Dear Chief Editor (ice) Mankoff,

We sincerely appreciate your patience and constructive comments, which are crucial to helping us improve the quality of our manuscript. We tried our best to address each of your comments as meticulously as possible and revised typos and grammar throughout the main text. We invite you to see our responses below.

Comments to the author:

Table 4: Good. Now can you discuss why one product is 10 - 15 % off of the other product? [Response] Yes, we agree with you that a discussion about the difference between the two products is important. Table 4 is put in the Results, so we discuss the discrepancy in Section 6.2 (Comparison of Sentinel- and Landsat-derived products), please read the analysis there. For clarity, we also revised the sentences as below:

'In each size class, the overlap ratios are greater than 85% in count and area, and there are also a higher number and larger area of glacial lakes from Sentinel than that from Landsat images. Sentinel-2 images (10 m) with a finer spatial resolution produce more glacial lakes than those from Landsat images (30 m). The discrepancy is mainly attributed to the inconsistency of spatial resolutions and image acquisition dates, as discussed in section 6.2.'

Figure 9: You replied, "As for the X-axis, we prefer using the means (as a compromise) of our mapping and the reference, rather than our mapping alone, to better reflect the scales of each of the lake areas. We hope the reviewer finds this reasonable and acceptable.".

I do not find it reasonable. The point of this graphic and section is *validation* - that is, independently comparing your product to some other product, so that you and your users can understand quality issues. By averaging your product with the validation product on the X axis, you are hiding the differences between the two that would appear on the Y axis.

[Response] Thank you for these constructive comments. We originally used the means of both methods on the x-axis, following the convention of the Bland-Atlman plot (please see the last figure in: <u>https://en.wikipedia.org/wiki/Bland%E2%80%93Altman_plot</u>). But we do fully agree that for the validation purpose, it makes better sense to not mix up the two on the same axis. Accordingly, we have adjusted the horizontal axis of Figure 9c-d, and the lake areas derived from high-resolution Google Earth images (the reference values alone) are shown on the x-axis.



Figure 9. Distribution of the validation sample (a), visual comparison of glacial lakes derived from Landsat and Sentinel-2 images overlaying Google Earth imagery (© Google Earth 2019) in a zoomed site (b), and differences between our glacial lake product (mapped from Landsat and Sentinel-2 images) and the validation reference (digitized from Google Earth images) (c and d).

As for "As exemplified in Figure 11d, the mapped glacial lake areas exhibit a substantial discrepancy, which is likely a joint consequence of both sensor difference and actual glacier lake dynamics that occurred during this short period of time." This seems like the 4th or 5th time I've asked you to adjust this sentence. If you can provide evidence that it is likely that there are any changes in the lake (it doesn't even have to be due to "dynamics") I would like to see that evidence. This graphic alone does not demonstrate that is "likely".

[Response] We sincerely appreciate your time and patience. We replaced the previous case in Figure 11d with another region that indeed shows supraglacial lake changes. And we also rewrote this paragraph for improved clarity.

It now reads 'In addition to the difference in image resolution, different acquisition dates between Sentinel-2 and Landsat images can also contribute to the discrepancy of those two glacial lake datasets. The total number of supraglacial lakes and ice-dammed glacial lakes are less than 300, but those lakes are controlled by glacier movement and temperature changes (Liu and Mayer, 2015; Miles et al., 2018), which vary faster with time than relatively stable glacial

erosion and moraine-dammed lakes. Acquiring same-day images from the two sensors was not always possible due to the impacts of cloud contaminations, topographic shadows, snow cover, and revisit periods (Williamson et al., 2018; Paul et al., 2020). Despite our efforts of leveraging all available high-quality images, the overlap of acquisition dates between Landsat and Sentinel-2 images for the same location is relatively low (only 7 scenes of Sentinel-2 images or 112 glacial lakes in 2020) in this study area, and the consequential temporal gaps led to a difference in the number and area of the derived glacial lakes. As exemplified in Figure 11d, the mapped supraglacial lakes in the same location exhibit a considerable discrepancy, which is likely a joint consequence of both sensor difference and glacier lake evolution.'



Glacial lake derived from Landsat

Glacial lake derived from Sentinel-2

Figure 11. The discrepancy of lake extents extracted from Landsat and Sentinel-2 images.

Table 5: Better. But now that you've matched other's MMU, can't you also match their time, at least when their baseline is a subset of yours? For example, Row 1 and 2 could be combined to common subsets of time and MMU, and then collapsed to 1 row:

1990 (1989-1993), Manual, 5400 (6), Count (them), Count (us), Ratio, Reference

Now we're validating your product against 9 other time periods, and 4 other products.

[Response] Thank you for suggesting this for the purpose of better clarity and we have improved Table 5 according to your advice. We are, however, unable to match the time of different datasets, because producers have to combine images acquired in different years to create the "baseline year" product covering the entire study area. If we filtered the baseline period, it leads to an incompetent coverage of the study area that reduces the comparability. To avoid any misunderstanding, we highlighted the baseline year by deleting the period information and revised the field to be the baseline year (them/us) for better clarity. Now, the revised Table 5 can better validate our product against 9 other periods and 4 other products.

datasets in the study area.						
Baseline year	Method	MMU	Count	Count	Ratio	Reference
(them/us)	(them/us)	m ² (pixels)	(them)	(us)	(%)	
1990/1990	Manual/Semi-automated	5400 (6)	1720	2069	83.13	Wang et al., 2020
1990/1990	Automated/Semi-automated	50000 (55)	145	363	39.94	Shugar et al., 2020
1990/1990	Manual/Semi-automated	4500 (5)*	622	2154	28.88	Zhang et al., 2015
2000/2000	Manual/Semi-automated	4500 (5)*	724	2184	33.15	Zhang et al., 2015
2000/2000	Automated/Semi-automated	50000 (55)	155	361	42.94	Shugar et al., 2020
2008/2000	Automated & Manual/Semi-automated	8100 (9)	1067	1800	59.28	Chen et al., 2021
2015/2020	Automated/Semi-automated	50000 (55)	148	364	40.66	Shugar et al., 2020
2017/2020	Automated & Manual/Semi-automated	8100 (9)	1063	1813	58.63	Chen et al., 2021
2018/2020	Manual/Semi-automated	5400 (6)	1956	2149	91.02	Wang et al., 2020

 Table 5. Comparison between our Landsat-based mapping and other third-party Landsat-based glacial lake datasets in the study area.

Note: MMU represents the minimum mapping unit that is possible to enable a valid comparison between our product and each of the third-party datasets. * The MMU in the dataset of Zhang et al. (2015) is 3 pixels, finer than 5 pixels in our product, so the MMU threshold of 5 pixels was used for this comparison.

I'm not asking for Table 4 level stats for each of Table 5 products. Table 5 as it is (but slightly more apples-to-apples) is, I think, good validation.

[Response] Thank you for your comments that help us a lot to improve the quality of our manuscript.

Finally, you appear to have more lakes than any of your validation products. This may or may not be good. If you can explain some of the differences, that would be helpful for your users to better understand your product. Are the more lakes an improvement and a good thing? Are the other products just looking at a smaller area?

[Response] Truly, our product has more lakes than the previous similar datasets, due to a lack of thorough cross-check quality assurance during their lake mapping over a larger study area. We inspected and validated this through a series of satellite images. According to your comment, we revised the main text in Section 6.3 as below:

'Our product has more lakes than each of the other 4 products at 9 time periods. By inspecting their dataset, we attributed this anomalous discrepancy to a range of glacial lakes that were missing due to a lack of thorough cross-check quality assurance during their lake mapping over a larger study area. And those more glacial lakes show an improvement of our product in contrast to the previous similar dataset. Our Landsat-derived glacial lake dataset has been visually cross-checked over three time periods after the step of threshold-based semi-automated lake mapping and has also been visually validated by Sentinel-derived glacial lakes.'

Here, we selected a region for example by comparing Wang's data in 2018 with ours to show why our product has more lakes. Please see:



Non-public comments to the Author: Dear Authors,

Just to be clear, I am not a reviewer. I am the Chief Editor (Ice) of the journal. My suggestions for changes should therefore be considered differently than a reviewer - although all 4 reviews requested major changes too! This has been a difficult process, and I'm tired of requesting the same change multiple times. I will reject this paper next time I see it if you continue to make claims that cannot be supported by the data, or do not properly validate the product. [Response] On behalf of my co-authors, I would like to express my apologies for the four-time revision that occupied your time.

Thank you very much for your great patience and valuable comments. Your comments significantly helped us improve the quality of our article, such as the presentations in Figures 8, 9, and 11, also Tables 4-5 as well as the corresponding main text. We learned a lot from this process. I would be more than happy to make further adjustments should you find them necessary.

Best wishes,

Yong