Reply to referee reports:

We thank both reviewers for their insightful reviews. In response, we would like to take this opportunity to improve the structure of the paper. The source of dispersion (excess to counting stats) in U-Pb SIMS data has been debated for decades; this debate can most succinctly be summarized by the arguments of Black and Jagondinski (2003) on one hand, who say that an excess error term needs to be applied if the source cannot be identified, and Compston (2000), who argues that all the scatter can be traced to bad zircons. We didn’t include either paper in our introduction as the actual data they were discussing (multi-grain pre-CA TIMS aliquots vs SHRIMP 1 data) are obsolete, but the ideas are still current. Indeed, and Black and Jagodzinski (2003) say in their subsection “The Way Ahead”:

> “ There are two alternative courses of action to adopt once it is accepted that uncertainties exceeding those predicted from counting statistics can be generated as part of the SHRIMP analytical process. The first is to empirically quantify the magnitude of variation by means of replicate analyses, and then to use SHRIMP only for those projects where such variation (e.g. 1–2%) is acceptable. The other approach is to delve as deeply and objectively as possible into the various sources of uncertainty in SHRIMP dating, so that they can be identified, understood and ultimately either minimised or removed altogether.”

Moving our treatment of uncertainty from the intro to the methods (as reviewer 2 suggests) will allow us to streamline the introduction to highlight how we can now examine both calibration issues and inhomogenous natural sample issues using our data of doubly dated zircons.

As both reviewers recommend removing figure 3 and minimizing the associated analysis, we can replace this with discussion about the calibration. In the current manuscript we skip that analysis and go straight to looking at natural mineral-based explanations, which is jarring when you consider how much of the introduction is spent explaining the calibration. We have re-examined the calibration-related data enough to know that there is information there to discuss, and this would make the paper more interesting and help tease out the competing factors of calibration quality vs geological accuracy.

This leads us to the relevance and usefulness of considering sub-percent SIMS data. It should not be surprising that most of our data are reported with sub-percent precision, as most current SIMS U-Pb data falls into this category. In both rounds of the recent G-Chron U-Pb proficiency testing (Webb et al., submitted), the reported median 2sig uncertainty for the $^{206}Pb/^{238}U$ SIMS age was 0.5% in round 1 and 0.6% in round 2. As our dataset contains many samples in this range, it is worth considering how SIMS data in that range of precision compares to CA-ID-TIMS.

This leads us to the questions about the source and application of uncertainty in central to the comments from reviewer 1 about the statistical validity of our data analysis.

Using the data as reported, even with the exclusion of the outliers discussed in section 4.2.2, the weighted mean of the reported data has an MSWD of 1.9 and a probability of fit of 0.002, indicating that it is not a homogenous population. Obviously the probability of fit can be increased by arbitrarily increasing the uncertainties of each age- increasing them to 2% give an MSWD of 0.35 and a 100% probability of fit, for example. But as Reiners et al. (2017) suggest in chapter 4, hiding dispersion by the use of excess error should only be done if a physical explanation cannot be found; this is why it is important to consider explanations related to the physical samples. As far as our hypothesis that the data contains two populations is concerned, we would like to point out that the mixture-modelling approach of Sambridge & Compston (1994) shows that the statistically most likely population split is similar to that derived from dividing the samples into intrusive and extrusive rocks.

Obviously we cannot accommodate Reviewer 1's request to both cull data and have a larger dataset. One result of compiling this data, which was first done in the late 2010's, is that it tells us which samples are not likely to yield geologically useful ages from SHRIMP analyses, and should be sent straight to a CA-ID-TIMS lab. As a result, the incidence of double dating has been reduced, and there is only one additional doubly dated sample which has appeared since 2016. As it wasn't available until the manuscript was in internal review, and it is yet another outcrop of the Emmaville Volcanics, which are already over-represented relative to the rest of the Australian continent in this study, we didn't add it in between reviewers. However, we would be happy to include it in the final paper. As for culling data, we feel that a discussion of the calibration behaviour and quality (see above) will help put various outliers in context.

We appreciate reviewer 1's suggestion of Burgess et al. (2019), and agree that it is a great paper on Pleistocene tuff dating. However, the papers on Permian tuffs cited by us which describe the zircons discussed in this paper are more representative of the problems we have in these rocks. For example, figure 12 in Metcalfe et al 2015, shows that most of the samples in this drill core which have been analysed by CA-ID-TIMS contain zircons crystals which predate the eruption age. A particularly striking example is the second lowest tuff, GA2122738, with an eruption age of 254.34 +/- 0.08 Ma. The next three tuffs above this in the drill core contain inherited (or contaminant) zircons whose age is within uncertainty of the GA2122738 eruption age. These grains are excluded from the weighted mean eruption age in each case because the TIMS is precise enough to identify them.

The uppermost tuff in this sequence (GA2122750) is one of the SHRIMP analyses in which the SHRIMP age is older than, and not within uncertainty of, the CA-ID-TIMS age. However, the old outlier grain in the CA-ID-TIMS analysis of GA2122750 is within error of the SHRIMP age, as are the next six tuffs lower down in the drill hole. The 2sig uncertainty on each individual SHRIMP spot for this sample is on the order of 5 Ma, so distinguishing spots on antecrysts instead of eruption-age zircons by U-Pb date alone is impossible, allowing accidental antecryst analyses to bias the SHRIMP ages older.

We are happy to incorporate the line-by-line corrections not addressed above as appropriate when revising the paper.

References cited:

Black, L. P., Jagodzinski, E. A. Importance of establishing sources of uncertainty for the derivation of reliable SHRIMP ages, Australian Journal of Earth Sciences, 50:4, 503-512, 2003. DOI: 10.1046/j.1440-0952.2003.01007.x

Compston, W. Interpretations of SHRIMP and isotope dilution zircon ages for the geological time-scale: 1. The early Ordovician and late Cambrian Mineralogical Magazine (2000) 64 (1): 43–57.

Reiners, P., Carlson, R., Renne, P., Cooper, K., Granger, D., McLean, N., Schoene, B. (2017). Interpretational approaches: making sense of data. In Geochronology and Thermochronology. Wiley. 10.1002/9781118455876.ch4.

Sambridge, M.S., Compston W.: Mixture modelling of multi-component data sets with application to ion-probe zircon ages. EPSL 128 373-390. 1994

Scientific Technical Report STR - Data 21/06 ISSN 2190-7110G-Chron 2019 – Round 1

Webb, P., Wiedenbeck, M., Glodny, J. An International Proficiency Test for U-Pb Geochronology Laboratories -Report on the 2019 Round of G-Chron based on Palaeozoic Zircon Rak-17 Submitted