Response to Reviewer #1 on "Co-registration and residual correction of digital elevation models: A comparative study"

Comment received: 19 Dec 2022

Key:

Reviewer comment (black)

Response (blue)

Li et al. compared different methods of correcting 3D-shifts and others biases in digital elevation models (DEM) with the ultimate goal to reduce uncertainties in glacier elevation change estimates. They first compared the widely-used Nuth & Kaab (2011) to the less popular Rosenholm and Torlegard (1988) algorithms for DEM coregistration. On top of a simple 3D shift, the latter algorithm also account for any rotation or scale differences between the DEMs. Further, they proposed an improved correction of the structured-biases between the DEMs that have a proper signature in the two directions of image acquisitions (along and across tracks). They go beyond fit by polynomials and sums of sinusoids by proposing a spline-based non parametric model.

I found nothing wrong with this study. However, I am not super convinced that this article, in its current form, fits well in The Cryosphere and its readership. I see two main reasons for that:

1/ DEM differencing is a popular technique to measure glacier changes. However the scope of the present study is really technical with no direct application to glacier changes. The study sites only marginally include glaciers.

2/ The added value of the proposed method is modest. I am not convinced that a gain of 0.2 m (5%) in standard deviation of the residuals between DEMs only covering a single (and not very challenging) test site is sufficient to convince the glaciological community to rethink the way they coregister DEMs. The added value of the spline-based correction of along track residuals is higher but would need to be confirmed in different settings.

Overall the paper is well written, the work is performed seriously but I missed some novel results that would make a real impact on the glaciological community.

Thank you for your constructive comments and suggestions.

In the experimental section of the previous manuscript, only the ASTER DEMs on the margins of the Greenland Ice Sheet (GrIS) were used for two main reasons:

1) There is commonly a large proportion of stable terrain in the scene, which is convenient to visually analyze DEM co-registration residuals.
2) Strong jitter-induced errors remain in the DEM co-registration results, which is desirable for the comparison of different residual correction methods.

Our experiments so far have been carried out on more than 200 DEM pairs from ZY-3, ASTER, SRTM, and Copernicus DEMs in GrIS, High Mountain Asia (HMA), and southern Alps (SALP). The comparative results show that the method of Rosenholm and Torlegard has a greater ability to remove DEM misalignments (83.3% maximum) than the method of Nuth and Kääb. Strong jitters can only be observed from the co-registration results of ASTER DEMs in GrIS. The GAM spline fitting method often yields a higher accuracy than the two parametric regression methods (high-order polynomials and the sum of the sinusoidal functions), but the improvement is relatively modest, because the magnitudes of the errors caused by satellite attitude jitters in ASTER DEMs are typically not significantly greater than those of random errors and unmodeled systematic errors. Considering that the residual correction is just a secondary work in our study, the GAM spline fitting method and related experimental results can be completely removed from the manuscript.

Major comments

1/ Do the conclusions apply in other settings? Map of elevation differences are constructed from several ASTER DEMs in western Greenland with a strong proportion of stable terrain. Ice-covered terrain is restricted to the eastern part of the images/DEM. It seems that images are almost cloud free. This site and the cloud-free images are appropriate to design and test the different methods but are not representative of real case scenario. In my experience, further challenges for DEM coregistration comes from: vegetation (changing with time), large fraction of glacier areas vs. stable terrain, gaps or unreliable data in the DEMs due to clouds, the rough topography leading to higher noise level in the DEMs. Authors did not explore these difficulties and thus their results are not representative of more complex and more realistic situations.

Followed by your suggestion, DEM co-registration algorithms have been compared in more complex and challenging scenarios. We present some representative examples below. Since the three versions of the method of Nuth and Kääb always produce similar co-registration results, we only compare the linear version of the method of Nuth and Kääb (L23 in Table 1) and the method of Rosenholm and Torlegard (L57 in Table 1) to make figures clear.

1) Large fraction of glacier areas vs. stable terrain

We select the DEM pair HMA-1 (Table R1) located in the northern Pamirs, excluding the erroneous observation over the bright snow-covered areas (e.g., upstream of Bol and Oktyabrskiy glaciers) due to ASTER saturation issues. The L57 algorithm can effectively remove the residuals in the east-west direction of the L23 results (Fig. R1a) and improve the co-registration accuracy by 17.5%.

Table R1. Characteristics of DEM pair HMA-1.

Pair ID	Data	Date	Res. (m)	Scene ID	MedAD (L23–L57)
HMA-1	SRTM	11–22 Feb 2000	30	N39E072, N39E073, N38E072, N38E073	9.768–8.060
	ASTER	22 Mar 2005	30	AST14DEM.003:2028219582	(17.5%)

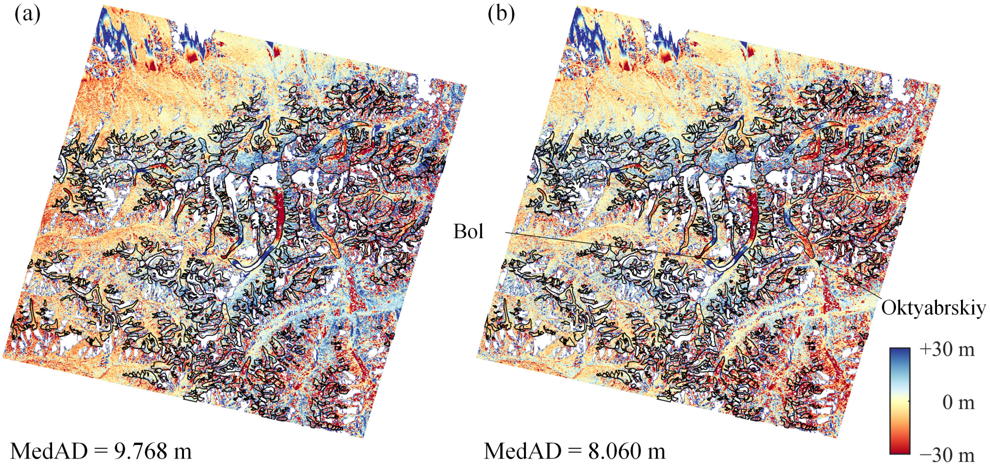

Figure R1. Co-registration results of DEM pair HMA-1: L23 (a) and L57 (b). The black lines mark the glacier boundaries.

2) Gaps or unreliable data in the DEMs due to clouds

As seen in Fig. R2, there are a lot of clouds distributed throughout the center of the ASTER DEM 20050822, and the rotation-induced biases of this image make the glacier in the southeast area of the L23 co-registration results exhibit a positive (Fig. R2a) or negative (Fig. R2b) pattern, which leads to the wrong estimation of the glacier elevation change.

Table R2. Characteristics of DEM group HMA-2.

Group ID	Data	Date	Res. (m)	Scene ID	MedAD (L23–L57)
HMA-2	SRTM	11–22 Feb 2000	30	N39E073	8.598–6.366 (26.0%)
	ASTER	22 Aug 2005	30	AST14DEM.003:2030590191	8.512–7.561 (11.2%)
	ASTER	07 Sept 2005	30	AST14DEM.003:2030819798	6.334–5.790 (8.6%)

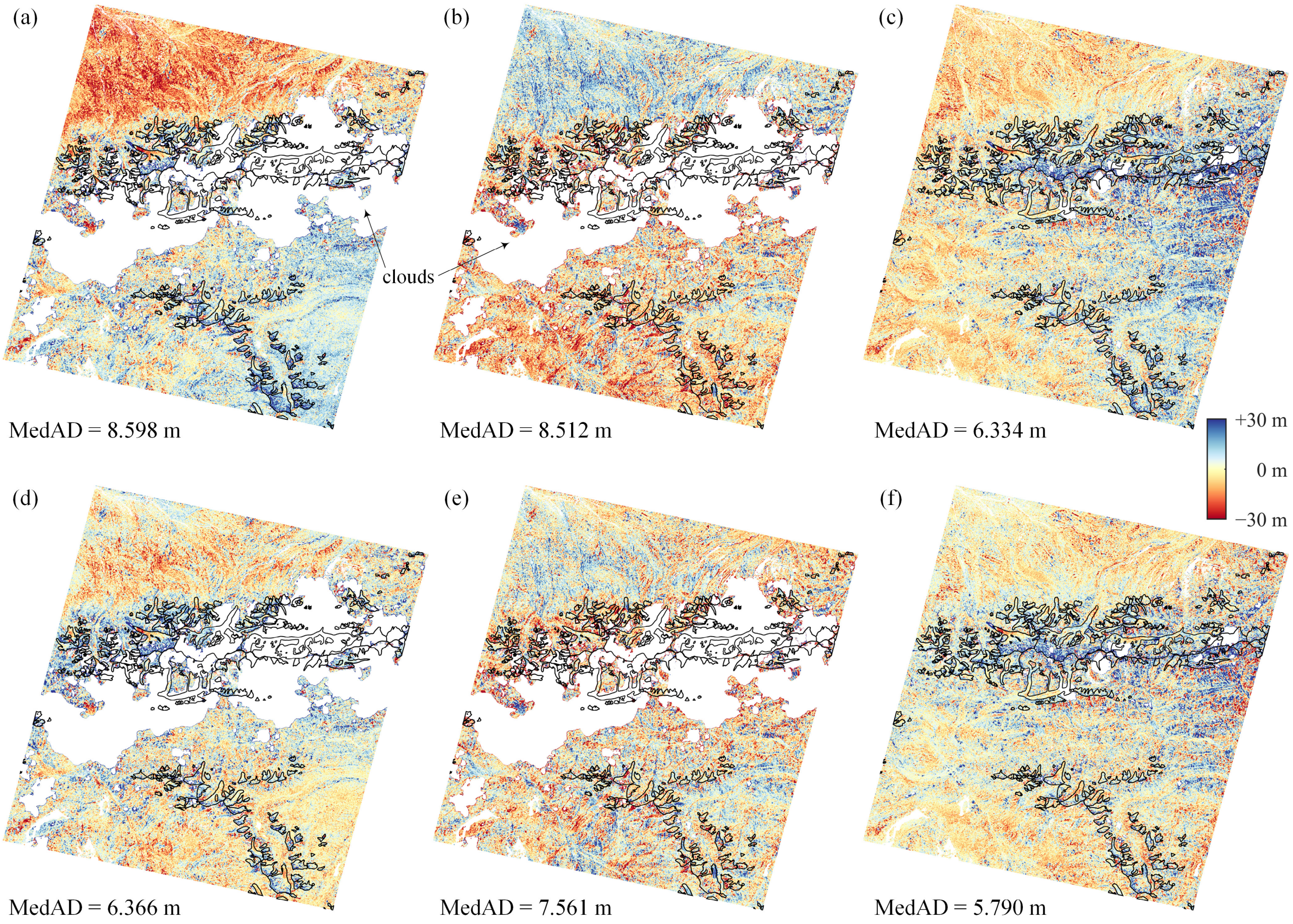

Figure R2. Co-registration results of DEM group HMA-2 based on L23 (top) and L57 (bottom). From left to right: ASTER DEM 20050822 to SRTM DEM, ASTER DEM 20050907 to ASTER DEM 20050822, and ASTER DEM 20050907 to SRTM DEM, respectively.

3) Vegetation

We choose the DEM pair SALP-1 (Table R3) with a large amount of vegetation on the west side (Fig. R3c). After removing the unstable pixels (forest land, water, wetland, and glacier and snow cover), the co-registration results of the L23 method and the L57 method are shown in Fig. R3 a and b. The L57 algorithm can reduce the error in the east-west direction and improve the accuracy by 13.0%.

Table R3. Characteristics of DEM pair SALP-1.

Pair ID	Data	Date	Res. (m)	Scene ID	MedAD (L23–L57)
SALP-1	SRTM	11–22 Feb 2000	30	S44E169, S44E170	7.898–6.871 (13.0%)
	ASTER	24 Feb 2003	30	AST14DEM.003:2011883607	

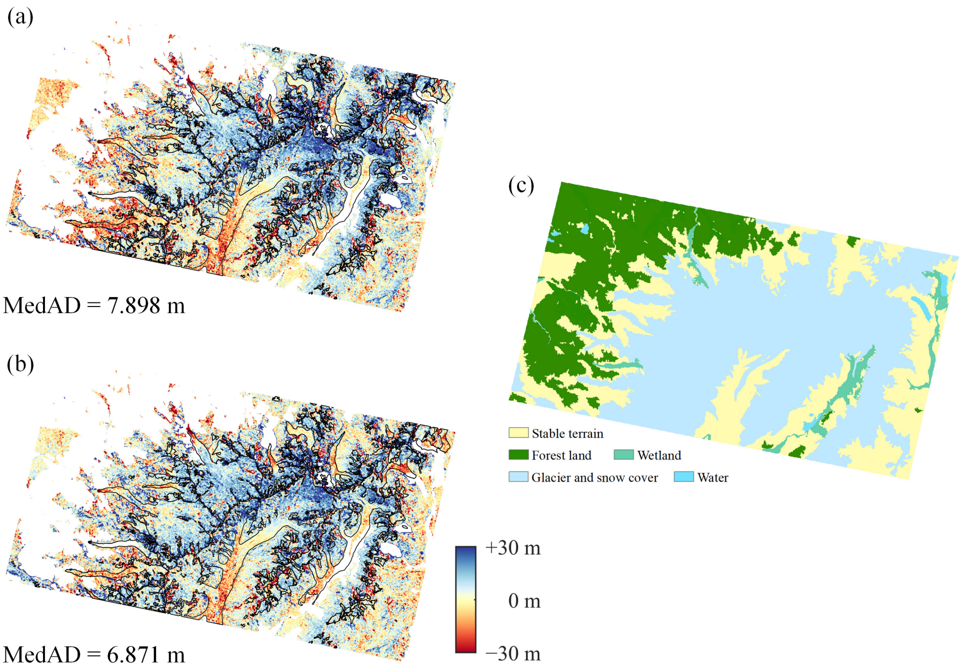

Figure R3. Co-registration results of DEM pair SALP-1: L23 (a), L57 (b), and the land cover map using GlobeLand30 product (c).

4) The rough topography leading to higher noise level in the DEMs

We select the DEM pair HMA-3 (Table R4) with a high noise level in Pamir, and the co-registration accuracy of the L57 method is 8.3% better than that of the L23 method.

Table R4. Characteristics of DEM pair HMA-3.

Pair ID	Data	Date	Res. (m)	Scene ID	MedAD (L23–L57)
HMA-3	ASTER	10 Oct 2017	30	AST14DEM.003:2280543414	7.220–6.621 (8.3%)
	ASTER	26 Oct 2017	30	AST14DEM.003:2281248034	

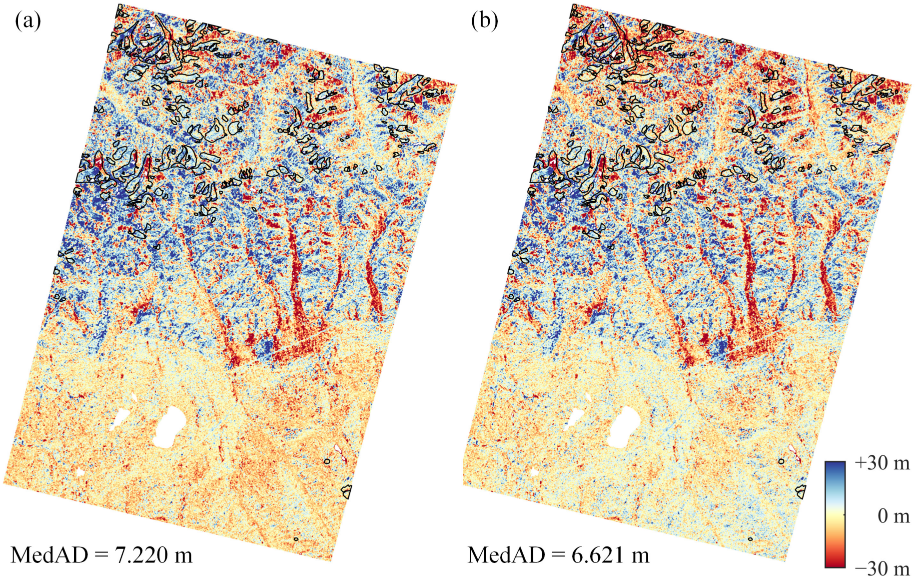

Figure R4. Co-registration results of DEM pair HMA-3: L23 (a) and L57 (b).

2/ Do the improvements over stable terrain percolate to ice-covered areas? To convince the readers (glaciologists, the readership of TC) of the added values of the proposed methods, authors would need to demonstrate real improvements over glacier terrain.

Such a validation is tricky, I reckon, because glacier elevations are constantly changing. I see two ways for the authors to demonstrate this

(a) apply their methods to DEMs derived from images acquired just a few days apart so that the assumption of no elevation change is almost valid. They would then be in position to coregister and bias correct their DEMs over the stable terrain and then check the improvements on glaciers (where no change should be measured over a few days).

Two DEM pairs have been used for validating the co-registration results over glacier terrain.

The first DEM pair is ZY-3 DEM 20171008 and ASTER DEM 20171010, two days apart (Table R5). The co-registration results of the L23 method show systematic errors in the southwest-northeast direction, resulting in a significantly negative bias (-11.826 m) in the northeast region (within the red circle in Fig. R5a). In contrast, the estimations of glacier elevation changes are much close to zero in the results of the L57 method.

Table R5. Characteristics of DEM pair HMA-4.

Pair ID	Data	Date	Res. (m)	Scene ID	MedAD (L23–L57)
HMA-4	ZY-3	8 Oct 2017	30	—	6.126–5.063 (17.4%)
	ASTER	10 Oct 2017	30	AST14DEM.003:2280543414	

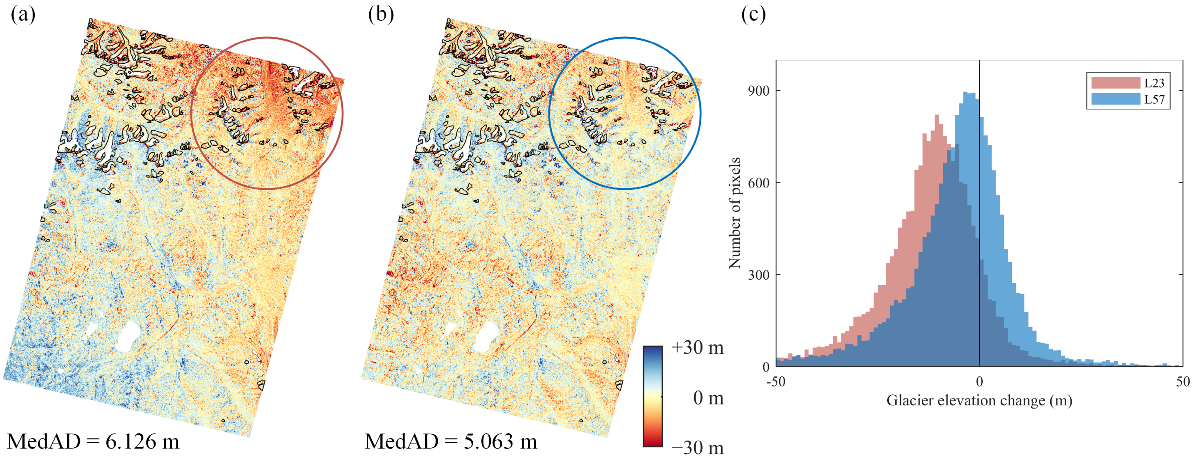

Figure R5. Co-registration results of DEM pair HMA-4: L23 (a), L57 (b), and the histogram of elevation change for glaciers within the circle (c).

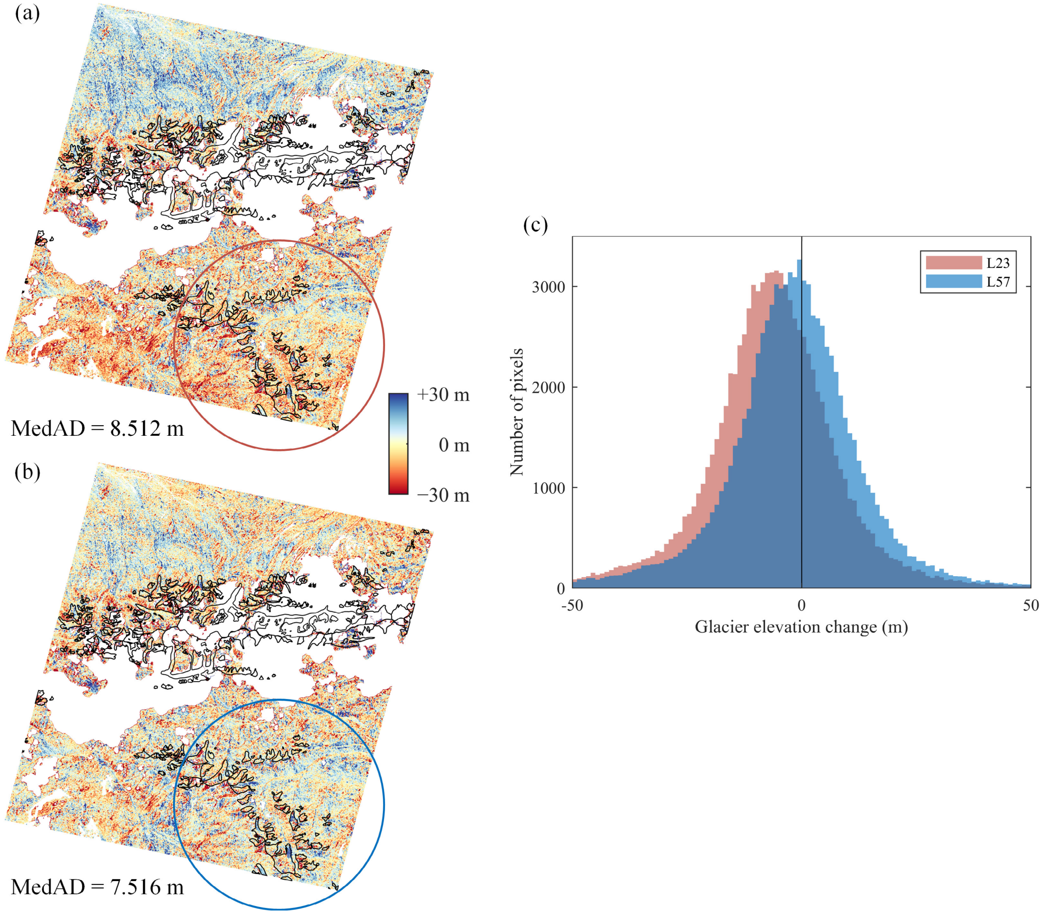

Figure R6. Co-registration results of ASTER DEM 20050907 and ASTER DEM 20050822 from DEM group HMA-2: L23 (a), L57 (b), and the histogram of elevation change for glaciers within the circle (c).

As shown in Fig. R6, the rotation-induced errors in the results of the L23 method can also be observed in the DEM pair of ASTER DEM 20050907 and ASTER DEM 20050822 (Table R2). We further compared the glacier elevation change estimations of ASTER DEM 20050822 to SRTM DEM and ASTER DEM 20050907 to SRTM DEM. Figure R7 shows that the discrepancy in the co-registration results of the L57 method is smaller than that of L23, with the mean value improving from -5.927 m to -1.669 m.

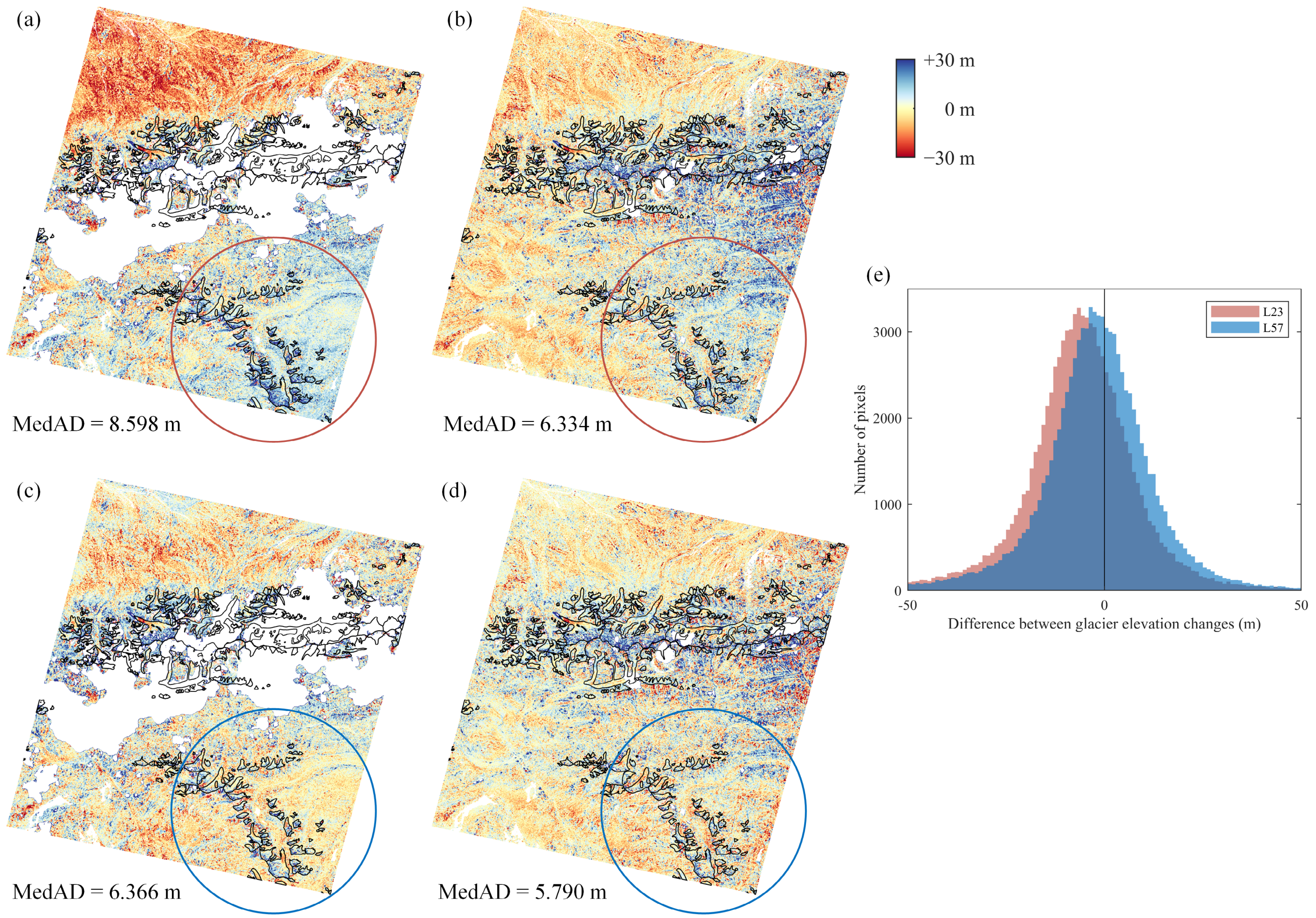

Figure R7. Co-registration results of DEM group HMA-2. ASTER DEM 20050822 to SRTM DEM: L23 (a) and L57 (c); ASTER DEM 20050907 to SRTM DEM: L23 (b) and L57 (d); the histogram of the difference between glacier elevation changes derived from the two DEM pairs (e).

(b) find sites where ASTER DEMs are acquired simultaneously to higher resolution DEMs (for example from the Arctic DEM project) so that a reference elevation change map is available. This second solution is more tricky to identify.

To date, we have not found any ArcticDEM data to meet the needs of validating ASTER DEM co-registration.

3/ The discussion is rather weak. There is a long part about the "extrapolation error" that is mostly unrelated to the rest of the article.

In the experiment and discussion sections, we investigated the choice of algorithms in accuracy and reliability, respectively. As shown in Fig. R8, when stable regions are located on one corner of the DEM, the method of Rosenholm and Torlegard is prone to produce incorrect predictions over glacier covered regions. Similar problems exist in all the application scenarios required large extrapolations, e.g., the residual regression example in Fig. 10.

Table R6. Characteristics of DEM pair GrIS-1.

Pair ID	Data	Date	Res. (m)	Scene ID	MedAD (L23–L57)
GrIS-1	ASTER	25 July 2019	30	AST14DEM.003:2344943025	4.237–3.657 (13.7%)
	ASTER	26 Aug 2019	30	AST14DEM.003:2346334895	

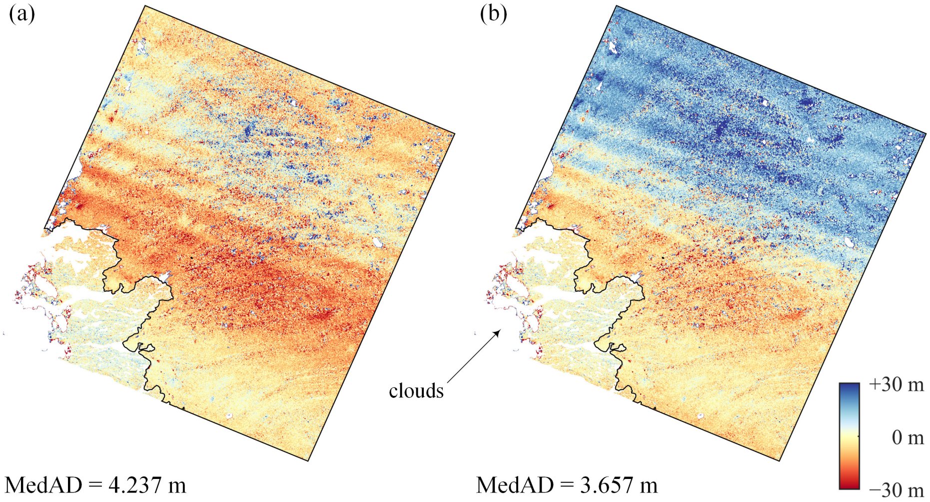

Figure R8. Co-registration results of DEM pair GrIS-1: L23 (a) and L57 (b).

4/ I was also a bit disappointed to see that the techniques are only apply to ASTER DEMs. This also reduce the scope/relevance of the results.

We have conducted the tests on a variety of DEMs, including ZY-3 DEMs, ASTER DEMs, SRTM DEMs, and Copernicus DEMs. Please see the reply to the previous questions for test results and more details.

Specific comments

L30. I find it unbalanced that three out of four references on the use of DEMs for glacier elevation change mapping are from Chinese colleagues. Others more seminal papers on the topic could be cited here.

We will replace the latter two references with Gardelle et al. (2013) and Pieczonka et al. (2013).

References:

Gardelle, J., Berthier, E., Arnaud, Y., and Kääb, A.: Region-wide glacier mass balances over the Pamir-Karakoram-Himalaya during 1999–2011, The Cryosphere, 7, 1263–1286, https://doi.org/10.5194/tc-7-1263-2013, 2013.

Pieczonka, T., Bolch, T., Junfeng, W., and Shiyin, L.: Heterogeneous mass loss of glaciers in the Aksu-Tarim Catchment (Central Tien Shan) revealed by 1976 KH-9 Hexagon and 2009 SPOT-5 stereo imagery, Remote Sens. Environ., 130, 233–244, https://doi.org/10.1016/j.rse.2012.11.020, 2013.

L31. Leprince et al., is about mapping surface displacements (using Cosi-Corr), I am not sure this reference is appropriate for DEM errors. Can authors double check?

We will remove this reference.

L50. What are these "scenarios"?

We will change it to "images".

L71. Authors need to explain why they need to revisit the Nuth & Kaab's method and why they present in details these flavours of their method. It is not straightforward for the reader what is the aim here. Also because in the end the results are almost identical...

The following sentences will be added at the beginning of Section 2.

> "In this section, we will demonstrate that the method of Nuth and Kääb (2011) and the method of Rosenholm and Torlegard (1988) are theoretically compatible, and the latter can be viewed as an extension of the former by additionally modeling the scale and rotation errors. As the original algorithms in the works of Nuth and Kääb (2011) and Rosenholm and Torlegard (1988) were presented in distinct forms, we will present detailed derivations of the equations used in their algorithms and variants."

Figure 1. I did not really understand the figure because I did not understand what were representing the different letters/segments. Annotation to be clarified.

The meanings of θ, ψ, a, b, c have been annotated in Fig. R9 (highlighted in the green rectangle), X, Y, Z represent the axes along the three dimensions, O,E,G... are points in 3D space, and dH_{XY} and dH_Z are the elevation differences induced by a horizontal shift and a vertical shift, respectively.

Figure R9b (i.e., Fig. 1b in the manuscript) was redrawn from Figure 2 (i.e., Fig. R10) of Nuth and Kääb (2011). It is easy to read but only illustrates the special case when $b = \psi$, where b and ψ are the aspect of the shift vector and the terrain aspect, respectively. Figure R9a presents a 3D illustration of a general case when $b \neq \psi$, which is of importance to interpret the $\cos(b-\psi)$ term in Equation 2 of Nuth and Kääb (2011).

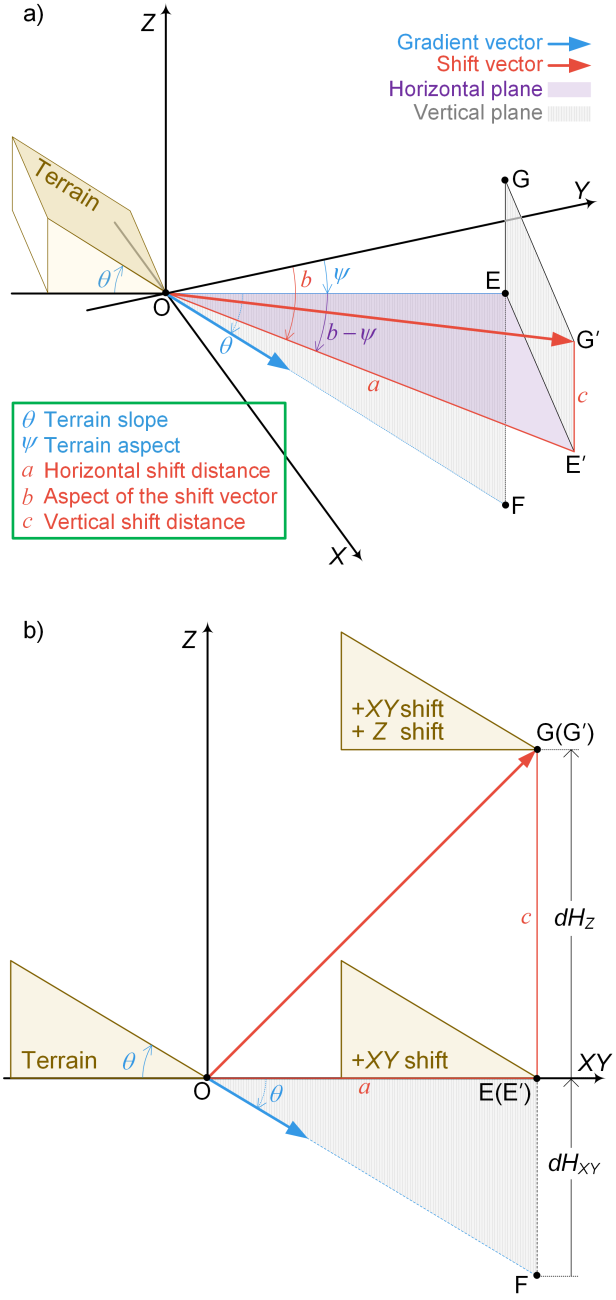

Figure R9. Elevation differences induced by DEM shift: **(a)** 3-D view when $b \neq \psi$. **(b)** 2-D view when $b = \psi$.

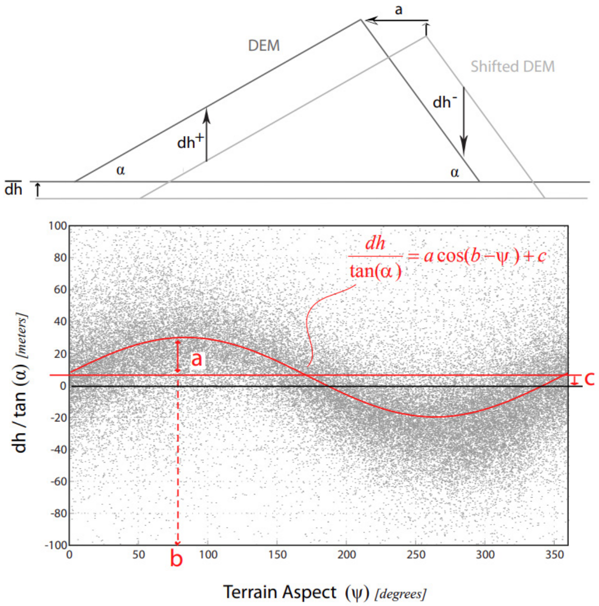

Figure R10. The Figure 2 of Nuth and Kääb (2011).

Figure 2. the terminology "master" and "slave" are not very the best ones for ethical reasons. "Reference" and "secondary" DEMs are better.

We will follow your suggestion.

L204. I do not understand why 23 DEM pairs are first mention and then only 2 DEM pairs are presented in detail in Table 2. Rather include an appendix with the dates and ID of all the DEMs so that the study can be reproduced. Also, as you read in my general comment, a more extensive study using a variety of study sites would be more convincing.

We only displayed the results of 2 DEM pairs to improve the manuscript readability. In the revised version, we will add the information of all DEM pairs in the appendix.

L208. The NDBI index is not often used in the glaciological community so need to be explained a bit more.

As listed in Table R7 (The Table 1 of Nguyen et al., 2021), there are more than 10 bareness indices available in literature. The NDBI index used in this work was adapted from Deng et al. (2015) by replacing the SWIR2 band by SWIR1. Our experimental results show that the revised index performs slightly better than Deng et al.'s version (i.e., the NDSI2 in Table R7) in terms of enhancing bare soil from other land cover features around the periphery of GrIS.

Table R7. The Table 1 of Nguyen et al. (2021).

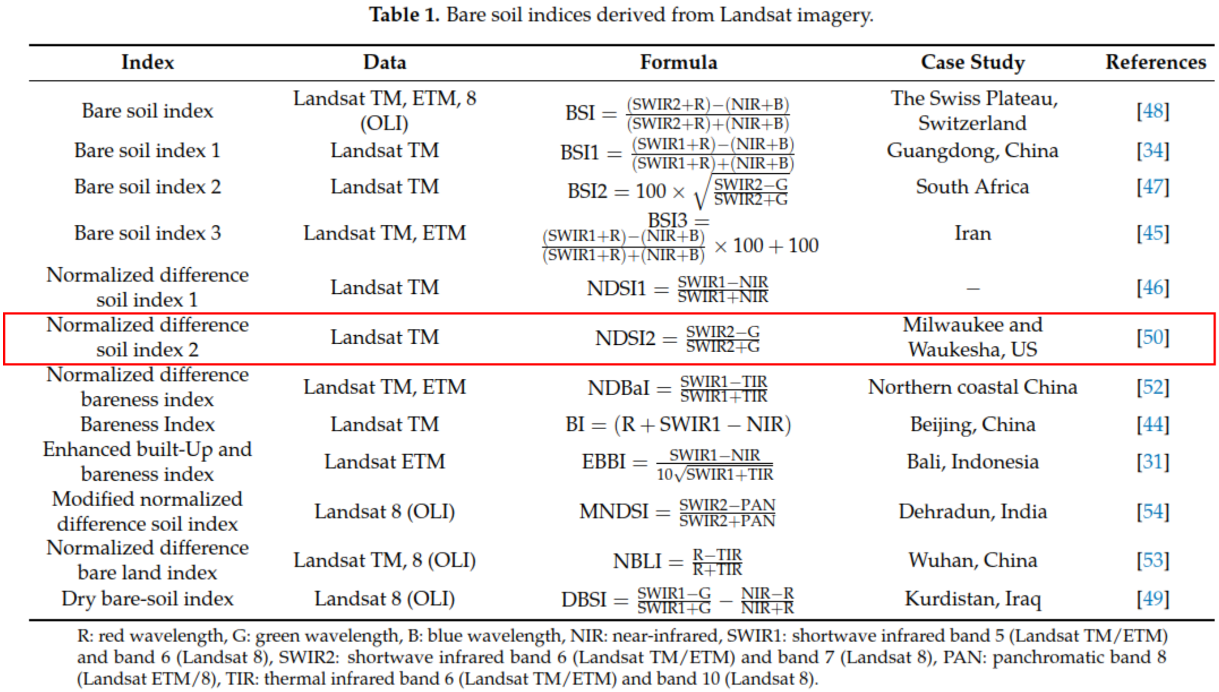

Table 1. Bare soil indices derived from Landsat imagery.

Index	Data	Formula	Case Study	References
Bare soil index	Landsat TM, ETM, 8 (OLI)	$\mathrm{BSI} = \frac{(\mathrm{SWIR2+R})-(\mathrm{NIR+B})}{(\mathrm{SWIR2+R})+(\mathrm{NIR+B})}$	The Swiss Plateau, Switzerland	[48]
Bare soil index 1	Landsat TM	$\mathrm{BSI1} = \frac{(\mathrm{SWIR1+R})-(\mathrm{NIR+B})}{(\mathrm{SWIR1+R})+(\mathrm{NIR+B})}$	Guangdong, China	[34]
Bare soil index 2	Landsat TM	$\mathrm{BSI2} = 100 \times \sqrt{\frac{\mathrm{SWIR2-G}}{\mathrm{SWIR2+G}}}$	South Africa	[47]
Bare soil index 3	Landsat TM, ETM	$\mathrm{BSI3} = \frac{(\mathrm{SWIR1+R})-(\mathrm{NIR+B})}{(\mathrm{SWIR1+R})+(\mathrm{NIR+B})} \times 100 + 100$	Iran	[45]
Normalized difference soil index 1	Landsat TM	$\mathrm{NDSI1} = \frac{\mathrm{SWIR1-NIR}}{\mathrm{SWIR1+NIR}}$	–	[46]
Normalized difference soil index 2	Landsat TM	$\mathrm{NDSI2} = \frac{\mathrm{SWIR2-G}}{\mathrm{SWIR2+G}}$	Milwaukee and Waukesha, US	[50]
Normalized difference bareness index	Landsat TM, ETM	$\mathrm{NDBaI} = \frac{\mathrm{SWIR1-TIR}}{\mathrm{SWIR1+TIR}}$	Northern coastal China	[52]
Bareness Index	Landsat TM	$\mathrm{BI} = (\mathrm{R} + \mathrm{SWIR1} - \mathrm{NIR})$	Beijing, China	[44]
Enhanced built-Up and bareness index	Landsat ETM	$\mathrm{EBBI} = \frac{\mathrm{SWIR1-NIR}}{10\sqrt{\mathrm{SWIR1+TIR}}}$	Bali, Indonesia	[31]
Modified normalized difference soil index	Landsat 8 (OLI)	$\mathrm{MNDSI} = \frac{\mathrm{SWIR2-PAN}}{\mathrm{SWIR2+PAN}}$	Dehradun, India	[54]
Normalized difference bare land index	Landsat TM, 8 (OLI)	$\mathrm{NBLI} = \frac{\mathrm{R-TIR}}{\mathrm{R+TIR}}$	Wuhan, China	[53]
Dry bare-soil index	Landsat 8 (OLI)	$\mathrm{DBSI} = \frac{\mathrm{SWIR1-G}}{\mathrm{SWIR1+G}} - \frac{\mathrm{NIR-R}}{\mathrm{NIR+R}}$	Kurdistan, Iraq	[49]

R: red wavelength, G: green wavelength, B: blue wavelength, NIR: near-infrared, SWIR1: shortwave infrared band 5 (Landsat TM/ETM) and band 6 (Landsat 8), SWIR2: shortwave infrared band 6 (Landsat TM/ETM) and band 7 (Landsat 8), PAN: panchromatic band 8 (Landsat ETM/8), TIR: thermal infrared band 6 (Landsat TM/ETM) and band 10 (Landsat 8).

References:

Deng, Y., Wu, C., Li, M., and Chen, R.: RNDSI: A ratio normalized difference soil index for remote sensing of urban/suburban environments, Int. J. Appl. Earth Obs. Geoinf., 39, 40–48, https://doi.org/10.1016/j.jag.2015.02.010, 2015.

Nguyen, C. T., Chidthaisong, A., Kieu Diem, P., and Huo, L.-Z.: A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8, Land, 10, 231, https://doi.org/10.3390/land10030231, 2021.

L215. I do not understand how this 3-sigma rule is applied to check outliers from the classification. Authors need to elaborate more.

We will edit this sentence as follows:

> " …and a three-sigma rule (i.e., more than three times the standard deviation) was employed on the elevation differences between two DEMs to remove erroneous data caused by misclassification of unstable terrain areas."

Figure 4 does not really bring much. I think it will be pretty obvious to most readers and can be explained in a few sentences in the text.

We will delete Figure 4 and the relevant descriptions.

L239. Example of why the application of the algorithm to a greater diversity of images is needed.

So far, the simplified and standard versions of the method of Nuth and Kääb have been compared with more than 200 DEM pairs from different sources, and the experimental results show that their performance is always close to each other. However, we still cannot reach a more definitive conclusion, because it lacks theoretical grounding.

L302. English is not really correct I think. Check

We will remove this sentence.

L320. This statement (and the rest of the paragraph) about extrapolation error and elevation error as a function of altitude comes a bit out of nowhere. Why discussing extrapolation when this was not mentioned before.

Please see the response to the third question of Major comments.

L328. This is exactly what the revised study should do: include cases where bare terrain is rare and see how the different methods compare.

There are two cases where bare terrain is rare:

1) Stable regions are fairly evenly distributed throughout the image

   Compared to the L23 method, L57 usually yields a higher co-registration accuracy by additionally modeling the scale and rotation errors (refer to Fig. R3).

2) Stable regions are located on one side or even on one corner of the image

   Due to data extrapolation, both the L23 and L57 methods are prone to yield unreliable co-registration results (refer to Fig. R8).

L349. I found the improvements rather modest. "outperformed" is a bit overselling.

The additional test results for HMA and SALP showed that the L57 algorithm could improve the co-registration accuracy by up to 83.3% more than the L23 method. Here we illustrate two examples in HMA (Figs. R11 and R12), where the DEM information used is shown in Table R8.

Table R8. Characteristics of the two DEM pairs.

Pair ID	Data	Date	Res. (m)	MedAD (L23–L57)
HMA-5	Copernicus	2011 to 2014	30	4.680–0.780 (83.3%)
	ZY-3	8 Oct 2017	30	
HMA-6	Copernicus	2011 to 2014	30	2.476–1.429 (42.3%)
	ZY-3	2 Dec 2016	30	

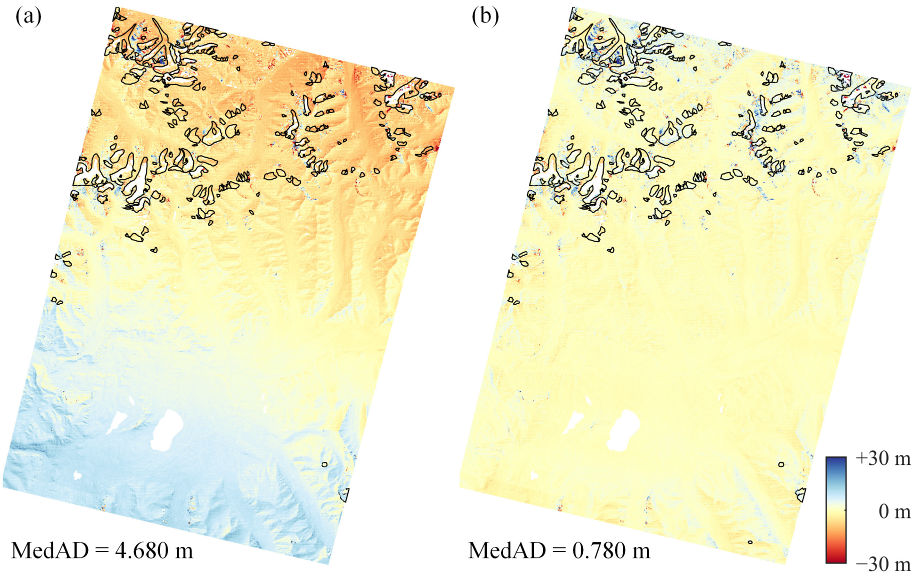

Figure R11. Co-registration results of DEM pair HMA-5: L23 (a) and L57 (b).

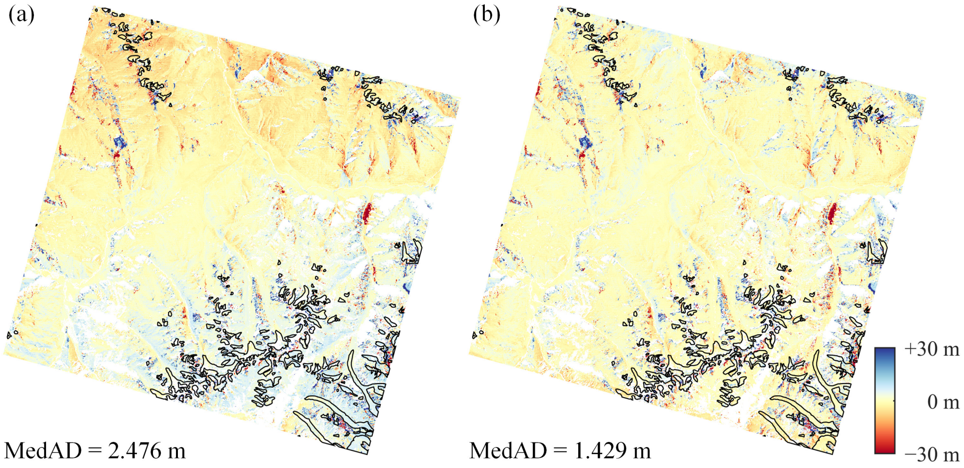

Figure R12. Co-registration results of DEM pair HMA-6: L23 (a) and L57 (b).