reference material for Rb-Sr dating, can produce inaccuracy in the Rb-Sr dates. This paper is well-written and overall well presented, although some of the figures present more as drafts and could benefit from some final editing/formatting.

This paper constitutes an important contribution to the Rb-Sr community as it highlights significant pitfalls of some commonly used data reduction strategies. My main comment is that there is no code/software package provided to apply the data reduction protocols recommended by the authors, making it hard for the inclined reader to replicate the approach recommended in this work. I would strongly encourage the authors to consider either i) expanding the description of their data reduction algorithm (including equations), ii) providing a reference to a detailed protocol, iii) or sharing the in-house code the authors used, which would greatly benefit the practical usefulness of this work.

Detailed Comments:

Line 23: "as secondary RM to correct for matrix-dependent fractionation" The term secondary RM is misleading here, as it refers to a RM used for calibration purposes and not one used for QA/QC. This terminology is also inconsistent with Table 1 where "secondary RM" is used in the "traditional" way.

Line 67: see comment above. In these two sentences "secondary RM" is actually referring to two different things. Perhaps the RM that is used to correct for matrix-dependent fractionation could be termed "matrix-calibration RM" or something similar to avoid confusion.

Lines 86-87: see comments above regarding terminology.

Lines 173-175: "using a customized data reduction algorithm that calculates error correlations from the raw isotopic ratios for each sweep in an analysis, in the same way as for U-Pb data reduction". It would be useful to describe this in more detail or provide a reference for the interested reader. As you show in Table 1, many publications do not report error correlation at all or use an estimate. Having a best practice approach described somewhere – or even a code publicly available - would be very useful for the community. Since rho affects the precision of isochron ages, I believe this would be a useful addition to this paper.

Line 176: "(1) correcting down-hole fractionation (DHF) against the primary RM" see comment above. To my knowledge there is no publicly available software to do this (such as a DRS for iolite or similar).

190-192: "All mica samples (including biotite samples and MDC phlogopite) were ablated with the laser ablating parallel to cleavage. The Bundarra samples were analysed in thin section and the Taratap sample as a rock block." From the petrographic description it sounds like biotite does not have a preferred orientation in the Bundarra and Taratap samples. Did you specifically target grains with the cleavage perpendicular to the surface for analysis in these samples? And why? Please clarify.

Lines 204-205: “Session-dependant correction factors (CF) were calculated from the measured 87Rb/87Sr ratio for MDC and Mica-Mg and compared to the reference value” Are these fractionation factors stable within an analytical session? If not, using a spline rather than a constant correction factor may yield better results. EDIT: This is clarified in section 4.2. It may be useful to add half a sentence to line 204 to justify the use of a constant correction factor.

Line 265: “As shown, analytical protocol (A) involving NIST as primary” Specify which NIST SRM.

Figures and Tables:

At least in the low -esolution version, some axes labels/legends are not readable (e.g. Figure 2) or seem to be cut-off on the side (e.g. Figure 4). Spelling is inconstant between and within Figures (e.g., “Mica-Mg”, “MicaMg”, “MicaMG”)

Table 1: What does “estimated formula” mean?

Figure 2: Please add mass numbers to axes labels

Figures 3&5: Color ramp on the side not labelled.

Figures 6&7: What is the uncertainty level of the error bars?

Reply

We thank the reviewer for their constructive comments and are pleased to read that the manuscript was well received. Below, we respond to each individual comment.

Data reduction strategy

As indicated in the manuscript, we use the commercially available software package “LADR” (https://norsci.com/?p=ladr), which has been used by geochronologists in many previous publications. The software can calculate quantified ratios and trace element compositions simultaneously, and the algorithm is explained in depth in the software manual and the cited paper: Norris A & Danyushevsky L (2018). 'Towards Estimating the Complete Uncertainty Budget of Quantified Results Measured by LA-ICP-MS', Goldschmidt, Boston, 2018-08-12. As the title says, the main advantage of LADR over other platforms is that the associated uncertainties are fully quantified and fully interrogatable. We believe it is not in the scope of this paper to outline the algorithms of LADR further beyond what is already presented in the manuscript. It uses standard equations that are cross-platform and the reader is referred to LADR documentation for further details. The only additional correction that is applied outside of LADR (as explained in the manuscript) is a matrix-dependant correction on the Lu-Hf ratio. We have added more detail in the revision.

Detailed comments

First 3 comments on RM terminology: We agree that this can be further clarified. NIST610 or Mica-Mg is used as a calibration reference material for normalization and drift correction. MDC is used as a calibration reference material to account for matrix-induced offsets (i.e. inaccuracy caused by use of non-matrix matched reference materials for normalisation and drift correction). We plan to introduce 3 levels of RMs in the revised manuscript. The primary RM (PRM) used for normalisation and drift correction, a matrix calibration RM (MCRM) used to correct the Rb/Sr ratio for matrix induced offsets between NIST/MicaMg and natural materials, and the secondary RMs (SRM) to verify the accuracy, as has been used for other decay systems.

Rho calculation: LADR calculates Rho values in the same way as Iolite does (and hopefully any other platform available). As the line says, this is the exact same as for traditional LAICPMS U-Pb data reduction. We don`t see much use in explaining error correlations further as it is (or at least should be) common practice in LAICPMS geochronology. We removed the `customized` in Line 173-175, as it isn`t really customized, we are just using an inbuilt algorithm in the software.

Correcting DHF: We are surprised about this comment, as down-hole fractionation is a common issue between most geochronological methods. For example, down-hole fractionation corrections need to be applied to accurately quantify U/Pb ratios. This is not different for Rb/Sr ratios. We have used the in-built standardized DHF algorithms to correct our Rb-Sr data and have indicated this in the revision.

Mica orientation: This is an important comment, as ablation parallel vs perpendicular to cleavage gives different apparent Rb/Sr ratios. Hence, RMs and unknowns need to be ablated in the same way. For the Bundarra and Taratap samples, the rocks were cut perpendicular to cleavage and inspected under optical microscopy to ensure ablation was only being done parallel to cleavage. We added an extra line to clarify this in the revised manuscript.

Correction factors: Yes, these are stable as the data is already drift-corrected prior to calibrating the Rb/Sr ratios for matrix-dependant fractionation. We clarified this in the revision.

NIST SRM is NIST-610. We adjusted this in the revision.

Figures: All fonts and keys appear easily readable in the submitted figures, so this comment might be a function of the resolution of the compiled document. We have ensured that all text is fully legible in the revised figures (and have increased font sizes were relevant). We have also ensured a constant spelling for Mica-Mg.

Fig. 2: masses are added as requested.

Fig. 3&5: colour ramps have been removed and simplified (see comment reviewer 2).

Fig. 6&7: All 2SE uncertainties – this is added in the captions.

Many thanks for your helpful review.

REVIEWER 2

This paper presents some issues with normalisation of LA-ICP-MS/MS Rb/Sr geochronology, namely, differences in downhole fractionation of different reference materials. The paper provides guidance on what gives the best data in terms of accuracy and precision using the most common RMs currently in circulation.

This contribution is most welcome. Clearly different strategies are in use by different labs, and highlighting potential issues with accuracy is needed as this method is rapidly growing.

It took the zircon community many years to try and adopt some of the best practise, and still there is massive variation in approach, and the methods and guidance continue to evolve. Thus, this paper will no doubt be surpassed in the future, but is highly relevant right now.

Although the data presented and the general message is fine, I feel some of the details need better communication and clarification.

The first major point is about how primary and secondary RMs are used, and the second is about uncertainty propagation. Both of these are highly relevant to other LA based geochronology methods, and are issues that equally need addressing and highlighting elsewhere.

Normalisation

As with U-Pb dating of common-lead bearing minerals (e.g. calcite), there are three main options for the normalisation of Pb/U ratios – 1) normalise to something 'homogeneous' such as NIST, and then conduct a secondary normalisation to a heterogeneous (but isochronous) matrix-matched mineral, 2) normalise directly to a heterogeneous matrix-matched mineral, or 3) normalise to something like NIST and then just check for accuracy using a matrix-matched mineral. The third should be avoided, but I have seen it done. Options 1 and 2 sound simple enough; however, there is some nuance here that is very important, and should be clarified. One method involves a downhole correction to NIST for the Pb/U ratios, and then just re-normalising the mean ratios offline to a secondary RM. The alternative is to use a downhole correction to the matrix-matched mineral. Again, I have seen both options used. The issue is that ablation data can be split up in Iolite and LADR type programs, such that a shorter ablation interval will have a different mean offset than the total ablation (leading to inaccurate results). I acknowledge that the authors know all this, and mention some of these issues; however, it is not clear of the exact procedure adopted here throughout. I would argue the terminology of primary and secondary RMs should be equally clear and consistent. If something is normalised first to NIST, and then to a matrix-matched mineral, this matrix-matched mineral is still a primary RM.

A figure outlining the data reduction protocol might be useful.

Uncertainties

Error correlations are not trivial calculations, and thus some description of how these are calculated by LADR might be useful.

On appearance, the uncertainty propagation seems reasonable; however, it is not entirely clear nor presented in a single section.

Is the reproducibility of the primary RM propagated as excess variance, or simply as its relative uncertainty? And on to the datapoints or the isochron? Best practise guidelines from the U-Pb community are to propagate the primary RM ratio uncertainties as 'excess variance' onto the datapoints. The relative age/isochron uncertainty of the primary RM should then be additionally propagated onto the age/isochron uncertainty (although admittedly this step is not part of the published guidelines in Horstwood et al., 2016).

The authors highlight some issues with the use of heterogeneous RMs, namely that if they are used for uncertainty propagation (i.e. increase the datapoint uncertainties to achieve a MSWD of 1 on the primary RM), then datapoints will have overestimated uncertainties.

There are issues with the current simplistic way of conducting data reduction, of which this is one. Arguably, datapoint uncertainties should not be estimated at all using heterogeneous RMs. Current data reduction programmes generally do not allow for such nuance.

I feel the authors could state these issues more clearly, rather than just listing which combination of RMs gives the most precise and accurate final data.

Data

I would expect a full analytical protocol table to be included.

I would expect to see at least one signal as cps or mV in the table.

Normally these tables also include concentration estimates too, although I realise these have a large uncertainty due to matrix-matching and the potential heterogeneity of the materials.

Figures 3 and 5 – The scale could be misleading, as there is no gradation in the data, they are discrete classes.

Figure 6 and 7 – These are lifted from excel by the looks of it, and could be tidied up a little.

Line 103 – state confidence limit for reference ages and ratios – i.e. 2s.

Line 176 – Presumably a background correction was calculated first?

Line 183 – What is the source of the ppm data?

Line 186 – At what stage is this done? Do you mean NIST-610 was used as the primary RM for both Rb/Sr and Sr/Sr ratios, followed by a second correction of the Rb/Sr ratios to the secondary RM?

Line 191 – see above comment on orientation.

Line 192 – surely not all micas are orientated the same way in a rock block? One should be explicit and honest – i.e. the micas were presumably orientated at varying angles, but those analysed have their cleavage close to vertical? Can the range in angles be estimated?

Line 210 - uncertainty propagation uses the relative uncertainty as % - i.e. not as excess variance?

Line 216 – can a supplementary figure be used to compare DHF across different sessions?

Line 228 – Are these with or without drift correction – please state?

Line 231 – Variations will also be due to drift in instrumental conditions that affect mass bias, such as plasma ionisation.

Line 240 – Right, so clarifying this earlier on might make it clearer for the reader.

Line 302-305 – I think there is more nuance here than is discussed – see above.

Line 315 – I would start with a simple phrasing – the datapoint uncertainties will be overestimated – then list the problems this may cause.

Line 317-323 – Some of this would be fixed if a more comprehensive data reduction strategy was employed, where uncertainty propagation and DHF are not necessarily directly linked – see above.

Line 324 – 342 – This gets very wordy. It might be better visualised with a figure in addition to the text. The message is that using RMs with heterogeneous ratios, or that cannot be measured precisely, leads to inaccurate propagation of uncertainties. This message could be clearer and stronger.

Normalisation: We agree that there is ambiguity in the presented manuscript around what each RM does. We added clarifications on that in the revised manuscript and define 3 types of RMs: (1) The PRM, which is used to normalise, correct for drift and DHF; (2) the MCRM, which is used to calibrated the Rb/Sr ratio (after data reduction) to a natural standard with heterogenous Rb/Sr ratio but constant age; and (3) the SRM, which is used for accuracy verifications.

We have demonstrated that there is no difference in DHF between the different analysed minerals and thus can use a phlogopite to correct any of the analysed materials (note, that is at the 67µm spot size, typically used for Rb/Sr dating. We will specify this as smaller spot sizes might break this argument). There is a difference in DHF behaviour in NIST-610, which we use to correct for DHF, and the natural micas. However, currently available software cannot deal with isochronous primary standards besides the U-Pb system (as acknowledged in the manuscript, this might change in the future), and therefore we correct to a homogenous material (NIST glass) and do the off-line correction for Rb/Sr fractionation. We have reworded this section slightly to accommodate the comment.

Uncertainties and error correlations: It is not the scope of the paper to outline how LADR does reduction. It is an accepted tool for geochronology that is widely used and it is well documented in the software manual / associated paper. In fact, LADR calculates the full uncertainty budget, which

can be demonstrated and inspected with the uncertainty tree that can be queried after data reduction has been completed. We have added an example of such uncertainty-tree in a supplementary file to illustrate how the final uncertainty is calculated. Indeed, uncertainties related to the PRM and normalisations are added as extra variance on each data point. However, the uncertainty on the matrix calibration reference material (now called MCRM) is not added to the individual datapoint, but to the regression itself. What we do here is simply calibrating the Rb-Sr ratios based on the age off-set between the RM and the measured values. Hence, this uncertainty is effectively an `age` calibration and should only be added to the age calculation, rather than as extra variance to individual data points. We added more explanation on this in the revision.

Data: Analytical conditions are presented in Table 2 and the analytical protocol is exactly what this paper is about. We are unsure why a table would be useful here, nor what to include in it.

Fig. 3 &5: We changed the colour gradient to discrete steps rather than a continuous band.

Fig. 6 & 7: We don`t see any issue with using excel figures as long as they are informative. Masses have been added as well as some minor tidying up.

Line 103: Uncertainty levels are better specified in the text as suggested.

Line 176: correct, this is added.

Line 183: These are the GeoREM preferred values (added in the revision).

Line 186: correct – we added a clarification here around the RM name usage.

Line 191: We didn`t find an `above` comment on orientation, unless you mean from the other reviewer? As explained in the reply to their review, rock chips were prepared with cleavage upright and inspected for variations with optical petrography.

Line 192: We have used optical microscopy (using birefringence) to only ablate micas parallel to cleavage. We don`t expect more than 10 degrees variation as that would be visible under the microscope. We added a line to explain the strategy for ensuring ablation parallel to cleavage.

Line 210: correct. As explained above, we are applying an age off-set correction, which effectively shifts the isochron. The uncertainty on that needs to be propagated onto the isochron age uncertainty, not as extra variance on the data-points themselves.

Line 216: That is A LOT of work for a figure in a supp file, to show there is no difference between sessions. When the same analytical conditions are used, why would there be a difference? It is not something we deem useful here.

Line 228: without drift correction, as these graphs essentially show the drift. We added some words to clarify.

Line 231: yes, but drift is essentially what is shown in this figure.

Line 240: we clarified that these are drift curves.

Line 302-305: We added some of the info you have provided above in the revision, but we don`t want to add too much detail as this is `common knowledge` in geochronology.

Line 315: good suggestion – we reworded accordingly.

Line 317-323: We don`t disagree, but current data reduction programs cannot separate these issues, leading to problems in currently published literature. The whole point of this paper is to point out issues with the tools the community currently use and strategies to get the most accurate data possible with those tools.

Line 324-342: We reworded slightly to make the message stronger indeed. The figure that this text refers to is fig. 5.

Many thanks for your helpful review.