Prof. Pieter Vermeesch
University College London
+44 (0)20 3108 6369
p.vermeesch@ucl.ac.uk

8 August 2022

To: Michael C. Sitar, Colorado State University

Dear Mr. Sitar,

Thank you for submitting your Technical Note entitled "colab_zirc_dims: a Google-Colab-based Toolset for Automated and Semi-automated Measurement of Mineral Grains in LA-ICP-MS Images Using Deep Learning Models" to *Geochronology*. Having considered your paper, the two reviews and your response to the reviews, I have decided that your manuscript is suitable for publication in *Geochronology* after revisions. In addition to the reviewer comments, I would like to add a few thoughts of my own.

1. I found the paper easy to read up to Section 3.2, when its complexity suddenly increases. Here the text contains a frightening number of acronyms and technical terms, including Mask RCNN, MS COCO, NMS, Detectron2, Swin-T, FPN, ResNet-50, ResNet-101, "backbone network", Centermask etc. There are three problems with this complexity. First, the AI jargon won't make sense to the vast majority of GChron readers, who are not experts in this field. Second, given the rapidly changing landscape in AI technology, it won't be long before the specific tools used in `colab_zircon_dims` are superseded by more performant alternatives. Therefore, even AI experts may have trouble understanding the paper in the future. Third, whilst it is easy to update software to keep up with technical developments, the same is not true for academic papers. If you swap out some components in `colab_zirc_dims`, then the notebook will be 'out of sync' with the *GChron* paper. In order to make the paper more future proof, I suggest rewriting the text in a more generic form. Please explain the AI segmentation algorithm in general terms and dedicate fewer words to the specific implementation. Technical notes are meant to be short anyway, so much of the specific details could be moved to the online documentation of the Jupyter notebook. I appreciate that it is not possible or desirable to remove all jargon from the paper. It would be useful to add a table to the revised manuscript, listing the remaining definitions and acronyms.
2. It is not clear how the apparent grain sizes measured by the AI algorithm on 2D images relate to the actual size distribution in 3D. The introduction mentions some published studies investigating age-size relationships in zircon U-Pb geochronology. Some of these studies (Lawrence et al., 2011) used sieves, whereas others (Cantine et al., 2021) used images. Are there any studies that have compared both approaches? I would imagine that their results can differ significantly. On a related note, the reviewers have already highlighted some issues caused by `colab_zirc_dims`'s reliance on reflected light images. As pointed out in your manuscript, reflected light images of polishing surfaces tend to underestimate the actual grain sizes. Cross sectional areas also depend on the polishing depth and on whether the grains are mounted as SIMS-style epoxy pucks, or on glass slides (Figure 1). As a consequence, cross sectional grain sizes can only be used to compare samples prepared in the same lab and by the same analyst. I do not think that they can be used to compare samples from different labs or prepared by different analysts. This greatly reduces their usefulness. The revised manuscript needs to assess these limitations upfront.

Prof. Pieter Vermeesch
University College London
+44 (0)20 3108 6369
p.vermeesch@ucl.ac.uk

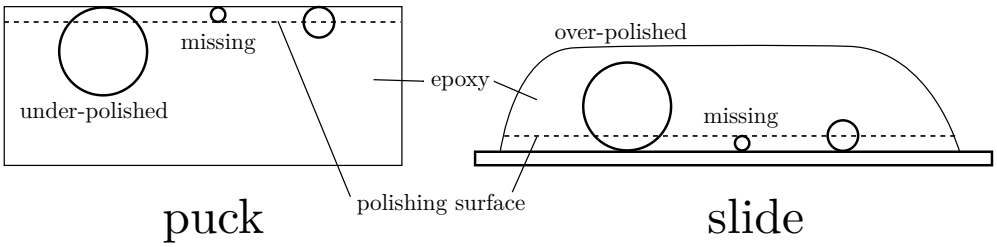

Figure 1: Some laboratories mount zircons in epoxy pucks (left), whereas others use glass slides (right). In both cases, the apparent grain size in cross section depends on the depth of the polishing surface (dashed line). Grain size variation leads to variable degrees of over- and under-polishing. This diminishes the ability to compare grain size distributions between analysts, and between laboratories.

3. According to Section 5.2 of the paper, it would take an estimated six hours to determine the grain size distribution of the Leary dataset. This is impractical. Unless the grain size measurements are truly effortless, I'm afraid that the AI approach won't get much use. I understand that the Jupyter notebook is a proof-of-concept product. What would need to be done to make it faster or easier and use?
4. Section 3 lists the advantages and disadvantages of Google Colab. It does not mention two problems. First, Colab requires an internet connection, yet many lab computers are not connected to the web due for security reasons (automatic OS updates are disabled on most lab computers). Second, Google products are not accessible from China, which is a huge 'geochronological market'.
5. In your response to Dr. Nachtergaele, you chose not to follow his suggestion to add a plot of grain size vs. U-Pb age to your paper, because such a plot is already scheduled to appear in an upcoming *JSR* paper by Leary et al. I would urge you to reconsider this decision, for two reasons. First, the Leary et al. study only presents the manual measurements, and not your automated results. Second, the paper will be stuck behind a paywall, so not all *GChron* readers will be able to check this useful figure. I suggest that you replace Figure 8 with a scatter plot of grain size vs. U-Pb age for a representative sample, with a box plot and KDE shown along the y- and x-axis, respectively.

Geochronology normally gives authors four weeks to complete the revision. I would be happy to extend this if you need more time to address my first comment. Please do not hesitate to contact me if you have any questions.

Sincerely yours,

Pieter Vermeesch
Associate Editor
Geochronology