Response to Reviewer #2 on "Co-registration and residual correction of digital elevation models: A comparative study"

Comment received: 11 Jan 2023

Key:

Reviewer comment (black)

Response (blue)

This paper deals with the co-registration of DEMs for determining surface elevation changes by DEM differencing. The differences between DEMs are affected by random measuring errors and various systematic errors due to imperfections of the sensors. The success of the simple differencing method depends on how well the systematic errors can be determined and be removed. The paper begins with comparing 3 variants of the Nuth and Kääb method with the lessor known method proposed by Rosenholm and Torlegard. The authors then introduce a non-parametric residual correction model and present results from a few experiments with ASTER DEMs of Western Greenland.

Thank you for your constructive comments and suggestions.

Major Comments

I agree with the authors that the Nuth and Kääb method is predominently used for co-registering ASTER DEMs by the cryospheric research community. The method has been improved over the years, particularly with handling systematic errors of ASTER DEMs (see reference Luc Girod et al., 2017). If one wants to correct Aster DEMs, as the authors do, then I think one should start the process on the level of 2017 (see reference above) and not on the original level of 2011. The main reason is that in the 2017 version a new DEM is computed (MMASTER) with superior image matching that increases the reliability of the DEMs.

The research objects of our work are the DEM co-registration and residual correction methods utilized by cryospheric research for DEM differencing. In the previous manuscript, we chose the ASTER DEMs as a test dataset, which existed obvious complex systematic errors induced by satellite altitude jitter, enabling us to simultaneously design co-registration and residual correction methods. In the additional experiment, we tested DEMs widely used by cryospheric research, including ZY-3 DEMs, SRTM DEMs, and Copernicus DEMs (please see the response to Reviewer #1 for specific experimental results).

The research of Girod et al. (2017) includes two points: the process from a single stereo pair (one ASTER L1A scene) to an ASTER DEM and the correction of DEMs differences (dDEMs). Since we are not involved in the ASTER L1A data processing, the first point is not compared. For the dDEMs process, Girod et al. (2017) first adopted the DEM co-registration method described by Nuth and Kääb (2011), then proposed a parametric regression model (the sum of the sinusoidal functions) to correct the jitter-induced bias, which we have compared in the previous manuscript (Sect. 3.1. Parametric regression).

Another comment is related to the 'master/slave' concept to co-register DEMs. As is apparent from Table 2, the authors use as a master another ASTER DEM. That makes all computations relative to the master (which is essentially affected by the same systematic sensor errors as the slave) and thus precluding comparisons to an accepted ground reference system. In the area of the test site in Greenland are alternative sources that would be much better suited for serving as a master DEM (e.g. ICESat-2, World View DEMs, ATM airborne laser altimetry).

We agree that refined ASTER DEMs, ICESat-2, and ArcticDEM would be better alternatives for the accurate estimation of glacier changes. However, it should be noted that the main goal of this work is not to access the accuracy of new DEM datasets, but to compare different DEM co-registration algorithms. So far, our experiments have been carried out on four different DEM sources, including ASTER, ZY-3, SRTM, and Copernicus DEMs. The georeferencing errors in these DEMs are commonly larger than those in up-to-date data, but much smaller than those in the historical data before 2000, which is adequate for the comparative analysis of different DEM co-registration methods.

The research results presented in this paper include a comparison between the methods proposed by Nuth/Kääb and Rosenholm/Torlegard. The results of these comparisons can be found in Table 3. The numbers confirm what other researchers have found. The question I have is the definition of AverageMed which is used throughout the paper. How does it compare with more traditional statistical error measures such as mean, median, standard deviation?

The mean and median are measures of location, while the standard deviation is a measure of scale (Rousseeuw and Hubert, 2018). There are two versions of MedAD (the median of all absolute deviations, also abbreviated as MAD) used in literature, i.e., MedAD around the median (Mcmillan et al., 2019) and around zero (Shen et al., 2021).

The MedAD around the median is a measure of scale and can be seen as a robust version of the standard deviation. It is calculated as follows:

$$\text{MedAD} = 1.4826 * \underset{i=1,\dots,n}{\text{median}}\left(\left|x_i - \underset{j=1,\dots,n}{\text{median}}(x_j)\right|\right) \tag{1}$$

where $x = H_{\text{Master}} - H_{\text{Slave}}$ in our manuscript. The constant 1.4826 is a correction factor which makes the MedAD consistent with the standard deviation at Gaussian distributions (Rousseeuw and Hubert, 2018).

In our manuscript, the MedAD is calculated around zero, and the constant 1.4826 is omitted:

$$\text{MedAD} = \underset{i=1,\dots,n}{\text{median}}\left(\left|x_i\right|\right) \tag{2}$$

This form of MedAD is a combined measure of location and scale, and it can be used as a robust alternative to the Root-Mean-Square Deviation (RMSD).

As shown in the last four rows of Table R1, the ratio of the standard deviation (Std) to the MedAD around zero is very close to 1.4826 in our experiments, because the distribution of DEM co-registration residuals is often nearly Gaussian with a zero mean.

Table R1. Co-registration results obtained with the 23 DEM pairs.

Method	ID	Average Median (m)	Average MedAD (m)	Average Mean (m)	Average Std (m)
Before co-registration	—	−3.391	12.043	−3.504	12.604
Nuth and Kääb standard version	N23	0.045	7.170	−0.014	10.910
Nuth and Kääb simplified version	N13	0.036	7.163	−0.017	10.913
Nuth and Kääb linear version	L23	0.045	7.170	−0.014	10.910
Rosenholm and Torlegard	L57	0.005	6.839	0.002	10.484

References:

McMillan, M., Muir, A., Shepherd, A., Escola, R., Roca, M., Aublanc, J., Thibaut, P., Restano, M., Ambrozio, A., and Benveniste, J.: Sentinel-3 Delay-Doppler altimetry over Antarctica, The Cryosphere, 13, 709–722, https://doi.org/10.5194/tc-13-709-2019, 2019.

Rousseeuw, P. J. and Hubert, M.: Anomaly detection by robust statistics, WIREs Data Mining and Knowledge Discovery, 8, e1236, https://doi.org/10.1002/widm.1236, 2018.

Shen, X., Ke, C.-Q., Yu, X., Cai, Y., and Fan, Y.: Evaluation of Ice, Cloud, And Land Elevation Satellite-2 (ICESat-2) land ice surface heights using Airborne Topographic Mapper (ATM) data in Antarctica, Int. J. Remote Sens., 42, 2556–2573, https://doi.org/10.1080/01431161.2020.1856962, 2021.

Another conclusion the authors make is that GAM spline fitting can be used to reduce complex systematic errors that are still present after geo-referencing. These research results are OK but limited to a specific sensor (ASTER, 25 years in space, outdated technology, complex suite of systematic errors that change in time). Moreover, since GAM spline fitting seems to play an important role in this paper I would strongly suggest to cover it in more detail and provide readers with explanations why you choose it.

Thank you for your suggestion. More technical details of the GAM spline fitting algorithm will be included in the revised manuscript.

Our experiments so far have been conducted on more than 200 DEM pairs from four different sources including ASTER, ZY-3, SRTM, and Copernicus DEMs, located in the Greenland Ice Sheet (GrIS), High Mountain Asia (HMA), and southern Alps (SALP). The experimental results show that strong jitters can only be observed in the ASTER DEMs of GrIS.

Though the paper is written well I have doubts that it is suitable for publishing in this journal in its current form. The methodology presented in this paper should be made more relevant to cryospheric research or might be better suited for a journal that is more focused on new methods and algorithms.

The paper of Nuth and Kääb (2011) was originally published in "The Cryosphere". Their algorithm is currently the most commonly used DEM co-registration method in glacial studies, but it has not been widely adopted in other geoscience applications. Our work aims to present a deep investigation on the algorithm of Nuth and Kääb, and, therefore, we submitted our manuscript to the same journal, "The Cryosphere".

As suggested by you and Reviewer #1, we will provide more experiments to investigate the choice of DEM co-registration algorithms on glacier change estimation. For more details, please see our replies to Reviewer #1.

Reference:

Nuth, C. and Kääb, A.: Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change, The Cryosphere, 5, 271–290, https://doi.org/10.5194/tc-5-271-2011, 2011.