Comments to the author:

We sincerely appreciate the patient and constructive comments from this reviewer, which are crucial to helping us further improve the quality of our manuscript. We have addressed each of the reviewer's comments as thoroughly and meticulously as possible, and believe have provided sufficient validation and justification for our datasets. We invite the reviewer and the editor to see our responses below.

+ Figure 9 X axis: Mean of lake areas from only one product, not two. I assume Landsat and Sentinel 2, *not Google Earth*. Google Earth is only included in the Y axis?

[Response] We are sorry for the confusion of the axis label. The Y-axis in Figure 9c/d shows the differences between our mapping results (from Landsat/Sentiel-2 images) and the validation reference (digitized from Google Earth imagery), and the horizontal axis reflects the means of glacier lake areas from both our mapping (Landsat/Setinel-2) and the reference (Google Earth). 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.

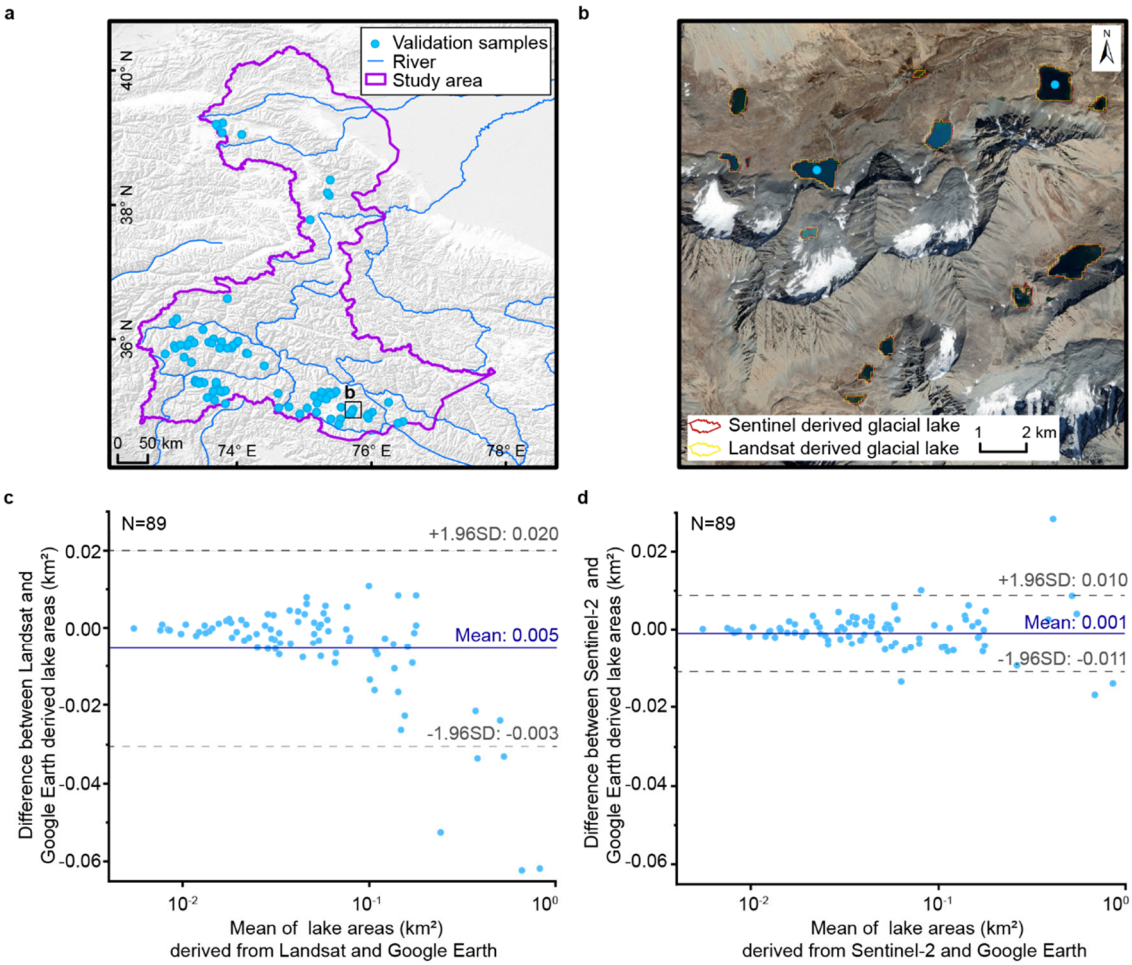

Figure 9. Distribution of the validation sample (a), visual comparison of glacial lakes derived from Landsat and Sentinel-2 images overlaying Google Earth imagery (© Google Earth 2019) in a zoomed site (b), and differences between our glacial lake product (mapped from Landsat and Sentinel-2 images) and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

images) (c and d).

+ L506 to 509. You cannot say that the lakes changed due to time, when you are comparing across sensors. Clearly c1/c2 are only due to sensor (same date). The same is possible of panel (d). We don't know.
[Response] Thank you very much for pointing this out. We agree that the lake area changes are attributed to not only temporal difference but also sensor difference. For improved accuracy, we have clarified the initial sentences of this paragraph as below:

"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. Acquiring same-day images from the two sensors were 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). 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."

+ Table 4 - you're finally getting towards a robust validation. I've repeatedly asked you to compare using properties that exist in both datasets. Why not remove the first row which is only S2 so that you can also compute useful statistics with the bottom Total row? If you want, you can of course mention in the text or figure caption that S2 column excludes 4969 lakes in the 0.0005 to 0.0045 km^2 range.
[Response] Thanks for your constructive suggestion, which we have adopted now. Please see the revised Table 4 below.

Table 1. Count and area of glacial lakes mapped from Sentinel-2 and Landsat images in 2020 in various size classes.

Lake size km^2	Glacial lakes from Sentinel-2 count (km^2)	Glacial lakes from Landsat count (km^2)	Overlap % (%)
0.0045-0.05	2182 (35.52±3.72)	1870 (31.47±9.57)	85.70 (88.60)
0.05-0.1	237 (16.37±0.89)	204 (14.07±2.18)	86.08 (85.95)
0.1-0.2	122 (16.88±0.68)	115 (15.91±1.83)	94.26 (94.25)
⩾0.2	50 (27.20±0.54)	45 (24.86±1.40)	90.00 (91.40)
Total	2591 (95.97±5.83)	2234 (86.31±14.98)	86.22 (89.93)

Note: Second column excludes 4969 (7.73±0.54) lakes in the 0.0005 to 0.0045 km^2 range. Overlap % (%) represent the ratios between our Landsat-derived dataset and Sentinel-derived product in count and area, respectively.

+ I would also like to see the Table 4 methods (comparing using 1:1 properties, so the comparisons are useful) done with 3rd party data - that is external validation. Table 5 (and associated analysis) should be reformatted. It may take a lot of work to for example filter Zhang et al., 2015 with an MMU of 2700/3 to match your 4500/5. Perhaps that is not needed. But it should be fairly easy, given your metadata and familiarity with your own product, to filter yours to match all the other products with MMU pixels > 5.
In my last comments to you I thought this request was clear. I apologize if it was not. I requested

"a like:like comparison" in the section "Comparison with previous similar dataset" which I took to mean external data sets, not comparison within your own dataset.

[Response] We really appreciate this constructive suggestion, and completely agree that a "apples-to-apples (like:like) comparison" is more useful and valid. After a careful deliberation that considered both scientific value and feasibility, we decided to perform a thorough comparison between our mapping and each of the third-party products (as listed in Table 5) based on the possible minimum mapping unit (MMU) for both datasets. For example, the MMU in the dataset of Zhang et al. (2015) is 3 pixels, finer than 5 pixels in our product, so a MMU threshold of 5 pixels was used for this comparison. The other comparisons in Table 5 all follow this MMU logic. For improved clarity, we have reformatted Table 5 to reflect this change. We kindly invite the reviewers to see our revised Table 5 below.

Table 2. Comparison between our Landsat-based mapping and other third-party Landsat-based glacial lake datasets in the study area.

Baseline year (period)	Method	MMU m^2 (pixels)	Count (km^2)	Other data/our product % (%)	Reference
1990 (1988-1993)	Manual	5400 (6)	1720 (89.68±13.69)	83.13 (105.87)	Wang et al., 2020
1990 (1989-1994)	*Semi-automated*	5400 (6)	*2069 (84.71±14.41)*		*This study*
1990 (1990-1999)	Automated	50000 (55)	145 (20.28)	38.77 (36.98)	Shugar et al., 2020
1990 (1989-1994)	*Semi-automated*	50000 (55)	*374 (54.84±5.49)*		*This study*
1990 (1989-1992)	Manual	4500 (5)*	622 (51.93±10.15)	28.88 (61.02)	Zhang et al., 2015
1990 (1989-1994)	*Semi-automated*	4500 (5)*	*2154 (85.10±14.66)*		*This study*
2000 (1999-2001)	Manual	4500 (5)*	724 (61.41±11.91)	33.15 (71.32)	Zhang et al., 2015
2000 (1996-2004)	*Semi-automated*	4500 (5)*	*2184 (86.10±14.83)*		*This study*
2000 (2000-2004)	Automated	50000 (55)	155 (22.35)	42.94 (40.70)	Shugar et al., 2020
2000 (1996-2004)	*Semi-automated*	50000 (55)	*361 (54.91±5.40)*		*This study*
2008	Automated & Manual	8100 (9)	1067 (65.45)	59.28 (78.08)	Chen et al., 2021
2000 (1996-2004)	*Semi-automated*	8100 (9)	*1800 (83.82±13.59)*		*This study*
2015 (2015-2018)	Automated	50000 (55)	148 (21.45)	40.66 (39.11)	Shugar et al., 2020
2020 (2016-2020)	*Semi-automated*	50000 (55)	*364 (54.84±5.41)*		*This study*
2017	Automated & Manual	8100 (9)	1063 (63.23)	58.63 (75.45)	Chen et al., 2021
2020 (2016-2020)	*Semi-automated*	8100 (9)	*1813 (83.80±13.63)*		*This study*
2018 (2017-2018)	Manual	5400 (6)	1956 (102.46±15.48)	91.02 (119.24)	Wang et al., 2020
2020 (2016-2020)	*Semi-automated*	5400 (6)	*2149 (85.93±14.74)*		*This study*

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 a MMU threshold of 5 pixels was used for this comparison. "% (%)" represents the ratios between the third-party dataset and our product in count and area, respectively.

For Table 4, our purpose here is to understand the difference in our own mappings caused by Landsat and Sentinel-2 sensors. To explore the sensor impact, we compared our own mapping under different lake size categories in Table 4. However, our validation with the third-party datasets (as in Table 5) are all consistently based on Landsat images. If we replicate Table 4 for

each of the third-party datasets, the validation will be very lengthy, and we are afraid there would be a lot of numbers (generated for each of the lake size categories and for each of the third-party datasets) that eventually become too overwhelming and disorienting for readers to grasp. Therefore, we prefer keeping our validation more succinct (as in Table 5), i.e., against each of the third-party datasets based on the minimum mapping unit (rather than each lake size category). We hope the reviewer finds it reasonable.

Once again, we very much appreciate the constructive and patient suggestions from the reviewer and apologize for not responding sufficiently well in the last revision. We sincerely hope that the reviewer finds our extended revision this time sufficient and to-the-point. We will be very happy to make further changes should the reviewer find more improvement necessary. Thank you.