My detailed comments on the manuscript can be found in the attached annotated pdf. The manuscript discusses a relevant topic (accuracy of different DEMs on the Tibetan Plateau and impact on ice thickness reconstruction) and overall I find that the methods are robust and useful conclusions are drawn. I do have quite a few comments (see attachment), most prominently related to 1) the way differences are calculated [I suggest to subtract the ICESat-2 data from the DEM datasets rather than the other way around; this would avoid a lot of confusion], 2) the large number of figures [some suggestions for removing figures are given], 3) clarification or explanation [in many places details about the methods and presented figures and tables are missing; additionally, numerous textual revisions may help to clarify the content], 4) transferability of results to other regions [some discussion/speculation on how the presented findings on DEM and ice thickness comparison may translate to other mountainous regions would be interesting to add].

Response: Many thanks to the reviewer for the critical comments. We have carefully addressed all the comments. 1) And we have changed the way differences are calculated, Tables and Figures are all updated. 2) We have deleted two figures and simplified two figures. 3) Details about the method and figures are added. 4) Further discussions are also given in the manuscript. Our point-by-point responses are attached below in blue, while the original reviewers' comments are in black.

1. "by ICESat-2" can be removed here

Response: Deleted

2. Line 16 Is this really considered?

Response: Yes, glacier elevation change data from Shean et al. (2020) is adopted in the evaluation, to distinguish the effect from glacier dynamics.

3. Why are the acquisition dates relevant in this sentence?

Response: Because glaciers elevation is changing all the time, if DEM across the large region is acquired in different periods, the estimated ice thickness by this DEM would not represent the real state of glacier. We revised this sentence.

"Considering the necessity of DEMs with consistency acquisition dates, NASADEM would be a best choice for ice-thickness estimates over the TP."

4. offered-> offers"

Response: Done

5. The effect of DEM grid resolution, giving more detailed slope information, on thickness estimation could be introduced here.

Response: We add one sentence here.

"Therefore, the DEM grid resolution could influence the thickness estimation from GlabTop, more detailed slope information could be provided by higher resolution DEM."

6. Slope or height?

Response: that's elevation error, we have corrected it.

7. Could be worth referring to Koldtoft et al. (2021) here.
Response: Done.

8. Could be good to shed some light on the relative significance of DEM errors compared to other uncertainties (particularly model physics).
Response: In this paragraph, we emphasized the fundamental role of DEM in determining the model physics, such as center flow lines, shear stress, apparent mass balance in line 42-44. Therefore, we add the sentence here as a conclusion after the literature review.
"DEM errors influence the determination of model physics and the final model outcomes."

9. Line 65 physiognomy?
Response: replaced with "landform"

10. Line 70 **something wrong with the grammar here.**
Response: We revised the sentence.
"publicly available DEM with high resolution**"**

11. I suppose the spatial coverage of ICESat-2 limits its use as a distributed DEM? It could be good to mention this, otherwise the reader may wonder why ICESat-2, rather than any of the 6 other DEMs, is not directly considered as the best DEM for thickness inversion.
Response: Yes, the reviewer is right, we revised this sentence.
"against ICESat-2 data which has been proven to have a high vertical accuracy and resolution but with sparse tracks (Fig. 1)."

12. That seems a bit superfluous after what is done in the previous assignment...
Response: Deleted.

13. Any indication of accuracy outside the Antarctic?
Response: We add one study in Qilianshan.
"The ATL06 product has better than 5 cm height accuracy and better than 20 cm surface measurement precision in the Antarctic and Qilianshan (Brunt et al. 2019; Zhang et al. 2020)."

Zhang, Y., Pang, Y., Cui, D., Ma, Y., & Chen, L. (2020). Accuracy Assessment of the ICESat-2/ATL06 Product in the Qilian Mountains Based on CORS and UAV Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 1558-1571.

14. Accurancy ->Accuracy
Response: Corrected.

15. Why fill gaps? It makes the comparison less independent.

Response: The data was provided with gaps filled by JAXA officially. In this article, we try to directly know the quantity of DEM product acquiring from the official. We didn't fill any gaps in this article.

16. This is not a sentence. Furthermore, it would be better to split this into 4), 5) and 6)

Response: We revised this paragraph. And split this into 4), 5) and 6)

17. Line 171 Farinotti-> Farinotti

Response: Corrected**.**

18. Line 174 what value is chosen here?

Response: fsl=0.8, we have added this value.

19. Line 179 Not sure why this is relevant here.

Response: Deleted**.**

20. Line 181 From what experience? Please clarify.

Response: The parameter is determined from previous study; we have revised this sentence.
" which is determined from Huss and Farinotti et al. (2012)."

21. Line 188 This is inverting the shallow ice approximation, right? If so, that could be mentioned.

Response: Yes, the reviewer is right, we have added this information.

"The Ice Thickness Inversion Based On Velocity (ITIBOV) model is inverting from the shallow ice approximation, it obtains the ice thickness by combining the surface velocity field with the Glen ice flow law (Gantayat et al. 2014; Glen 1955):"

22. Line 191 What year or period of velocity data are used?

Response: Mean velocity over 1985-2019 was used. We have added this information.
"We used the mean velocity over 1985-2019 from ITSLIVE dataset (Gardner 2019)"

23. Line 195 So some other shared parameters are different?

Response: Yes, some parameters are unique for specific model. For example, the apparent mass balance parameter is only used in HF model.

24. Line 200 one thickness per point? How is this weighting determined?

Response: We add these sentences here to explain our method to determine the weight.
"First, the ensemble ice thickness was the sum of the four models with four weights w1, w2, w3, and w4, respectively. The sum of four weights equals to 1. 70% of the GPR result are adopted as calibration data. 30% of the GPR result are adopted as validation data. Then, the four weights iteratively changed to achieve the minimal mean absolute error between calibration data and model result. Finally, the MAE between ensemble ice thickness and validation data are calculated."

25. Why no R values?

Response: R is usually used to estimate the linear correlation between variables. Here we used NMAD comparing with ME to assess the influence from extreme errors as Höhle and Höhle (2009) and Gdulová et al. (2020). Also the R^2 in Figure3 shown little difference among DEMs. Therefore, we didn't use R in this study.

26. Line 205 It is very confusing that the elevation differences and the ME, MAE are all defined by taking H_ICESat-2 minus H_DEM. Naturally, I would expect the opposite (H_DEM - H_ICESat-2) and I would suggest to change this. That would change the sign of all the "differences" presented in Figs. 3-7, but it would make more sense since the DEMs are compared against the ICESat-2 data, not the other way around.

Response: We redo the analysis as Reviewer's suggestion. The Tables and Figures are all updated

27. These do not seem to be excluded yet in Fig. 3 and 4. Please mention.

Response: Fig.3 shown the differences of different filter ranges. Only the values in the range of 4std are used for further analysis. The fit result and overall difference statistics in Fig.3 and 4 are all based on filtered data.

28. Line 210 All outliers in one DEM or all of them?

Response: Ration of outliers relative to each DEM is less than 1%. And this sentence is revised.

"The four standard deviations (that is 4 std) was chosen to filter on the differences between ICESat-2 and DEMs to filter out extreme outliers. Ration of excluded outliers relative to record of each DEM is less than 1% .**"**

29 Line 211 = regular? What is meant?

Response: We revised this sentence.

"Overall, there is no irregular"

30 Line 215 The ME is a more direct indicator for bias / systematic shifts

Response: Yes, the review is right. The ME is a more direct indicator that we used it in the follow analysis. Here, we set the slope coefficient close to 1, so the intercept value is equal to the ME as shown in Figure 4.

31. Line 225 but this is not shown right?

Response: By comparing the ME and Median, we found that they are almost same, which indicated that the influence from extreme value are little after the application of 4 std filter.

32. Line 295. please clarify what it exactly implies that STD is substantially larger than NMAD for Tandem-X.

Response: This indicated that outliers and noise may exit in the TanDEM and we discuss this in the Sect 4.1 and Figure 12b. That indicate that obvious errors exit in the steep region in the TanDEM product.

"indicating larger discrepancies due to the DEM errors and noise**"**

33. Line 225.It could be worth mentioning that the mean error (ME) is something that can easily be corrected for by applying a bias correction. With that in mind, the STD (rather than RMSE) might be the best measure of performance? This still gives the NASADEM as the best performer, but also AW3D30 has a low STD.

Response: Yes, the reviewer is right. When calculating the ME, the positive and negative biases cancel each other, making the error smaller. We used STD not only to prevent this, but also to measure the variation or dispersion of the error. We add one sentence at the end of Section 2.4 "When calculating the ME, the positive and negative biases cancel each other, making the error smaller; Therefore, the STD together with ME could be a complementary indicator for assessment."

34. Line 235.It is somewhat striking that four of the DEMs have a ME that is between -31 and -33 m. Is there an obvious explanation for this large similar bias with the ICESAT-2 data? Based on Fig. 5 it seems that even the spatial distribution of the bias is very consistent between these four DEMs.
Response: We checked the references of six DEMs and found that their references are different. AW3D30, SRTM-GL1, SRTM v4.1 and MERIT are above EGM96 geoid. And ICESat-2, NASADEM and TanDEM is above WGS84 ellipsoid (Table 1). We have unified them to WGS84 ellipsoid. All figures and tables are updated.

35. Line 275.Is the effect that is visible in Fig. 6c), with most negative values at low elevations, the effect of more rapid glacier thinning at these elevations?
Response: Yes, apart from the elevation error in DEM, we thought that this is due to effect of rapid glacier thinning. The error in the low elevation region (Zone 1) is largely reduced after removing the effect of glacier elevation change (Table 3).

36. Line 280. Is there really a need to assign ablation zone, accumulation zone and transition zones here?
Response: Yes. We think it is necessary. The glacier elevation change is always changing and has different characteristics in ablation and accumulation zones. So, assessments in previous studies exclude the glacier terrain. The six DEMs in this study are acquired in different months and years. The glacier elevation change indeed influences the elevation differences between DEMs and ICESat-2. We divide the glacier terrain into 3 zones to estimate the effect of glacier dynamic.

37. Line 286.This seems counterintuitive to me. The six DEMs are all older than the ICESat-2 data, meaning that the ablation area in the six DEMs is likely to have higher elevations than in the ICESat-2 data (assuming the ablation areas have thinned). That would give a more positive difference in zone 1 and 2 than in zone 3 and 4, where less surface height changes over time are expected.
Response: The reviewer is right. Previously, we subtract the DEMs elevation from ICESat-2, so there is a more negative difference in zone 1 and 2 than in zone 3 and 4. In this version, we have subtracted ICESat-2 from the DEMs elevation as Reviewer's suggest. Now, there is a more positive difference in zone 1 and 2 than in zone 3 and 4.

38. Line 295. It would be good to reformulate this. What I think Fig. 7 mainly shows is that the observed shift in the difference from zone 1 to zone 4 is a sign that thinning between the time of collection of the six DEMs and the ICESat-2 data is most pronounced in zones 1 and 2. This could be mentioned here. To some extent this effect can already be seen in Fig. 6c as well if I am correct.
Response: Yes, the reviewer is right. We add one more sentence to reformulate it.
"The observed shift in the difference from zone 1 to zone 4 is a sign that thinning or accumulation between the time of collection of the six DEMs and the ICESat-2 data is most pronounced in zones

1 and 2."

38. Line 302 The performance of the ITIMs for the different DEMs will depend on the model parameters.

Response: Yes, model parameters would influence the model output. However, we adopt same model parameters for the specific one model, but using different DEMs. By this, we explored the influence of DEMs on the outcome of the ITIMs. And we found that even with the same parameters, the same model using different DEMs have different outcomes (Figure 8 and 9). The DEM indeed influence the performance of the ITIMs.

39．Line 314 criterion-> criteria

Response: Corrected.

40. Line 326 A quick summary of how the weights for the different models are determined would be helpful.

Response: We add this in Section 2.3

"First, the ensemble ice thickness was the sum of the four models with four weights w1, w2, w3, and w4, respectively. The sum of four weights equals to 1. 70% of the GPR result are adopted as calibration data. 30% of the GPR result are adopted as validation data. Then, the four weights iteratively changed to achieve the minimal mean absolute error between calibration data and model result. Finally, the MAE between ensemble ice thickness and validation data are calculated.**"**

41 Line 330 Why not as a final experiment make a composite bed estimate with only the best four ITIM-DEM combinations?

Response: Here, we try to estimate the effect of DEM on ITIMs. So only the ice thickness from four ITIMs using same DEM is ensembled.

42. Line 355 Fix column width

Response: Done**.**

43. Might be good to split into two subsections (4.1 and 4.2) focusing on glacier elevation change and terrain separately.

Response: Done**.**

44. Line 366 Please clarify "from the same original data,"

Response: We means that the NASADEM and SRTM-GL1 are both generated from NASA's Shuttle Radar Topography Mission. We have deleted this sentence in this version, because the large difference is due to the reference difference, not the vertical shift.

45.Line 381 This error correction method should have been introduced in the methods. The related results (Table 3) belong to the Results rather than the Discussion. Also, more details are needed. Have the DEMs been corrected or the ICESat-2 dataset?

Response: We add more details about this method in Section 2.4. Only the ICESat-2 Data is corrected by the glacier elevation change data. Table 3 is a direct proof of the effect of glacier

elevation change on the assessment of DEM. We thought it may be included in this Discussion.
“Glacier elevation changed a lot at -21－17m/yr over the TP during 2000-2018 (Shean et al. 2020). Therefore, the disparity of acquiring date between ICESat-2 and six DEM (Table 1) could introduce large error due to this glacier dynamic. TanDEM-X and AW3D30 are acquired in different months and years (Table 1), it’s hard to analyse the impact of glacier dynamic on accuracy assessment. However, the other four DEMs are produced from NASA's Shuttle Radar Topography Mission during the 11-day mission in February 2000. We selected ICESat-2 data acquired in February 2019 and 2020. Then the glacier elevation dynamic magnitude during February 2000 and February 2019/2020 are subtracted from the selected ICESat-2 elevation based on the mean glacier elevation change data from Shean et al. (2020). By comparing the elevation from the four DEMs and adjust ICESat-2, we could exactly know the impact from glacier dynamic.”

46. Line 389 The figure is confusing. What is adjusted ICESat-2? It would make more sense to adjust the six DEMs and compare with the original ICESat-2 only. With DEMs collected in different years it is currently not clear for what period the ICESat-2 data are adjusted. The comparison currently does not make sense to me.
Response: Because the collect year of NASADEM, SRTM GL1 , SRTM v4.1 and MERIT is on Feb. 2000. Here we adjust the one track ICESat-2 data on the glacier to the year of 2000 using the glacier elevation change data over 2000-2018 from Shean et al. (2020). By this figure, we want to conclude that the difference between DEMs and ICESat-2 would be reduced after adjusting ICESat-2 elevation. However, in this version, we removed this figure as reviewer’s suggest.

47. Line 399 Not sure what you want to say here. I suppose the ICESat-2 dataset gives average elevation over the 2018-2020 period, so no seasonal dependence. Am I right that the other DEMs reflect one point in time, e.g. Feb 2000. That can of course give some bias. It could be good to rephrase a bit here.
Response: The reviewer is right. We rephrased this sentence.
“When taking all points from different seasons into consideration, ICESat-2 dataset gives average elevation over the 2018-2020 period, the seasonal effects could also partly cancel each other out.”

48. Line 403 Any rough idea on magnitude of this error?
Response: Corrections of these two errors require information about cloud structure and ice-surface conditions that are not available when ATL06 is processed. It remains an active avenue of research.

49 Line 409 Why show this? It is hard to see any differences between the panels this way.
Response: By this we try to show that there are serious errors in the steep region in TanDEM in ROI A region denoted in this figure. Of course, difference among the other five DEMs show little difference.

50. Line 416 The main thing I see in Fig. 13 is that there are many more measurements with north aspect, so both for steep and gentle slopes.
Response: If we fixed slope axis, we can find that there more measurements in north aspect at the same slope. That’s to say that, there more measurements with steep slopes in the north aspect.

51. Line 433 Could be worth highlighting which DEM suffers most/least from slope effects. Fig 10 gives an indication but only focuses on extreme outliers rather than mean differences. I am also curious how accurate ICESat-2 is in steep terrain. I can only find estimates for (flat) Antarctica in this study.

Response: We add one sentence in Setction 3.2.

"Overall, relative to the other DEMs, AW3D30 and NASADEM behaves best against slope in terms of spread and median value."

A study in Qilian Shan in north TP shown a less than 20 cm accuracy. We have added this reference.

52. Line 433 It would be good to explain briefly what is meant with misregistration.

Response: We added one sentence to explain "misregistration"

"Pixel of different DEMs at same location may mismatch each other."

53.Line 439 I am currently not sure how to interpret Fig. 14. What are "offset pixels"? Is it a measure of how many pixels (of 30 or 90 m) a DEM is shifted relative to ICESat-2 within a 1 by 1 degree grid cell? Please explain.

Response: Yes, the reviewer is right. We estimate the offset distance of DEM by 1×1 degree relative to ICESat-2. Then this offset distance is converted to offset pixels according to the resolution of a DEM. After unifying the reference, the shift pixels are all within one pixel, and show little spatial difference. We have updated Figure 14.

54. remove "except for" since these seem to be the DEMs that are affected. (?)

Response: We have deleted this paragraph after updating the data.

55. I can't follow this reasoning. Please clarify.

Response: We have deleted this paragraph after updating the data. The shift pixels are all within one pixel and has little influence on the assessment.

56. Line 459 This is not shown in Fig. 16. Maybe in Fig. 14, but only in red?

Response: We have deleted this Figure after updating the data. The shift pixels are all within one pixel and has little influence on the assessment.

57. I find this section rather chaotic and parts of it are of limited value. The discussion of Fig. 9 is relevant and should be kept, but the discussion of slope and elevation range perturbations in the first paragraph and Fig. 16 do not add much new insight and in my opinion can be removed. The experiments are very hypothethical as a homogeneous slope perturbation and a elevation range perturbation are not something that one can expect to happen for real DEMs. From the methods in Section 2 it can already be concluded which models would be most sensitive to certain types of terrain errors.

Response: We discussed the reviewer's suggestions. Yes, the reviewer is right, we can conclude that which model would be sensitive to certain types of terrain errors from the methods in Section 2. However, we could not conclude the degree of sensitivity of different models to terrain factors. For example, we know that GlabTop2 would be sensitive to elevation and slope, we found that small size glaciers would be more sensitive than big size glacier in our test. In the formula of HF and OGGM in the method, we may guess that these two models would be sensitive to the slope, but actually they have a good robustness to the accuracy of input DEM. We have deleted some irrelevant portions and readjusted the paragraph structure to solve reviewer's comments.

58 Line 464 where is this shown? It is not in Fig. 10 it seems.
Response： It's shown in Figure 8.

52 Line 465 and 45 % of what?
Response：We rephased this sentence.
"Generally, the outcome with GlabTop2 and ITBOV using 30-m DEMs is 51% and 43% better than with the 90-m DEMs in mean error, respectively."

53.Is this really needed? I would suggest to remove this part.
Response： Removed.

54.Line 505 please clarify or reformulate.
Response： We reformulate it as follow.
"When the results from different models are ensembled, "

55.Line510 I am missing discussion on how the results found in this study can potentially be of use for others using these DEMs, potentially as input in an ice thickness estimation model, in other mountainous (glacierized) regions than the Tibetan Plateau. Is it likely that the same conclusions could be drawn in other regions? Why (not)?
Response： We add further discussion here.
"However, it should be noted that the result may be not suitable for studies in other glacierized mountainous regions. Because various errors exit in DEMs, such as speckle noise, stripe noise and absolute bias; they behave different across the Earth (Yamazaki et al. 2017). But our method to assess the accuracy of DEMs is repeatable in different regions, combining with the recent released glacier elevation change data on Earth (Hugonnet et al. 2021)."

Towards ice thickness inversion: an evaluation of global DEMs ~~by ICESat-2~~ in the glacierized Tibetan Plateau

Wenfeng Chen[1,3], Tandong Yao[1], Guoqing Zhang[1], Fei Li[1,3], Guoxiong Zheng[2,3], Yushan Zhou[1], Fenglin Xu[1,3]

[1] State Key Laboratory of Tibetan Plateau Earth System Science, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China
[2] Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China
[3] University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence to: Wenfeng Chen (chenwf@itpcas.ac.cn)

Abstract. Accurate estimates of regional ice thickness, which are generally produced by ice-thickness inversion models, are crucial for assessments of available freshwater resources and sea level rise. Digital elevation model (DEM) derived surface topography of glaciers is a primary data source for such models. However, the scarce in-situ measurements of glacier surface elevation limit the evaluation of DEM uncertainty, and hence its influence on ice-thickness modelling over the glacierized area of the Tibetan Plateau (TP). Here, we examine the performance over the glacierized TP of six widely used and mainly global-scale DEMs: AW3D30 (30 m), SRTM-GL1 (30 m), NASADEM (30 m), TanDEM-X (90 m), SRTM v4.1 (90 m) and MERIT (90 m) by ~~using~~ comparing with ICESat-2 laser altimetry data while considering the effects of glacier dynamics, terrain, and DEM misregistration. The results reveal NASADEM as the best performer, with a small mean error (ME) of ~~–1.0~~0.9 and a root mean squared error (RMSE) of 12.6 m~~. A systematic vertical offset existed in~~, followed by AW3D30 (~~–35.3~~2.6 ME and ~~34.~~11.3~~9~~ m RMSE)~~, although it had a similar relative accuracy to NASADEM (~13 m STD)~~. TanDEM-X also performs well (–0.1 ME and 15.1 m RMSE), but suffers from serious errors and outliers on steep slopes. SRTM-based DEMs (SRTM-GL1, SRTM v4.1, and MERIT) (~~all ~36 m~~13.5-17.0 m RMSE) had an inferior performance to NASADEM. Errors in the six DEMs increased from the south-facing to the north-facing aspect and become larger with increasing slope. Misregistration of DEMs relative to ICESat-2 footprint in most glacier areas is small (less than one pixel). An intercomparison of four ice-thickness models: GlabTop2, Open Global Glacier Model (OGGM), Huss-Farinotti (HF), Ice Thickness Inversion Based on Velocity (ITIBOV), show that GlabTop2 is sensitive to the accuracy of both elevation and slope, while OGGM and HF are less sensitive to DEM quality and resolution, and ITIBOV is the most sensitive to slope accuracy. Considering the necessity of DEMs with consistency acquisition dates~~Considering the inconsistency of DEMs acquisition dates~~, NASADEM would be a best choice for ice-thickness estimates over the TP~~, followed by AW3D30, and TanDEM-X (if steep and high elevation terrain can be avoided)~~. Our assessment figures out the performances of mainly global DEMs over the glacierized TP. This study not only avails the glacier thickness estimation with ice thickness inversion models, but also ~~offered~~ offers references for other cryosphere studies using DEM.

1 Introduction

The Tibetan Plateau (TP), which includes the Pamir, Hindu Kush, Karakoram, Himalaya, and Tibet regions, covers an area of ~3 million km^2 and has a mean elevation of more than 4000 m a.s.l. (Fig. 1). It accounts for more than 82% of the Earth's land surface area above 4000 m a.s.l. (Fielding et al. 1994), and is often referred to as the Third Pole of Earth or the Asian Water Tower (Yao et al. 2012) due to its high elevation and abundant water resources in the form of glaciers, snow, permafrost, lakes, and rivers. The TP has a glacierized area of ~8.3×10^4 km^2 (RGI Consortium, 2017) with an ice volume of ~6.2×10^3 km^3 (Farinotti et al. 2019), mainly distributed in the Karakoram and Himalaya regions.

Ice thickness is a crucial parameter for assessing the contribution of glaciers to global sea level rise (Kraaijenbrink et al. 2017), quantifying regional water availability (Huss and Hock 2018; Immerzeel et al. 2020), and evaluating cryosphere-related hazards (Linsbauer et al. 2016; Zheng et al. 2021). In the TP, owing to the lack of in-situ ice thickness measurements (Welty et al. 2020), regional glacier thickness is mainly estimated by ice-thickness inversion models (ITIMs) using open access digital elevation models (DEMs) (Farinotti et al. 2009; Farinotti et al. 2019; Frey et al. 2014). The DEM is a fundamental part of most regional ITIMs (Farinotti et al. 2017), and is often used to determine center flow lines (Maussion et al. 2019), shear stress (Frey et al. 2014; Wu et al. 2020), apparent mass balance (Farinotti et al. 2009), and for ice-thickness interpolation (Huss and Farinotti 2012). In addition to its use in ITIMs, the DEM has been an essential input for a wide range of TP glaciology studies, such as glacier inventory (Bhambri et al. 2011; Frey et al. 2012; Ke et al. 2016; Mölg et al. 2018), glacier mass change (Brun et al. 2017; Shean et al. 2020; Zhou et al. 2018), glacier related disasters (Allen et al. 2019; Kääb et al. 2018; Zhang et al. 2019) and projections of glacier or glacial lake evolution (Kaser et al. 2010; Kraaijenbrink et al. 2017; Zheng et al. 2021). The uncertainty in the DEMs can lead to different ITIM outcomes (Frey and Paul 2012; Fujita et al. 2017; Furian et al. 2021; Kääb 2005), especially for those ITIMs in which the DEM is a crucial input. For example, the sensitivity of the glacier bed topography (GlabTop) model to slope increases for shallower slopes (Paul and Linsbauer 2012), and slope overestimated by ~10% would result in an underestimation of ice thickness of ~32% (Linsbauer et al. 2012). Therefore, the DEM grid resolution could influence the thickness estimation from GlabTop, more detailed slope information could be provided by higher resolution DEM. Localized elevation errors and data gaps could affect the estimated ice thickness by 5−25% (Huss and Farinotti 2012). DEM errors influence the determination of model physics and the final model outcomes. Therefore, it is imperative to choose a suitable DEM source for regional glacier thickness modelling (Koldtoft et al. 2021). Farinotti et al. (2017 and 2021) intercompare the performance of most ITIMs and suggest that consideration of the uncertainty in the input data could improve the model output. However, to our knowledge, the uncertainty in different open access DEMs and its influence on various ITIM outputs over the TP has not been evaluated.

Currently, open-access DEMs covering the whole TP are mainly created by stereo mapping sensors such as ALOS AW3D30 (Tadono et al., 2015), C- or X-band interferometry synthetic aperture radar (InSAR) such as TanDEM-X, and SRTM-C based products such as NASADEM (Crippen et al. 2016). Shadows and the layover effect of InSAR technology (González and Fernández 2011), along with the deficient orientation of photogrammetrically stereo images (Mukherjee et al. 2013) or low stereo-correlation (Hugonnet et al. 2021) propagated during DEM production, may introduce errors and voids. Filling these voids with other data could result in increased uncertainty (Liu et al. 2019). Additionally, the rugged terrain of glaciers and the low contrast of snow cover can often lead to geometric distortion and missing data (Reuter et al. 2007; Takaku et al. 2020). Estimates of the accuracy of DEMs in different terrains and ~~physiognomy~~landforms, and for different vegetation coverage and land use have been conducted outside the TP using Global Navigation Satellite Systems (GNSS) measurements or high-resolution DEMs (González-Moradas and Viveen 2020; Grohmann 2018; Hawker et al. 2019; Uuemaa et al. 2020). The performance of specific DEMs varied in these studies, indicating that the local terrain and land cover influenced the DEM accuracy. In the TP, glaciers are distributed across different climatic zones and have a wide range of elevations with rugged and complicated terrain (Fielding et al. 1994; Thompson et al. 2018). GNSS measurements are not accessible for most glaciers, and publicly ~~access~~ available ~~high-resolution~~ DEM with high resolution is also ~~a~~of limitation due to its long temporal coverage (Shean 2017). The assessment of DEM accuracy in specific regions with limited GNSS measurements and high-resolution DEM is not enough to determine the performance of global DEMs across the whole glacierized TP.

Liu et al. (2019) evaluated the performance of seven public freely-accessed DEMs over the TP with sparse ICESat altimetry data and suggested that AW3D30 has a high degree of accuracy. However, ICESat data with a footprint of 70 m (larger than the resolution of their estimated DEMs) could result in intra-pixel errors in steep slopes (Uuemaa et al. 2020). Besides, glacial regions were not considered in their studies, due to the variations of glaciers over time. Misregistration among DEMs, which may lead to evaluation bias (Han et al. 2021; Hugonnet et al. 2021; Van Niel et al. 2008), was also neglected. Bearing these issues in mind, and considering the limitations of optics sensors in rugged terrain and the glacier accumulation area (Chen et al. 2021), it is clear that a further assessment of the performance of AW3D30 is required. Recently, TanDEM-X (released in 2017) and NASADEM (released in 2020) have been reported to have large improvements in accuracy relative to previous DEM products for various land-cover types (Wessel et al. 2018), floodplain sites (Hawker et al. 2019), slightly undulating terrain (Altunel 2019), and mountain environments (Gdulová et al. 2020). Nonetheless, their performance over the rugged and glacierized TP remains unclear.

The purpose of this study is to evaluate the optimal DEM to use for regional ice thickness estimation over the TP. We first evaluated the performance of six widely used DEMs: AW3D30, SRTM-GL1, NASADEM, TanDEM-X, SRTM v4.1, and MERIT which are derived from different sensors and have different resolutions, against ICESat-2 data which has been proven to have a high vertical accuracy and resolution but with sparse tracks (Fig. 1). The elevation differences between these DEMs and the ICESat-2 are systematically analyzed with regard to aspect, slope, elevation, and glacier zones. The influence on the accuracy assessment of glacier elevation changes, terrain and misregistration among DEMs is then quantified. Finally, we compare the performance of ice thickness estimates derived using the six DEMs against in-situ measurements of ice thickness using Ground Penetrating Radar (GPR). ~~The influence of DEM uncertainties on the model outcomes is also analyzed.~~[陈1][陈2]

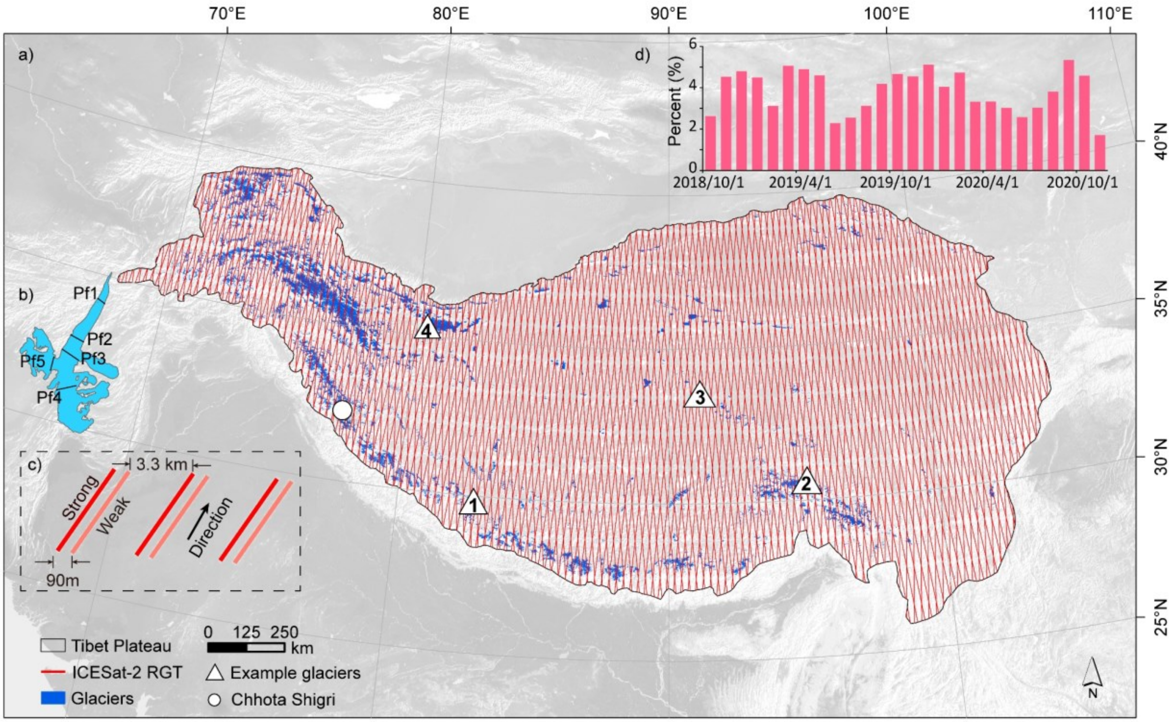


Figure 1. Location of the TP and its ICESat-2 reference ground tracks (RGTs). a) ICESat-2 tracks over the TP intersecting with glaciers. The numbered labels refer to glaciers used as examples in Fig.12. b) Location of Ground Penetrating Radar (GPR) profiles over the Chhota Shigri Glacier which is used as an example. c) Relative location of six beams when Advanced Topographic Laser Altimeter System (ATLAS) has backward orientation. Distance between RGTs is 28.8 km. d) Percentage of ICESat-2 data in different months from October 2018 to November 2020. The boundary of the TP is derived from SRTM above 2500 m a.s.l (Zhang et al. 2013).

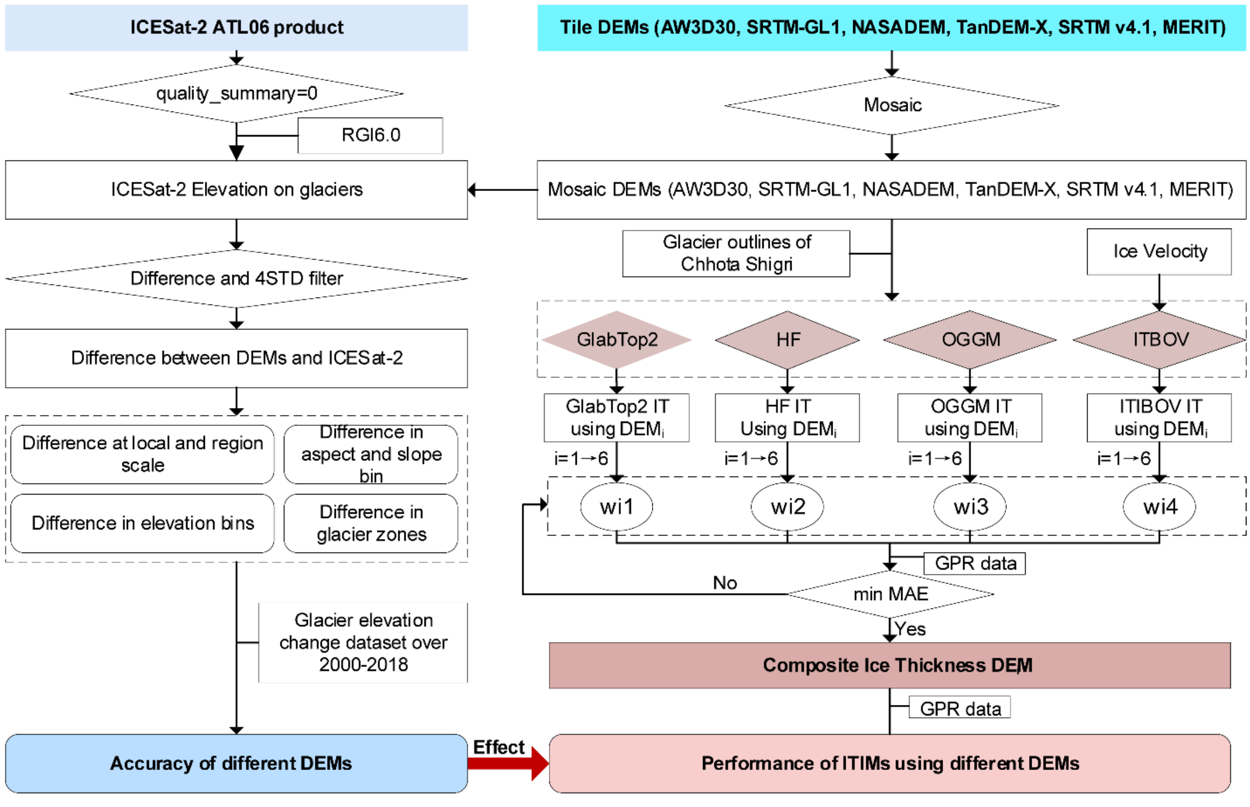
2 Data and Methods

2.1 ICESat-2 elevation data referenced

ICESat-2, a follow-on mission to the Ice, Cloud, and land Elevation Satellite (ICESat), was launched on 15 September 2018, with the goal of acquiring Earth's geolocated surface elevation that referenced to the WGS84 ellipsoid at the photon level. ICESat-2 ATLAS (Advanced Topographic Laser Altimeter System) emits a pulse every 0.7 m along the track covering a horizontal circular area with ~17 m diameter and 0.5 m in vertical extent. We used the ICESat-2 Level-3A land-ice ATL06 product. ATL06 heights are median-based heights derived from a linear-fit model over each segment corrected for first-photon bias. The segment has a length of 40 m centered on reference points at 20-m intervals along the track. The ATL06 product has better than 5 cm height accuracy and better than ~~13~~ 20 cm surface measurement precision in the Antarctic and Qilian Shan

(Brunt et al. 2019; Zhang et al. 2020). The product also contains land background points. The RGI6.0 glacier inventory (RGI Consortium, 2017) was used to extract points falling on the glaciers (Fig. 2).

ICESat-2 ATL06 data covering the TP from October 2018 to November 2020 was downloaded from https://earthdata.nasa.gov/ (Fig. 1). There are 2436 files containing about 100 GB of data in total. The fields: Location (latitude, longitude), surface elevation (h_li), elevation uncertainty (h_li_sigma) and quality (atl06_quality_summary) were used. By combining the quality field (atl06_quality_summary=0) (Smith et al. 2019) with the glacier inventory, a total of 3.5 million points out of 0.16 billion records over the TP were selected (Fig. 1). The slope, aspect and elevation value of the cell center of the DEMs were extracted for the ICESat-2 footprints.


ICESat-2 ATL06 product
Tile DEMs (AW3D30, SRTM-GL1, NASADEM, TanDEM-X, SRTM v4.1, MERIT)
quality_summary=0
Mosaic
RGI6.0
ICESat-2 Elevation on glaciers
Mosaic DEMs (AW3D30, SRTM-GL1, NASADEM, TanDEM-X, SRTM v4.1, MERIT)
Glacier outlines of Chhota Shigri
Ice Velocity
Difference and 4STD filter
GlabTop2
HF
OGGM
ITBOV
Difference between DEMs and ICESat-2
GlabTop2 IT using DEMi
HF IT Using DEMi
OGGM IT using DEMi
ITIBOV IT using DEMi
i=1→6
i=1→6
i=1→6
i=1→6
Difference at local and region scale
Difference in aspect and slope bin
Difference in elevation bins
Difference in glacier zones
wi1
wi2
wi3
wi4
GPR data
No
min MAE
Yes
Glacier elevation change dataset over 2000-2018
Composite Ice Thickness DEM
GPR data
Accuracy of different DEMs
Effect
Performance of ITIMs using different DEMs

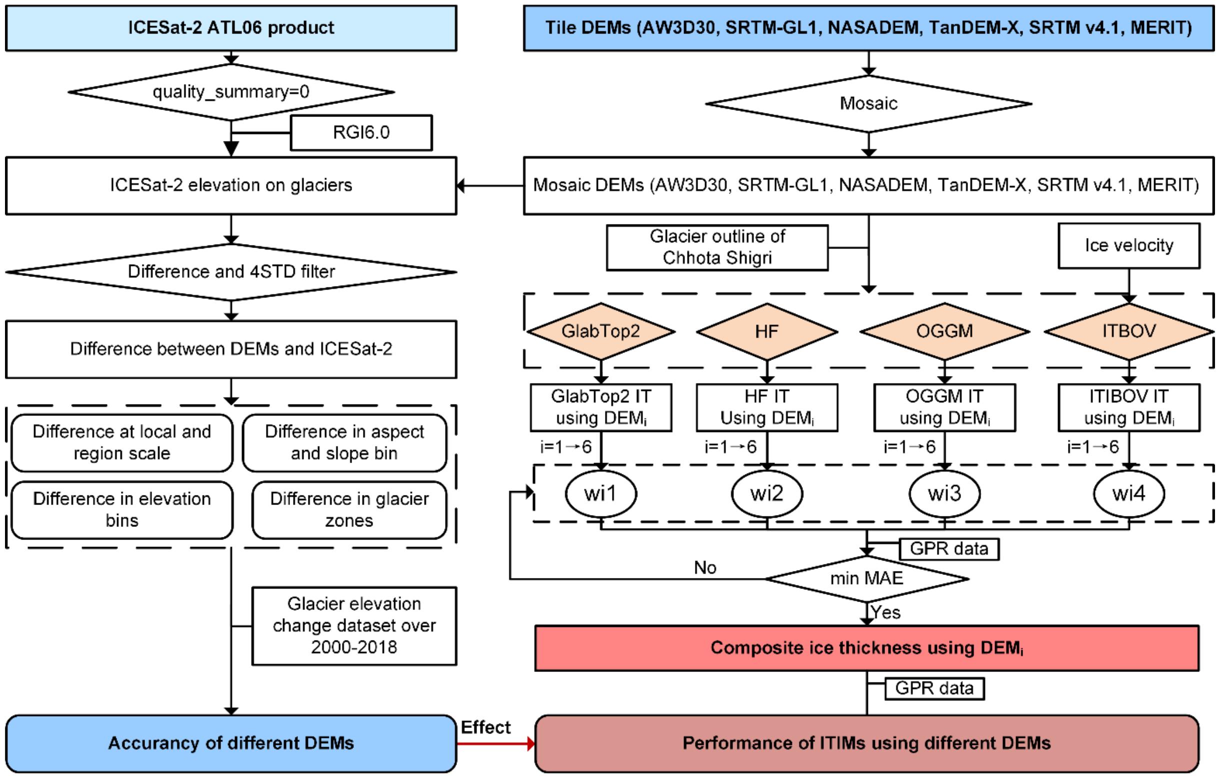


Figure 2. Flow chart showing the targets and methods used in this study including accuracy evaluation of DEMs and their effects on ice-thickness inversion models. The wi1, wi2, wi3 and wi4 denote the weight of each ~~modeled~~modelled ice thickness, i from 1 to 6 are six different DEMs, and the number 1−4 are the four ice thickness inversion models.

2.2 DEMs evaluated (AW3D30, TanDEM-X, NASADEM, SRTM)

Six global-scale DEMs were selected for evaluating their influences on ITIMs, based on popularity, data source, resolution and sensor type (optics or SAR) (Table 1).

1) ALOS World 3D - 30 m (AW3D30) is acquired by the optics stereo sensor loaded on the Advanced Land Observing Satellite (ALOS) which operated from 2006 to 2011 with a horizontal resolution of 30 m. Data gaps are filled with SRTM, ASTER GDEM v3, ArcticDEM v3, and TanDEM-X 90. Data was ~~acquired~~ available at~~from~~ https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm after user registration.

2) TanDEM-X 90 m DEM (hereafter TanDEM-X) is a product derived from the first bistatic X band SAR mission of the world which took place from 2014 to 2016 (Bachmann et al. 2021). It is a pixel-reduced product of the global TanDEM-X DEM with a pixel spacing of 0.4 arcseconds (12 m). The official reported absolute vertical and horizontal accuracy is better than 10 m at the 90% confidence level. It is noted that the current release is a non-edited version: areas with outliers, noise

and voids remain. The original data was collected during different seasons and years, and the influence of ablation and accumulation of glaciers should also be noted. Data was acquired from https://download.geoservice.dlr.de/TDM90/.

3) NASADEM is a new product released in 2020, which is derived by reprocessing the original SRTM signal data using updated interferometric unwrapping algorithms and auxiliary data, such as ICESat, to reduce voids and improve vertical accuracy (Crippen et al. 2016). Remnant voids are filled mainly by Global Digital Elevation Model (GDEM) v3 data. This data was downloaded from https://search.earthdata.nasa.gov/.

4) Other SRTM based DEMs (SRTM-GL1, SRTM v4.1, MERIT). SRTM-GL1 (30 m) is an extensively used DEM in ITIMs. The first open-access ice-thickness database of global glaciers also adopted SRTM-GL1 as its DEM source (Farinotti et al. 2019). Voids were primarily filled by ASTER GDEM2. SRTM v4.1 and MERIT were selected to compare with TanDEM-X, and simultaneously estimate the influence of DEM resolution on ITIMs.

5) SRTM v4.1, with a spatial resolution of 90 m, is produced by the method proposed by Reuter et al. (2007), including merging tiles, filling small holes iteratively and interpolating across the holes using a range of methods, according to the size of hole, and the land type surrounding it (https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1/). SRTM v4.1 was also used to compare against the performance of SRTM-GL1 to estimate the influence of resolution.

-6) MERIT is also widely used with a spatial resolution of 90 m. It was developed by removing absolute bias, stripe noise, speckle noise, and tree height bias from the existing spaceborne DEMs (SRTM3 v2.1 and AW3D30 v1) using multiple satellite datasets and filtering techniques (Yamazaki et al. 2017). Its accuracy was significantly improved, especially in flat regions (Yamazaki et al. 2017). The overall accuracy is similar to TanDEM-X in floodplain sites (Hawker et al. 2019), but lower in short vegetation. The dataset was downloaded from http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/.

Elevation of ICESat-2 data, NASADEM_SHHPv001 and TanDEM-X are based on WGS84 ellipsoid reference, and the other four DEMs are all based on EGM96 geoid (Table 1). The geoidheight function provided by MATLAB was used to calculate geoid height to unify their references.

Table 1 Details of the selected DEMs.

Item	Version	Acquisition time	Reference	Release time	Resolution (m)	Sensor type	Description	Source
AW3D30	v2.2	2006−2011	EGM96 geoid	Apr. 2019	30	Optical	Generated from its original version processed at 5 m or 2.5 m grid spacing. Voids were filled with other open-access DSMs such as SRTM, ASTER GDEM, ArcticDEM, etc.	https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm
SRTM-GL1	v003	Feb. 2000	EGM96 geoid	Sep. 2015	30	SAR C-band	ASTER GDEM2, USGS GMTED2010 or USGS National Elevation Dataset were used for voids filling	https://earthexplorer.usgs.gov/
NASADEM	SHHPv001~~v1~~	Feb. 2000	WGS84 ellipsoid	Feb. 2020	30	SAR C-band	Reprocessing of the original SRTM radar signal data and telemetry data with updated algorithms and auxiliary data such as ASTER GDEM2, ICESat, AW3D30	https://search.earthdata.nasa.gov/
TanDEM-X	v1.0 (Non-edited version)	2010−2015	WGS84 ellipsoid.	Feb. 2019	90	SAR X-band	A product variant of the 12 m (0.4 arcsec) DEM product in version 1.0 from the world's first bistatic SAR mission	https://download.geoservice.dlr.de/TDM90/
SRTM v4.1	v4.1	Feb. 2000	EGM96 geoid	Nov. 2018	90	SAR C-band	Reproduced using the method of Reuter et al. (2007).	https://drive.google.com/drive/folders/0B_J08t5spvd8RWRmYmtFa2puZEE
MERIT	v1	Feb. 2000	EGM96 geoid	Oct. 2018	90	SAR C-band	Improved by removing multiple error components from the existing spaceborne DEMs (SRTM3 v2.1 and AW3D-30m v1).	http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/

2.3 Ice thickness inversion models

Tiles of six DEMs (AW3D30, TanDEM-X, NASADEM, SRTM-GL1, SRTM v4.1, and MERIT) were used to form a mosaic of terrain data covering the whole TP. Four ice-thickness inversion models (GlabTop2, HF, OGGM, ITBOV) were used. The Chhota Shigri Glacier located in western Himalaya for which the GPR data were available (Fig.1) was selected as an example to evaluate the influence of DEM uncertainty on the ITIMs. Full details of the ITIMs are given below:

GlabTop (Glacier bed topography) is based on the theory that glacier thickness is mainly determined by the slope of the terrain (Linsbauer et al. 2012; Linsbauer et al. 2009; Paul and Linsbauer 2012). It is assumed that the glacier is an ideal plastic fluid, with bottom slip being ignored. Based on the empirical relationship between mean shear stress along the centerlines and the range of glacier elevation (Haeberli and Hoelzle 1995) (Eq. 1), the actual basal shear stress τ can be determined.

$$\tau = 0.005 + 1.598\Delta H - 0.435\Delta H^2$$
$$\tau = 150kPa, if\, \Delta H > 1600 \qquad (1)$$

where ΔH is the elevation range of glacier. The ice thickness h can then be determined from Eq. (2)

$$h = \frac{\tau}{f \rho g \sin\alpha} \qquad (2)$$

where f is the shape factor, ρ is glacier density (850±60 kg/m^3) (Huss 2013), g is the acceleration due to gravity (9.8 m/s^2) and α is the slope. Glabtop2 is an automated method for calculating ice thickness, similar to GlabTop, but avoiding digitizing the branch lines. For details refer to Frey et al. (2014).

HF (Huss- FarinottiFainotti) model is based on the mass balance principle which relates the surface mass balance of the glacier (b) to the ice flux and variation in the glacier thickness. Given the ice flux, ice thickness can be calculated according to Glen's ice flow law (Farinotti et al. 2009a; Huss and Farinotti 2012).

$$h = \sqrt[n+2]{\frac{q(1-fsl)}{2A}\frac{n+2}{\left(f \rho g \sin\alpha\right)^n}} \qquad (3)$$

where h is the mean elevation of band thickness, q is the ice flux, fsl=0.8 is the basal slip correction factor, n=3 is the exponent of flow law, ρ is glacier density (850±60 kg/m^3) (Huss 2013), g is the acceleration due to gravity (9.8 m/s^2), f is the valley shape factor (0.8) (Cuffey and Paterson 2010) and A is the Glen flow rate factor (3.24×10^{24} Pa^{-3} s^{-1}) (Cuffey and Paterson 2010; Gantayat et al. 2014).

This method defines a new variable $\tilde{b} = \dot{b} - \rho g \frac{\partial h}{\partial t}$, where $\tilde{b}$ is the apparent mass balance, $\dot{b}$ is the glacier surface mass balance, and $\frac{\partial h}{\partial t}$ is the glacier surface elevation change. $\tilde{b}$ is linearly related to the elevation and has nothing to do with whether or not the glacier is in a stable state. In the absence of mass balance data and thickness change data on the surface of a glacier, the ice flux q can be obtained by estimating $\tilde{b}$, which is determined from experience (Huss and Farinotti et al. (2012). Ice thickness in each elevation band can then be determined by substituting into Equation (3). Finally, h is extrapolated, in combination with the slope, to obtain the distributed ice thickness, according to the parameters in Huss and Farinotti (2012).

The Open Global Glacier Model (OGGM) is based on the same concept as HF, but has two main differences (Maussion et al. 2019). Firstly, the method described in Kienholz et al. (2014) is used to automatically obtain the middle streamlines and watershed division. Secondly, the apparent mass balance data are reconstructed from the local climatic dataset from variables such as precipitation and temperature.

The Ice Thickness Inversion Based On Velocity (ITIBOV) model is inverted from the shallow ice approximation, it obtains the ice thickness by combining the surface velocity field with the Glen ice flow law (Gantayat et al. 2014; Glen 1955~~; McNabb et al. 2012~~):

$$h = \frac{n+1}{2A}\frac{(1-k)u_s}{\left(f\rho g\sin\alpha\right)^n} \tag{4}$$

where h is ice thickness, u_s is glacier surface velocity, and k is the contribution ratio of basal slip velocity relative to the u_s. We used the mean velocity over 1985-2019 from ITSLIVE dataset (Gardner 2019) as the u_s input. We assumed that basal slip only occurred during the warm seasons, and k was calculated by dividing the annual glacier velocity by winter glacier velocity (Wu et al. 2020). Data from the Global Land Ice Velocity Extraction from Landsat 8 (GoLIVE) dataset with a date separation length of less than 96 days are used to estimate the monthly velocity (Fahnestock et al. 2016; Scambos 2016), allowing the winter velocity (December, January and February) and annual mean velocity to be calculated. Basal factor k was calculated as 0.80 (Fig. S1). ~~Some~~ The shared parameters, such as creep factor, shape factor and basal creep factor are the same in all four models.

It is possible that an ensemble of the output from different models can improve the ~~modeled~~modelled thickness (Farinotti et al. 2017; Farinotti et al. 2021). Therefore, after calculating the ice thickness from four models using different DEMs, we calculated an ensemble ice thickness using the same DEM but with different models. ~~The weighting given to each model is iteratively calculated to achieve a minimal mean absolute error (Fig. 2).~~ First, the ensemble ice thickness was the sum of the four models with four weights w1, w2, w3, and w4, respectively. The sum of four weights equals to 1. 70% of the GPR result are adopted as calibration data. 30% of the GPR result are adopted as validation data. Then, the four weights iteratively changed to achieve the minimal mean absolute error between calibration data and model result. ~~[陈3][陈4]~~Finally, the MAE between ensemble ice thickness and validation data are calculated.

2.4 Accuracy assessment

The error in the DEMs is considered to be the difference between the DEM elevation and the ICESat-2 measurement. To remove the influence of outliers, elevation differences outside four standard deviations were removed. Mean error (ME), mean absolute error (MAE), median error, root mean square error (RMSE), standard deviation (STD), and normalized median absolute deviation (NMAD) were calculated for the error assessments. NMAD and ME were used to assess the disturbance from extreme errors (Höhle and Höhle 2009; Gdulová et al. 2020). When calculating the ME, the positive and negative biases cancel each other, making the error smaller; therefore, the STD together with ME could be a complementary indicator for assessment.

$$ME = \frac{1}{\mathrm{n}}\sum_{1}^{n}(H_{DEM} - H_{ICESat\text{-}2}) \tag{5}$$

$$MAE = \frac{1}{\mathrm{n}}\sum_{1}^{n} abs(H_{ICESat\text{-}2} - H_{DEM}) \tag{6}$$

$$RMSE = \sqrt{\frac{1}{\mathrm{n}}\sum_{1}^{n}(H_{ICESat\text{-}2} - H_{DEM})^2} \tag{7}$$

$$STD = \sqrt{\frac{1}{\mathrm{n\text{-}1}}\sum_{1}^{n}(H_{ICESat\text{-}2} - H_{DEM} - ME)^2} \tag{8}$$

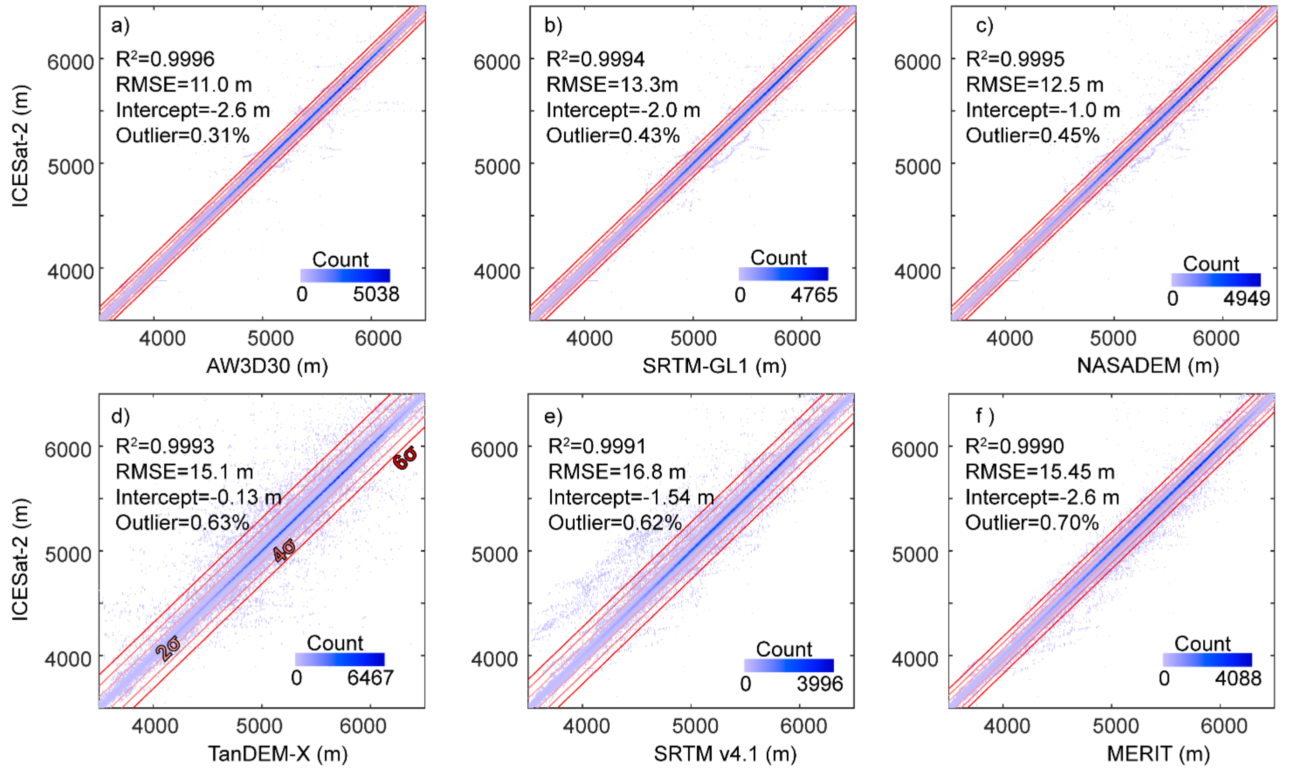
$$NMAD = 1.4826 * median(abs(H_{ICESat\text{-}2} - H_{DEM})) \tag{9}$$

Glacier surface elevation changed at -21—17m/yr over the TP during 2000-2018 (Shean et al. 2020). Therefore, the disparity of acquiring date between ICESat-2 and six DEM (Table 1) could introduce large error due to the glacier dynamic. TanDEM-X and AW3D30 are acquired in different months and years (Table 1), it's hard to analyse the impact of glacier dynamic on accuracy assessment. However, the other four DEMs are produced from NASA's Shuttle Radar Topography Mission during the 11-day mission in February 2000. We selected ICESat-2 data acquired in February 2019 and 2020. Then the glacier elevation dynamic magnitude during February 2000 and February 2019/2020 are subtracted from the selected ICESat-2 elevation based on the mean glacier elevation change data from Shean et al. (2020). By comparing the elevation from the four DEMs and adjust ICESat-2, we could exactly know the impacts on accuracy assessment from glacier dynamic.

3. Results

3.1 Accuracy of DEMs

Figure 3 shows a comparison of elevation from the six DEMs with the ICESat-2 data. The four standard deviations (that is 4 std) filter was chosen to filter on the differences between ICESat-2 and DEMs used to filter outexclude extreme outliers. Ration of excluded outliers relative to record of each DEM is excluded less than 1% of the data. Overall, there is non- irregular deviation existed among these DEMs and ICESat-2 elevation after filtering. The ICESat-2 vs DEMs values are distributed tightly around the fit line with a slope coefficient close to of 1, with no obvious differences among the R^2. NASADEM and TanDEM-X performed the best in terms of intercept and fit RMSE, with very little difference to the ICESat-2 data. For the other four DEMs, there are obvious systematic shifts which can be inferred from the high R^2 values, but high intercept values.


a)
R²=0.9996
RMSE=11.0 m
Intercept=-2.6 m
Outlier=0.31%
b)
R²=0.9994
RMSE=13.3m
Intercept=-2.0 m
Outlier=0.43%
c)
R²=0.9995
RMSE=12.5 m
Intercept=-1.0 m
Outlier=0.45%
d)
R²=0.9993
RMSE=15.1 m
Intercept=-0.13 m
Outlier=0.63%
e)
R²=0.9991
RMSE=16.8 m
Intercept=-1.54 m
Outlier=0.62%
f)
R²=0.9990
RMSE=15.45 m
Intercept=-2.6 m
Outlier=0.70%
Count
0 5038
0 4765
0 4949
0 6467
0 3996
0 4088
2σ
4σ
6σ
ICESat-2 (m)
AW3D30 (m)
SRTM-GL1 (m)
NASADEM (m)
TanDEM-X (m)
SRTM v4.1 (m)
MERIT (m)

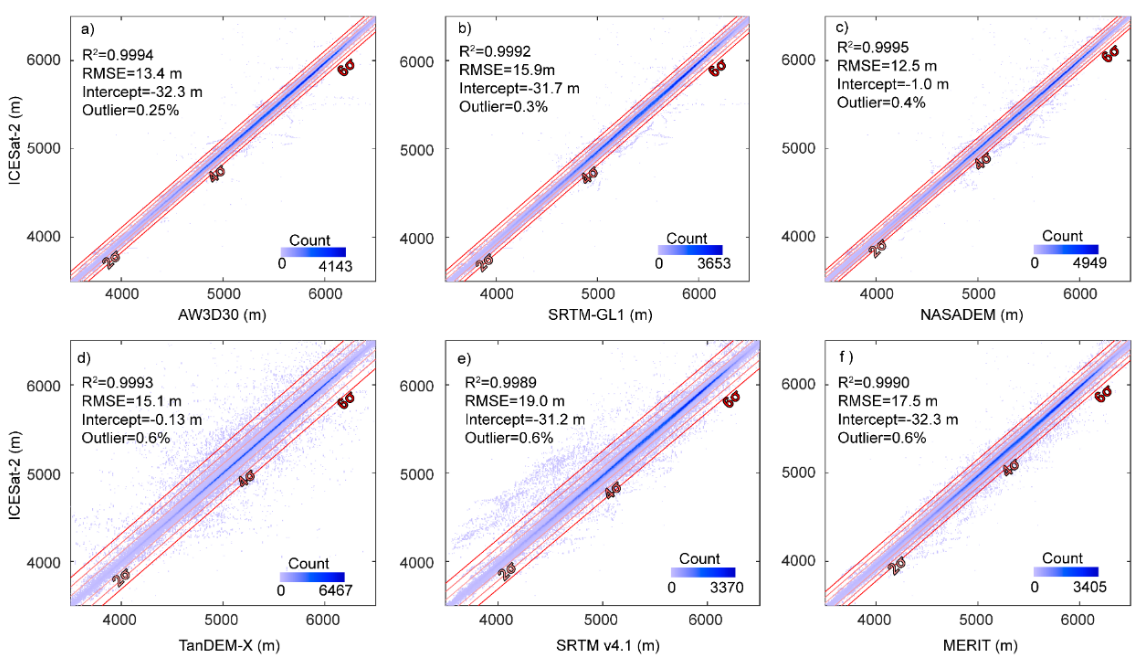
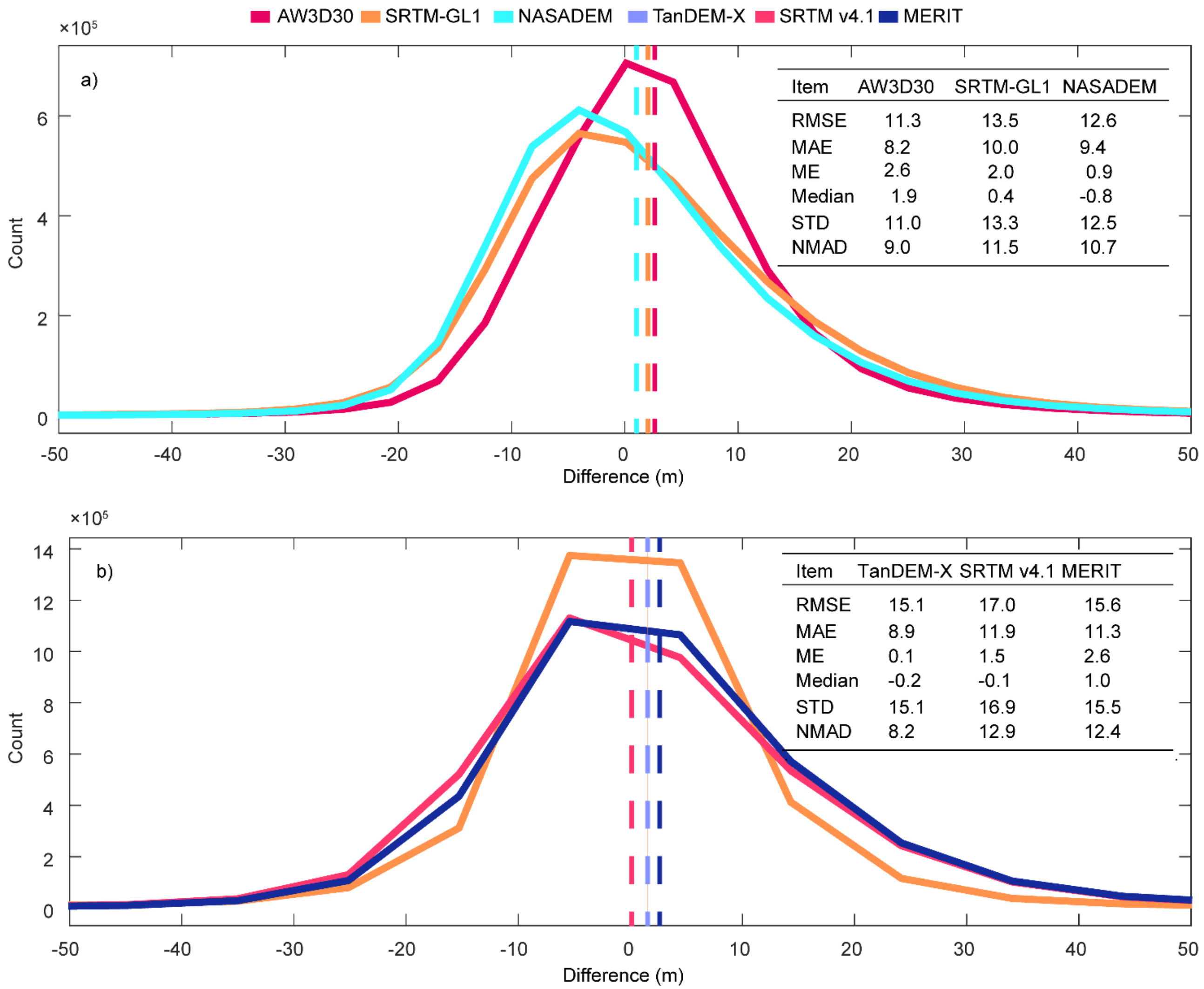
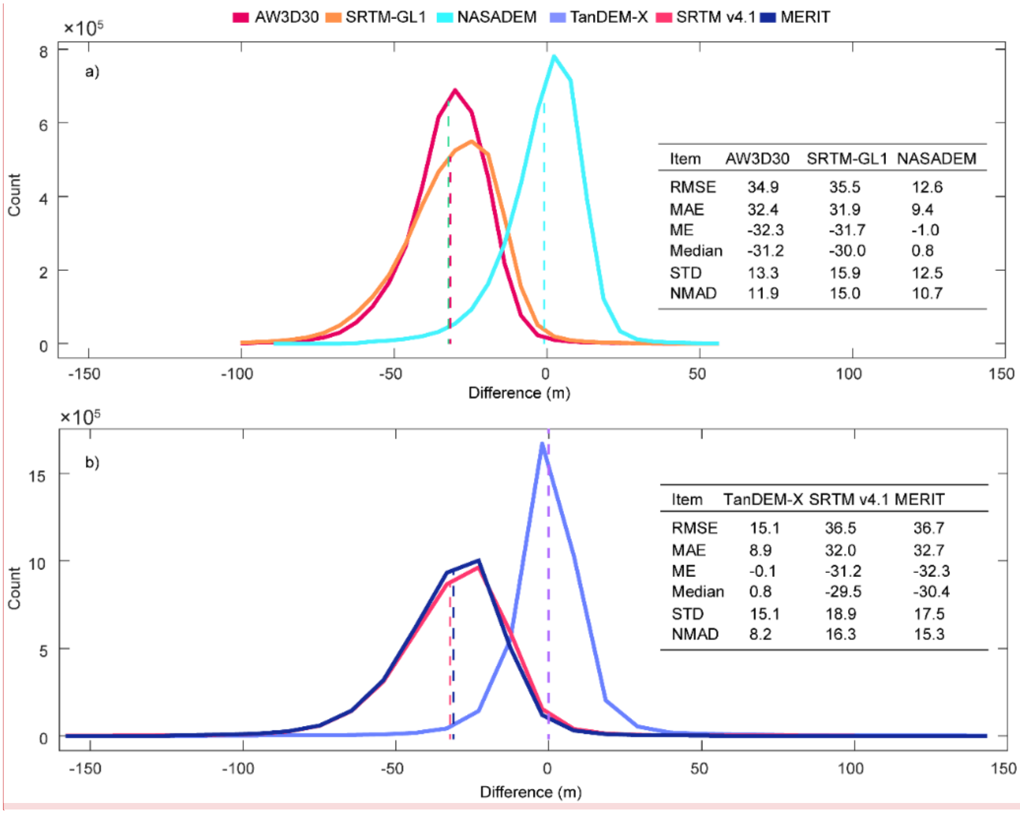

a)
R²=0.9994
RMSE=13.4 m
Intercept=-32.3 m
Outlier=0.25%
b)
R²=0.9992
RMSE=15.9m
Intercept=-31.7 m
Outlier=0.3%
c)
R²=0.9995
RMSE=12.5 m
Intercept=-1.0 m
Outlier=0.4%
d)
R²=0.9993
RMSE=15.1 m
Intercept=-0.13 m
Outlier=0.6%
e)
R²=0.9989
RMSE=19.0 m
Intercept=-31.2 m
Outlier=0.6%
f)
R²=0.9990
RMSE=17.5 m
Intercept=-32.3 m
Outlier=0.6%
Count
0 4143
0 3653
0 4949
0 6467
0 3370
0 3405
2σ
4σ
6σ
ICESat-2 (m)
AW3D30 (m)
SRTM-GL1 (m)
NASADEM (m)
TanDEM-X (m)
SRTM v4.1 (m)
MERIT (m)

Figure 3. Differences between six DEMs and ICESat-2 elevation. a) AW3D30, b) SRTM-GL1, c) NASADEM, d) TanDEM-X, e) SRTM v4.1, and f) MERIT. The gradually lighter red lines denote the range within 2, 4 and 6 std of the mean. The text at the top left of each panel gives the fit results for data within 4 std of the mean. 'Outlier' denotes the proportion of outliers relative to the total records. 'R2', 'RMSE', and 'Intercept' are fit results when the slope coefficient is set to 1. Elevation range was cut to 3500−6500 m, the range in which most elevations values are located, to show clearly the effect of using different multiples of the std from the mean.

The difference statistics for the six DEMs are presented in Figure 4. Statistically, Median and ME differed little, which indicated that and extreme values did not influence the ME much after the 4 std filter was applied. STD was slightly larger than NMAD, especially for TanDEM-X, indicating larger discrepancies due to the DEM errors and noise (Höhle and Höhle 2009). NASADEM performed far better than the other two 30-m resolution DEMs in , with smaller RMSE (12.6 m), MAE (9.4 m), and ME (−1.0 m). AW3D30 behaved best in RMSE (11.3 m), MAE (8.2 m). has a lower absolute accuracy (RMSE: 34.9 m, ME: −32.3 m), but a similar relative accuracy to NASADEM because of the similar overall dispersion (~13 m) and spatial scale. SRTM-GL1 and NASADEM are both produced from same original SAR data, but have large differencesdiffered in RMSE (3513.5 vs 12.6 m), MAE (31.910.0 vs 9.4 m), and ME (−31.72.0 vs −1.00.9 m). The new algorithm and auxiliary data applied in NASADEM do indeed greatly improve the absolute accuracy of the product over glacierized terrain. The quality of TanDEM-X was the best out of the 90-m resolution DEMs with smallest RMSE (15.1 m), MAE (8.9 m), ME (−0.1 m), and STD (15.1 m). SRTM v4.1 and MERIT are both error-reduced products from SRTM3 v2 (Reuter et al. 2007; Yamazaki et al. 2017), and they have similar behavior, with ME (about −32 1.5 m vs 2.6 m) and RMSE (~3717.0 m. vs 15.6 m).



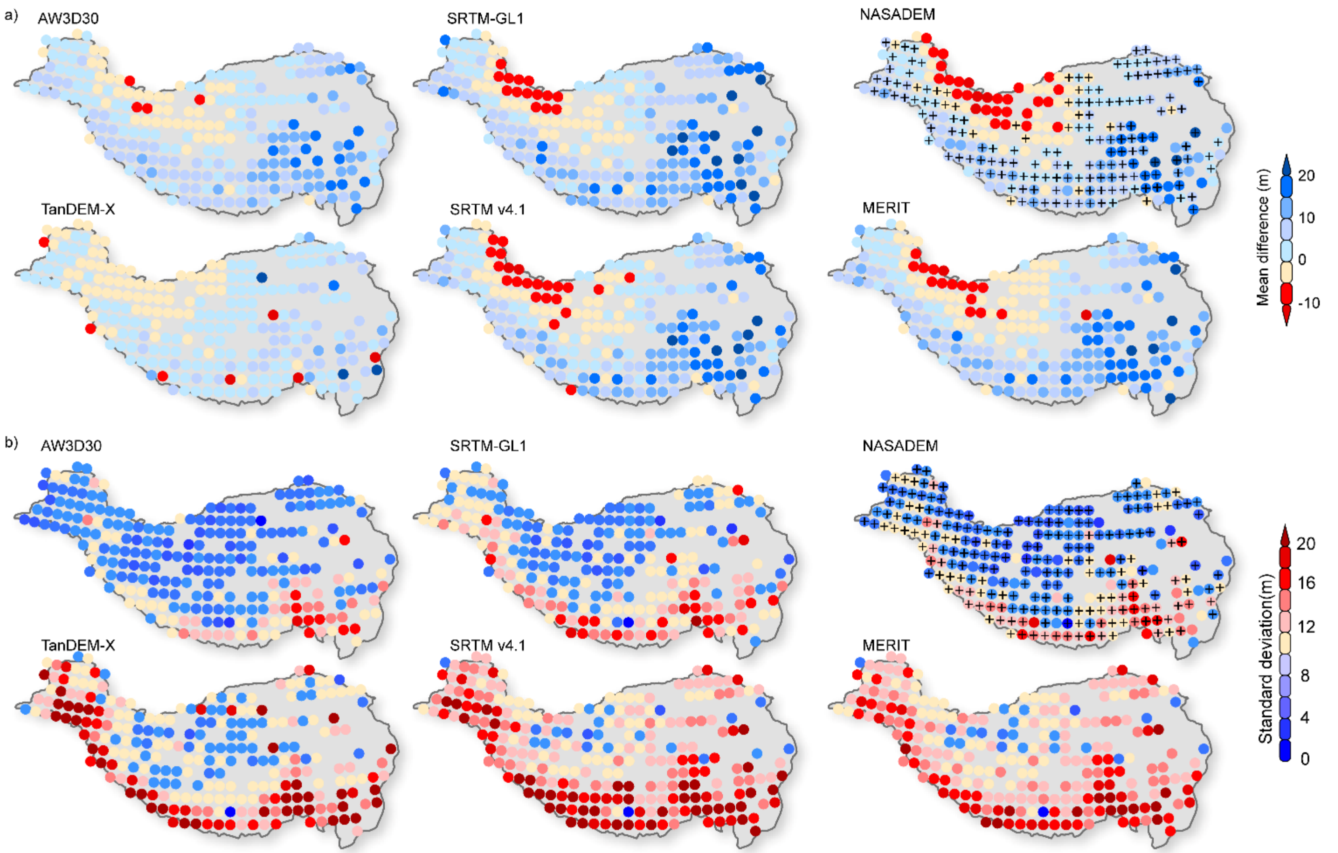
AW3D30 SRTM-GL1 NASADEM TanDEM-X SRTM v4.1 MERIT
×10⁵
a)
Count
Difference (m)
Item AW3D30 SRTM-GL1 NASADEM
RMSE 11.3 13.5 12.6
MAE 8.2 10.0 9.4
ME 2.6 2.0 0.9
Median 1.9 0.4 -0.8
STD 11.0 13.3 12.5
NMAD 9.0 11.5 10.7
×10⁵
b)
Count
Difference (m)
Item TanDEM-X SRTM v4.1 MERIT
RMSE 15.1 17.0 15.6
MAE 8.9 11.9 11.3
ME 0.1 1.5 2.6
Median -0.2 -0.1 1.0
STD 15.1 16.9 15.5
NMAD 8.2 12.9 12.4



[陈5][陈6]

Figure 4. Overall difference (m) statistics between six DMEs and ICESat-2 elevation. a) 30-m-resolution DEMs, AW3D30, SRTM-GL1 and NASADEM. b) 90-m-resolution DEMs, TanDEM-X, SRTM v4.1 and MERIT. The vertical dash line denote the mean difference of each DEM between ICESat-2.

The spatial distribution of the ME and STD are shown in Figure 5. ~~AW3D30, SRTMGL-1, SRTM v4.1 and MERIT all have large negative ME over the TP, while the ME of NASADEM and TanDEM-X are mostly within the range −10 to 10 m. For these two DEMs, t~~The ME ~~in~~ in southeast Tibet ~~the Himalaya~~ is more ~~negative~~ positive than that in ~~in southeast Tibet~~the Himalaya, and it is slightly ~~positive~~ negative in western Kunlun and the Karakoram mountains. It is worth noting that in the Himalaya and southeast Tibet, the ME of ~~NASADEM~~ these four DEMs is more ~~negative~~ positive than that of TanDEM-X and AW3D30. ME of TanDEM-X are mainly at ±5 m, but with some large values in several regions. SRTMGL-1, NASADEM, SRTM v4.1 and MERIT have nearly same distribution of ME, and all show negative ME values in the West Kunlun and Karakoram. ME of NASADEM is smaller than SRTM-GL1 in most regions of TP, but is bigger in West Kunlun and Karakoram. Overall, STD of 30-m resolution DEMs is much better than that of 90-m resolution DEMs (Fig.5b). STD along the Hindu Kush-Himalaya and southeast Tibet was larger than that in other regions. Thereinto, STD in southeast Tibet was relatively larger (>12 m). Specifically, the STD of AW3D30 ~~and NASADEM~~ was minimum and spatially relevant. Relative to ME, STD of NASADEM improved over the most part of TP, comparing with that of SRTM-GL1. This indicate that some disturbances from noise and errors may exit in the SRTM-GL1, SRTM v4.1 and MERIT in the West Kunlun and Karakoram. TanDEM-X performs well in overall statistics (Fig. 4b) and ME (Fig.5a), but ~~it is not stable~~didn't show much advantage in STD. ~~Spatially, it performed worse than SRTM-GL1 in terms of STD (Fig.5b).~~ The STD and ME of SRTM v4.1 and MERIT are almost the same in space (Fig.5b), corresponding to their similar overall STD (both ~~~18~~ 15 m) and ME (both ~~~3~~2 m) values (Fig.4b).


a)
AW3D30
SRTM-GL1
NASADEM
TanDEM-X
SRTM v4.1
MERIT
Mean difference (m)
20
10
0
-10
b)
AW3D30
SRTM-GL1
NASADEM
TanDEM-X
SRTM v4.1
MERIT
Standard deviation(m)
20
16
12
8
4
0

285

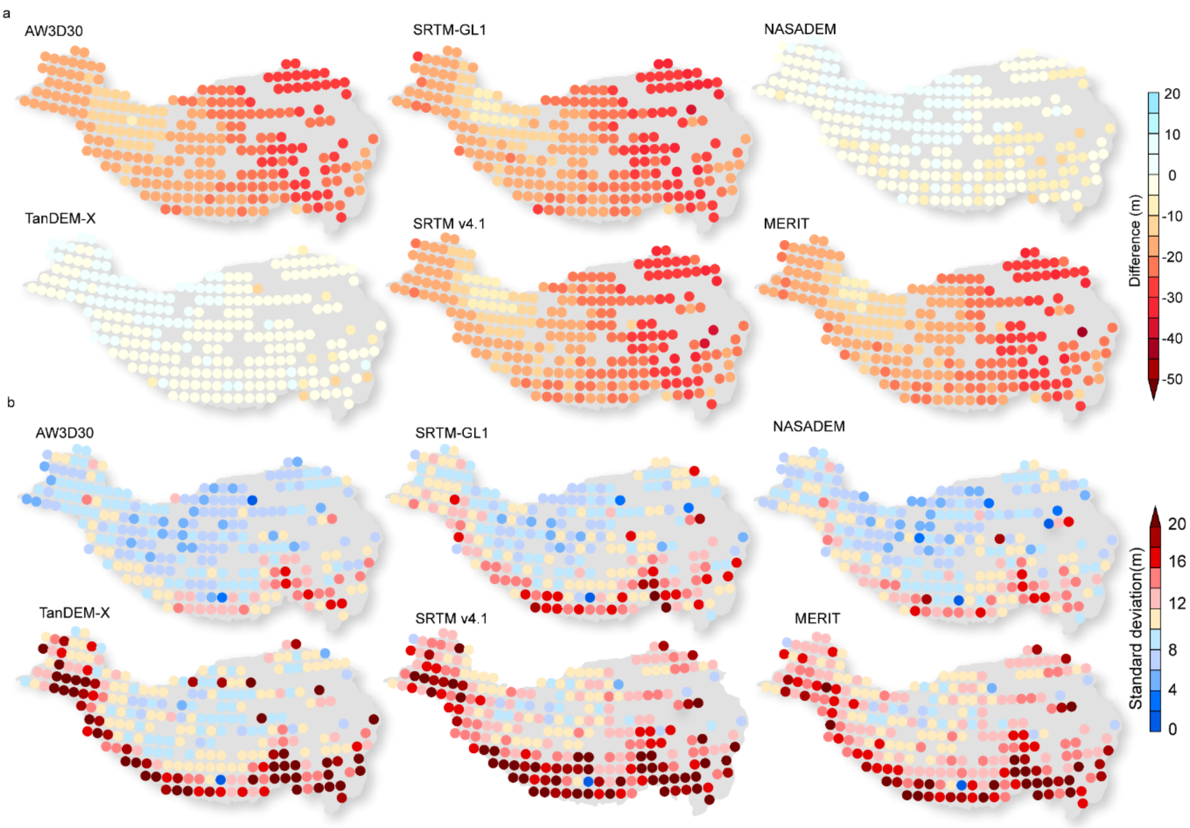


Figure 5. Aggregated spatial mean error (ME) (a) and standard deviation (STD) (b) between six DEMs and ICESat-2 elevation for 1°×1° cells across the TP. The cross symbol denotes that NASADEM performs better than SRTM-GL1 in ME or STD

3.2 Differences between DEMs and ICESat-2 in aspect, slope and elevation

The influence of terrain factors on the differences between the DEM and ~~ICEsat~~ICESat-2 elevations for the six DEMs are presented in Figure 6. The influence of aspect is most apparent for SRTM-GL1, with a median value of about ~~–25~~ 5 m in the south aspect which increased in magnitude gradually towards to the north aspect (~ ~~–3~~5 m). A similar pattern, but with a smaller amplitude ~~(±5 m) is apparent~~ for the NASADEM ~~and~~ , TanDEM-~~X~~X, MERIT (±1 m) and AW3D30 (0 – 2 m) (Fig.6a).

~~For NASADEM and TanDEM-X, the differences plotted against slope are distributed around zero, with mean median values of about –1.6 and –1.0 m, respectively (Fig.6b). For the other DEMs, the differences with slope are mainly less than zero, with a median range of –30 to –48 m and a mean upper quartile of ~20 m.~~ The median differences of the 30-m DEMs generally increased along the slope. However, for the 90-m DEMs, the difference increased with slope at first, but then decreased on steep slopes. NASADEM and TanDEM-X had minimum mean median values of about 0.9 and 1.2 m, respectively (Fig.6b). For all DEMs, the spreads of differences become larger as the slope becomes steeper. This increase is most obvious for

TanDEM-X and SRTM v4.1, with rates of ~~0.71~~1.29 m/degree (r=0.~~96~~97, p<0.01) and ~~0~~1.~~60~~ 11 m/degree (r=0.~~83~~89, p<0.01). This indicated that errors of both DEM suffered from serious slope effect. AW3D30 and NASADEM has similar mean spread (19.2m vs 20.8m). On slopes of less than 20°, TanDEM-X has the best quality with a mean median value of -0.2~~±0.39~~ m and mean spread of 11.7 m, ~~and 5.8±2.2 m, but increased disparity on steeper slopes~~respectively. MERIT shows a slight advantage over SRTM v4.1 with a reduced spread for steep slopes. —Overall, relative to the other DEMs, AW3D30 and NASADEM behaves best against slope in terms of spread and median value. ~~MERIT shows a slight advantage over SRTM v4.1 with a reduced spread for steep slopes.~~

The differences for all DEMs generally ~~increased~~ decreased with elevation, with fluctuations around zero at very high elevations (Fig.6c). AW3D30 has smaller difference at low elevation relative to NASADEM and SRTM-GL1. For NASADEM and ~~TanDEM-X~~SRTM-GL1, the differences ~~are negative at lower elevations and slightly positive at higher~~ along the elevations show similar distribution~~; NASADEM~~ and varied ~~from −70 to 20 m over the full elevation range, and~~ from –10 to 10 m over the range 4500−6500 m, where measurements are concentrated; TanDEM-X varied ~~from −15 to 20 m over the full elevation range, and~~ from around –5 to 5 m between 4500 and 6500 m. ~~For the other four DEMs, the differences all remained below zero for the full range of elevations.~~ The ~~SRTM-GL1,~~ SRTM v4.1 and MERIT differences changed ~~similarly~~ almost same from –~~100~~ 20 to 40 m~~. In comparison, the differences for AW3D30 were smaller for lower elevation bins, ranging from −70 to 0 m.~~, but show difference at high elevation reigon.

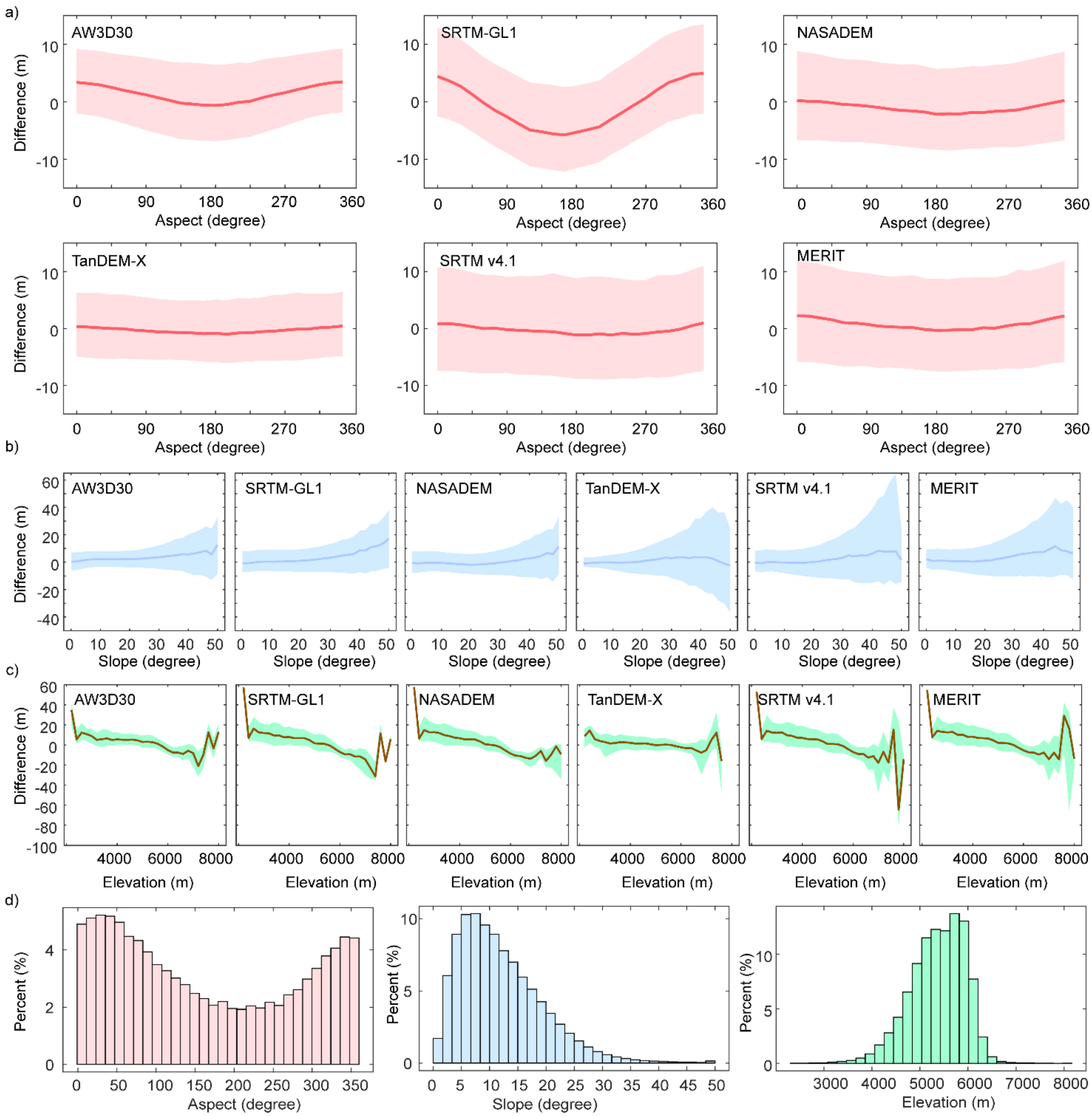


Figure 6. Differences between six DEMs and ICESat-2 with terrain factors. (a) 5° aspect bin. (b) 2° slope bin. (c) 200 m elevation bin. (d) Percentage (%) of data in each aspect, slope and elevation bin.

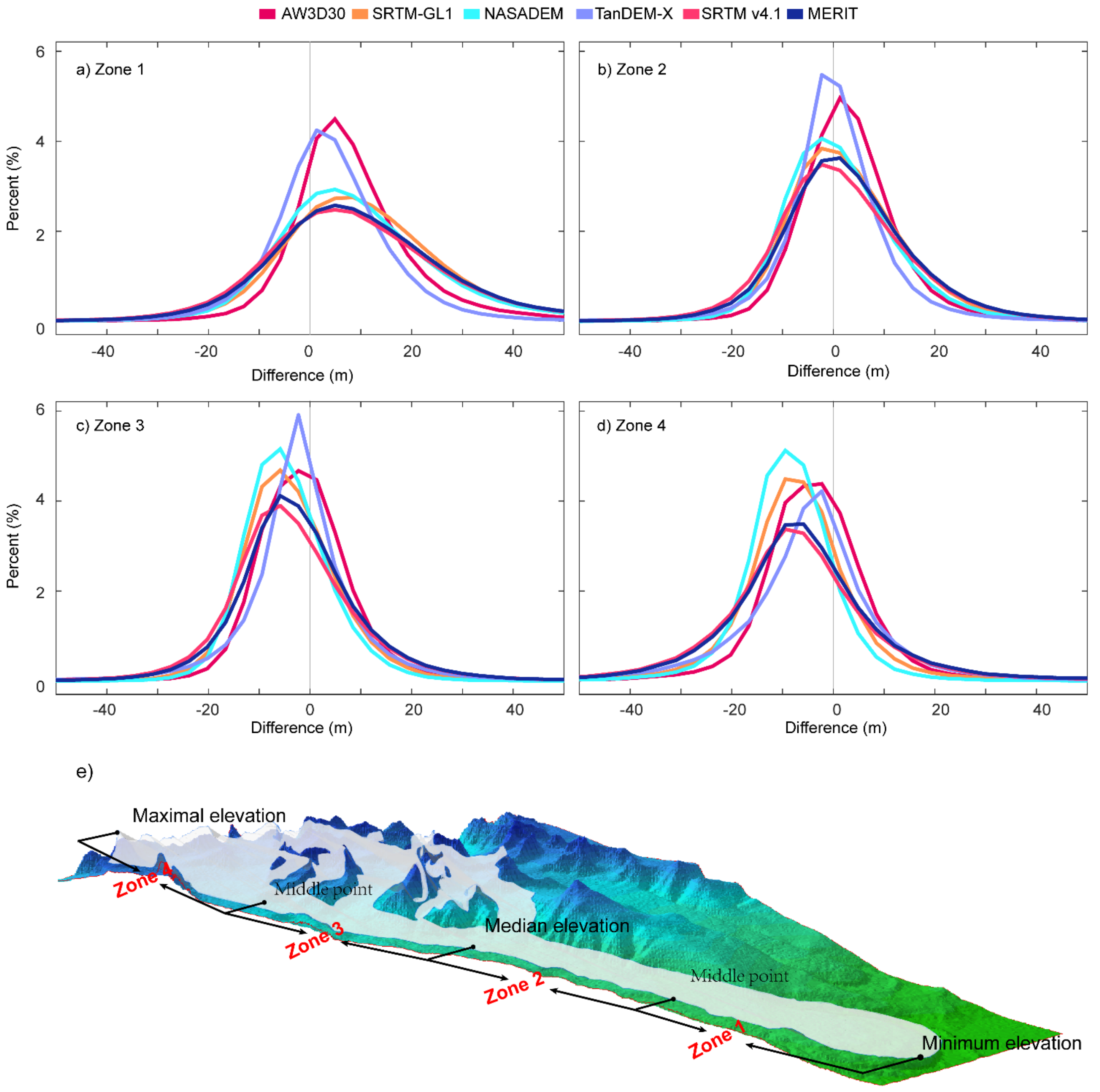
320

3.3 Differences between DEMs and ICESat-2 in different glacier zones

Differences in different glacier zones were also estimated and are shown in Figure 7a-d. We divided it into four sub-zones using the maximal, median and minimum elevation from the RGI glacier inventory (Fig. 7e). Here we consider Zone 1 to be the ablation area and Zone 4 the accumulation area. Zone 2 and Zone 3 are transition areas. Crests of the probability distribution of differences located in the ~~negative~~ positive axis range in Fig. 7a move to the ~~right~~ left in Fig. 7b−d. Correspondingly, ME, MAE and RMSE all decrease when moving from Zone 1 (ablation area) to Zone 2 (transition area) (Fig.7 and Table S1). Spatially, areas in the glacier terminus are subject to more melting (Brun et al. 2017) leading to this decrease. The ME of the SRTM based products SRTM-GL1, SRTM v4.1, NASADEM and MERIT are all around 10 m in Zone1 and decreased similarly by 8.~~7~~1, 7.6~~3~~, ~~8~~7.~~1~~ 5 and 7.~~8~~ 2 m towards Zone2, respectively (Table S1). Temporally, the ~~values~~ ME of the DEM acquired in earlier periods ~~decreased~~ is bigger~~more~~. The ME ~~decreased by a larger value (5.6 m) for~~is 8.1m for AW3D30, which was acquired in 2006−2011, ~~than~~bigger than for that of TanDEM-X (3.~~5~~ 9 m), which was acquired in 2010−2015.

ME, MAE and RMSE in Zone 3 and Zone 4, near or in the accumulation area, are almost all smaller than the corresponding values in Zone 1 ~~and Zone 2~~ (Fig. 7 and Table S1). ~~For TanDEM-X and NASADEM, which have better absolute accuracy than the other DEMs,~~ ME of all DEMs ~~changed~~decreased to ~~positive~~negative values ~~in~~in Zone 3 and Zone 4. Usually, in the accumulation area, glaciers have a positive or less negative elevation change (Li and Lin 2017; Maurer et al. 2019; Rankl and Braun 2016), therefore, accumulation may be concerned with changes in Zone3 and Zone 4. The observed shift in the difference from zone 1 to zone 4 is a sign that thinning or accumulation between the time of collection of the six DEMs and the ICESat-2 data is most pronounced in zones 1 and 2.

In terms of STD, NASADEM performed best in ~~all glacier zones, except for~~ Zone ~~1~~3 and Zone 4, with values ranging from ~~9.2~~8.8 to ~~16~~11.~~5~~1 m (Table S1). AW3D30 had the best performance of all DEMs in Zone 1 and Zone 2~~, and was the next best performer overall,~~ , with a STD varying from 1~~2~~0.0 to ~~20.7~~12.3 m. The STD of TanDEM-X was better than that of SRTM-~~GL1~~v4.1 and MERIT in Zone 1 and Zone 2, but worse in Zone 3 and Zone 4 where it suffered from discrepancies. ~~MERIT showed slight improvements in STD at ~2% relative to SRTM v4.1.~~



AW3D30 SRTM-GL1 NASADEM TanDEM-X SRTM v4.1 MERIT
a) Zone 1
b) Zone 2
c) Zone 3
d) Zone 4
Percent (%)
Difference (m)
e)
Maximal elevation
Middle point
Median elevation
Middle point
Minimum elevation
Zone 4
Zone 3
Zone 2
Zone 1

345

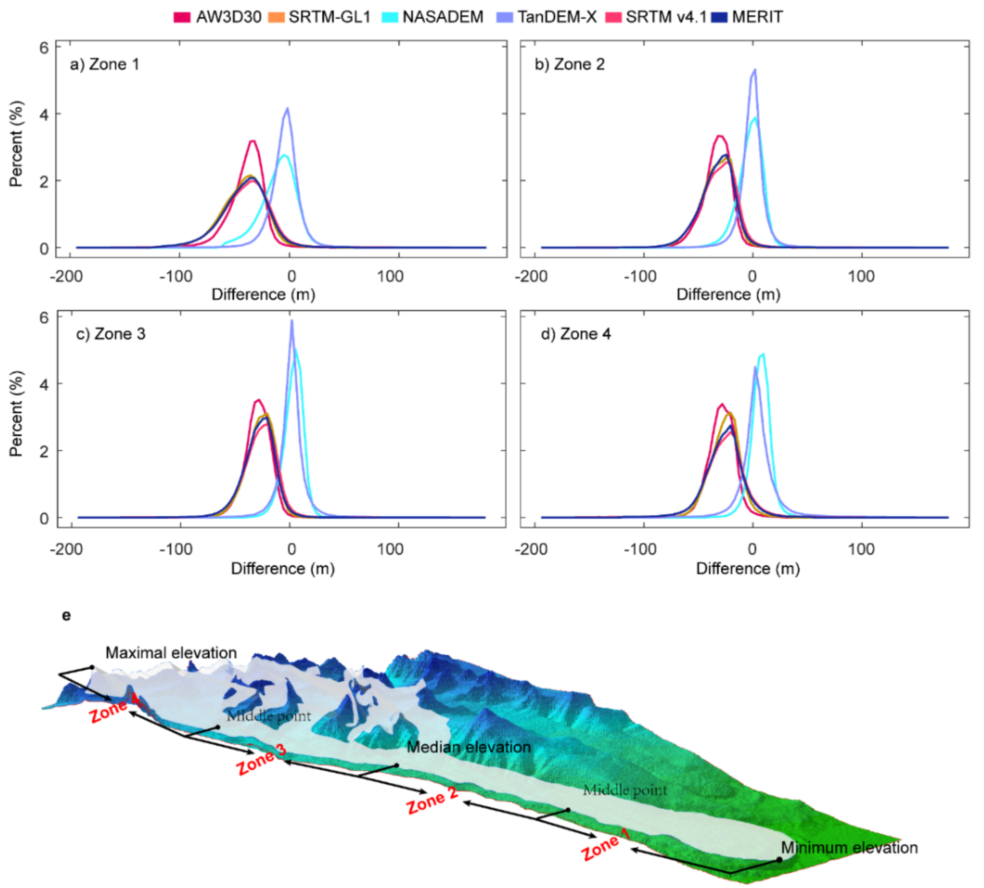


Figure 7. Probability distribution of the difference between six DEMs elevation and ICESat-2 elevation in different glacier zones (a−d) that defined in panel (e).

ME, MAE and RMSE in Zone 3 and Zone 4, near or in the accumulation area, are almost all smaller than the corresponding values in Zone 1 and Zone 2 (Fig. 7 and Table S1). For TanDEM-X and NASADEM, which have better absolute accuracy than the other DEMs, ME changed to positive values in Zone 3 and Zone 4. Usually, in the accumulation area, glaciers have a positive or less negative elevation change (Li and Lin 2017; Maurer et al. 2019; Rankl and Braun 2016), therefore, accumulation may be concerned with changes in Zone3 and Zone 4. [陈7]

In terms of STD, NASADEM performed best in all glacier zones, except for Zone 1, with values ranging from 9.2 to 16.5 m (Table S1). AW3D30 had the best performance of all DEMs in Zone 1, and was the next best performer overall, with a STD varying from 12.0 to 20.7 m. The STD of TanDEM-X was better than that of SRTM-GL1 in Zone 1 and Zone 2, but worse in Zone 3 and Zone 4 where it suffered from discrepancies. MERIT showed slight improvements in STD at ~2% relative to SRTM v4.1.

3.4 Comparisons of ice thickness modelled by DEMs

The models are not adjusted independently according to the difference between the output and GPR results. Therefore, the results are not indicators of the performance of the models but rather references for examining the influence of different DEMs on specific ITIMs. The effect of the DEMs on the model outcomes are presented in Figure 8 and are quite obvious. Mean ice thickness differs, according to the DEM used, by up to ~~88~~134%, ~~4~~6%, 47% and ~~7~~19% for GlabTop2, HF, ITIBOV and OGGM, respectively. The deepest ice thickness differs by up to ~~55~~53%, 25%, ~~16~~13% and ~~18~~13% for GlabTop2, HF, ITIBOV and OGGM, respectively.

The mean ice thicknesses from GlabTop2 and ITIBOV using the 90-m DEMs (that is TanDEM-X, SRTM v4.1 and MERIT) are ~30 m less than those obtained from using 30-m DEMs (that is AW3D, SRTM-GL1 and NASADEM) (Fig. 8). GlabTop2, HF and OGGM using AW3D30, and ~~HF and~~ ITIBOV using NASADEM output the maximal mean thickness. GlabTop2 and ITIBOV using TanDEM-X, ~~ITIBOV and~~ OGGM ~~using TanDEM-X,~~ and HF using SRTM-GL1 output the minimum mean thickness.

The influence of different DEMs on ITIMs can also be identified when making a comparison with the GPR results (Fig. 8 and Table 2). If median error is used as the criterion~~criteria~~, GlabTop2 and ~~using~~ITIBOV using NASADEM, HF using AW3D30, ~~ITIBOV using NASADEM~~ and OGGM using ~~NASADEM~~ SRTM v4.1 achieved the relatively best simulation (Fig. 9). If RMSE was used, GlabTop2 using NASADEM, HF using SRTM-GL1, ITIBOV using ~~NASADEM~~ AW3D30 and OGGM using TanDEM-X performed best (Table 2).

In different glacier zones, each DEM-model combination has its merits and weakness (Table 2). Totals of 9, ~~3~~2, 4~~7,~~ and 3 ~~and 1~~ output achieved the minimum RMSE in profiles by different ITIMs using AW3D30, SRTM-GL1, NASADEM ~~,~~ and TanDEM-X ~~and MERIT~~, respectively. Overall, ~~Overall,~~ NASADEM~~, as input to GlabTop2 and ITBOV, performs best, with RMSE values of 75.4 and 61.3 m, respectively; SRTM-GL1 performed the best in HF with an RMSE of 50 m; TanDEM-X performed the best in OGGM with an RMSE of 52.8 m (Table 2).~~ and AW3D30 performed better in different glacier zones in all models.

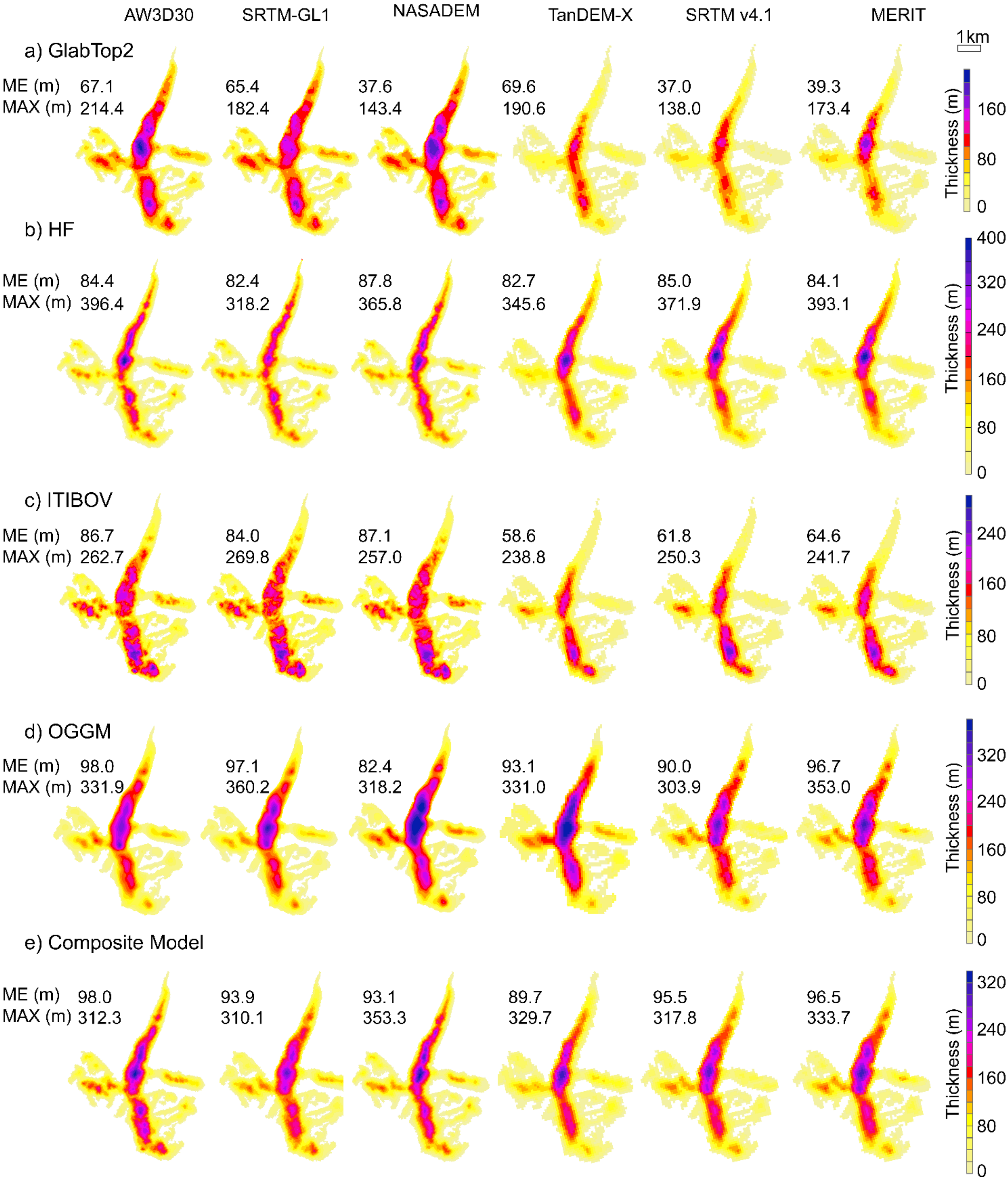

AW3D30
SRTM-GL1
NASADEM
TanDEM-X
SRTM v4.1
MERIT
1km
a) GlabTop2
ME (m) 67.1 65.4 37.6 69.6 37.0 39.3
MAX (m) 214.4 182.4 143.4 190.6 138.0 173.4
b) HF
ME (m) 84.4 82.4 87.8 82.7 85.0 84.1
MAX (m) 396.4 318.2 365.8 345.6 371.9 393.1
c) ITIBOV
ME (m) 86.7 84.0 87.1 58.6 61.8 64.6
MAX (m) 262.7 269.8 257.0 238.8 250.3 241.7
d) OGGM
ME (m) 98.0 97.1 82.4 93.1 90.0 96.7
MAX (m) 331.9 360.2 318.2 331.0 303.9 353.0
e) Composite Model
ME (m) 98.0 93.9 93.1 89.7 95.5 96.5
MAX (m) 312.3 310.1 353.3 329.7 317.8 333.7
Thickness (m)
0
80
160
240
320
400

Figure 8. Distribution of modelled ice thickness of Chhota Shigri Glacier (location shown in Fig.1) using AW3D30, SRTM-GL1, TanDEM-X, SRTM v4.1, NASADEM and MERIT. (a) Glabtop2; (b)HF; (c) ITBOV; (d) OGGM; (e) composite result. Mean (ME) and maximum (MAX) modelled ice thickness are given in each panel.

~~Following~~ As similar as the procedure of Farinotti et al. (2017), results from the four models are further composed to achieve the minimum MAE between the modelled and GPR thicknesses (Fig. 8e). The weights for each model in ten experiments are shown in Table S2. After composition, the mean thickness using different DEMs ranged from ~~77~~ 90 (acquired based on TanDEM-X) to ~~91~~ 98 m (acquired based on ~~NASADEM~~AW3D30). NASADEM and AW3D30 achieved minimum MAE, which are 36.7m and 44.1m, respectively. ~~The thickness error of the results based on NASADEM is best (median value 2.3 m), followed by TanDEM-X (median value -7.5 m). The minimum RMSE is for AW3D30 (45.9 m), followed by NASADEM with a RMSE of 47.7 m.~~The mean errors and median errors of all DEMs at the range of ±10m, except for that of AW3D30 and TanDEM-X at a level of around 20m. The spreads of error of 30-m DEMs are 33% smaller than those of 90-m DEM. Error spread from NASAM was minimum (75.1m), followed by AW3D30 (77.3 m).

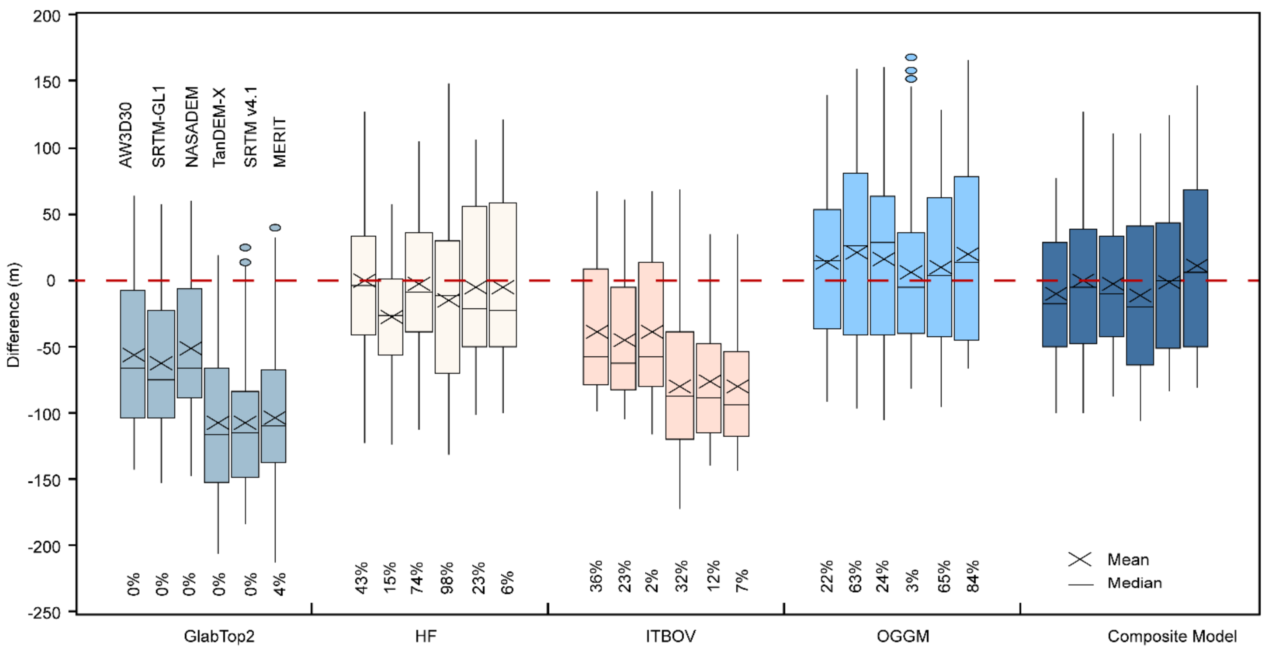


Figure 9. Point-by-point deviation comparison between the modelled and measured ice thickness from GlabTop2, HF, ITIBOV, OGGM and the composite result. In each group, the boxes are plotted in the order: AW3D30, SRTM-GL1, NASADEM, TanDEM-X, SRTM v4.1 and MERIT. Different models using the same DEM are aggregated by weights (~~labeled~~labelled at the bottom) to achieve minimum mean absolute error.

Table 2 RMSE (m) of modelled ice thickness compared with ground penetrating radar (GPR) measurements on each profile. Bold numbers denote the best model performance on each profile.

Model	DEM	No. of ground penetrating radar profiles					
		1	2	3	4	5	All
GlabTop2	AW3D30	56.6	84.5	**45.4**	**102.4**	103.3	79.7
	SRTMG-GL1	**54.8**	94.6	62.4	104.3	90.6	83.3
	NASADEM	53.8	**75.8**	48.5	104.6	**84.4**	**75.3**
	TanDEM-X	103.5	143.3	104.8	137.0	115.0	122.0
	SRTM v4.1	101.9	133.6	93.2	132.3	135.1	118.7
	MERIT	110.0	120.7	70.9	154.7	132.2	118.1
HF	AW3D30	**27.2**	29.8	83.7	30.7	92.5	61.6
	SRTM-GL1	28.4	58.4	**33.3**	30.6	88.1	**50.0**
	NASADEM	37.2	**26.5**	62.7	**30.1**	82.2	51.8
	TanDEM-X	50.7	63.4	60.9	66.5	**83.9**	65.4
	SRTM v4.1	49.3	36.4	74.9	44.8	86.4	61.3
	MERIT	51.4	33.6	87.6	42.4	86.2	65.8
ITBOV	AW3D30	69.3	61.9	**45.6**	**66.4**	70.8	**61.4**
	SRTMG-GL1	70.7	71.7	52.6	70.6	61.4	64.8
	NASADEM	**67.3**	**61.3**	58.3	80.3	**57.1**	65.3
	TanDEM-X	98.3	116.4	61.1	91.2	109.7	94.0
	SRTM v4.1	102.9	114.5	51.8	73.4	111.3	88.3
	MERIT	108.3	117.6	54.8	80.4	112.7	91.9
OGGM	AW3D30	**32.0**	47.2	**80.5**	**62.5**	**34.6**	58.9
	SRTMG-GL1	33.5	56.9	101.6	65.1	41.1	70.2
	NASADEM	32.2	55.3	92.3	68.8	43.6	67.1

	TanDEM-X	42.7	**35.6**	86.1	**52.2**	35.4	**58.4**
	SRTM v4.1	37.4	46.1	81.4	68.7	39.9	61.6
	MERIT	38.3	52.1	104.1	53.8	49.7	69.7
Composite	AW3D30	37.2	27.9	**44.4**	45.8	68.3	**45.8**
	SRTMG-GL1	39.1	29.8	64.3	63.3	**47.5**	52.7
	NASADEM	**34.9**	**22.1**	65.4	**41.1**	70.6	51.3
	TanDEM-X	50.9	46.3	60.5	57.9	69.0	57.5
	SRTM v4.1	41.8	41.4	59.3	59.5	55.1	53.1
	MERIT	42.0	49.1	85.3	53.8	55.3	62.5

4. Discussion

4.1 Influence of glacier elevation change ~~Factors related to~~on the ~~differences~~ assessment of DEMs~~: glacier elevation change, terrain~~

The identified extreme outliers (Fig. 3) are mostly located in the glacier terminus, high elevation and high slope regions (Fig. 10a−b). Extreme glacier melt, such as in ~~southeastern~~south-eastern Tibet, and surges, as observed in the Karakoram, can also lead to dramatic elevation changes, resulting in large difference (Fig. 10c). This glacier elevation change effect is also reflected in the spatial distribution of difference (Fig. 5), elevation bins (Fig. 6c) and glacier zones (Fig. 7). The differences at lower elevations are ~~negative~~positive, and generally ~~increase~~ decrease with elevation, consistent with the fact that glaciers melt at lower elevations and accumulate at higher elevations (Cuffey and Paterson 2010). The differences ~~in the NASADEM and TanDEM-X data~~ of all DEMs with elevation and glacier zones comply with these features (Fig. 6c and Fig. 7). NASADEM was acquired in 2000 and TanDEM-X was acquired in 2010−2015, and the value of NASADEM is more ~~negative~~ positive than TanDEM-X in the ablation zone, as would be expected. ~~By making a comparison between SRTM-GL1 and NASADEM from the same original data,[阴8] we conclude that the negative differences of the other four DEMs through the elevation bins may be related to absolute vertical shift. MERIT shows less improvement over SRTM v4.1 in glacierized terrain than in the flat regions in terms of both absolute and relative accuracy (Yamazaki et al. 2017).~~ The relatively more ~~negative~~ positive and larger values of ME and STD along the Hindu Kush-Himalaya, southern Tibet (Fig. 5) and negative ME values in the West Kunlun and Karakroam ~~and in glacier zones~~ (Fig. ~~6~~5) are also related to glacier elevation change (Hugonnet et al. 2021)~~(Hugonnet et al. 2021)~~.

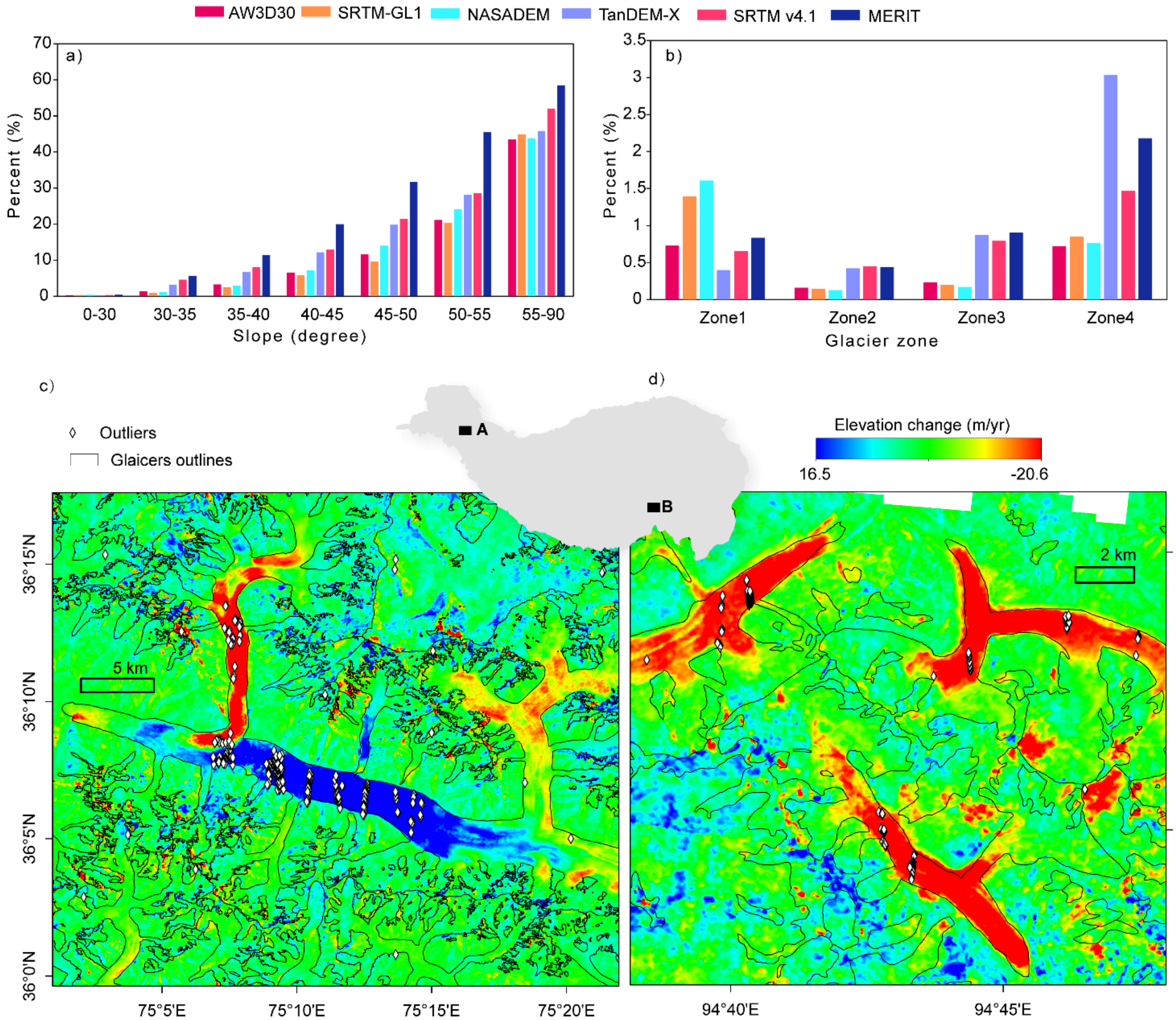
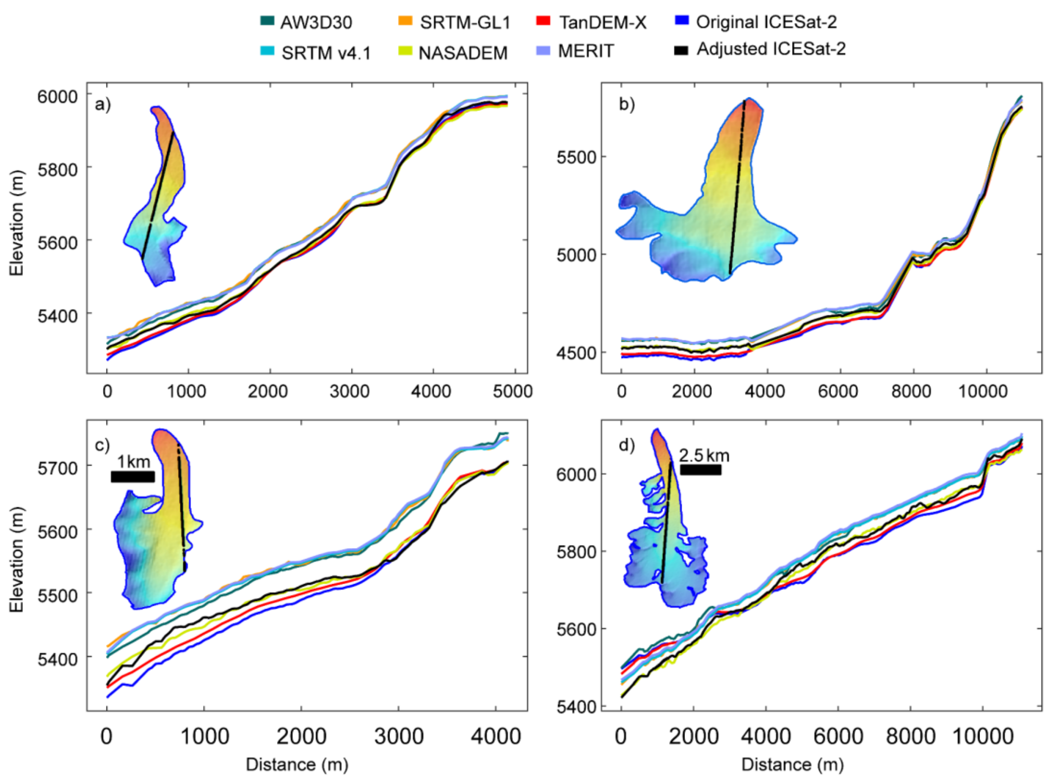


Figure 10. Distribution of excluded extreme outliers. Proportion of outliers accounting for total number in slope bins (a) and each glacier Zone (b). Examples of locations of excluded points overlaid with glacier surface elevation change in Karakoram (c) and southern TP (d). Locations of these two examples are labeled A and B in the central insert. Glacier elevation change data covering 2000−2019 is from Shean et al. (2020).

Table 3 Comparisons of differences between four SRTM based DEMs and ICESat-2 elevation over glacier zones before and after adjustment. ICESat-2 data acquired in February are used to calculate the differences. Glacier zones are defined according to Fig. 8e.

Item	Zone	Before (m)				After (m)			
		SRTM-GL1	NASADEM	SRTM v4.1	MERIT	SRTM-GL1	NASADEM	SRTM v4.1	MERIT
Mean error	1	27.8	26.6	23.9	27.4	13.4	12.1	9.7	12.8
	2	10.0	9.1	8.3	10.5	2.9	2.0	1.7	3.6
	3	-0.2	-0.8	0.7	1.7	-3.5	-3.9	-2.2	-1.4
	4	-12.4	-12.0	-8.5	-6.0	-9.3	-10.1	-7.5	-5.3
Absolute mean error	1	33.8	33.0	36.3	34.3	17.8	16.8	20.5	18.5
	2	16.4	16.1	19.0	18.0	9.0	8.6	11.9	10.9
	3	11.2	10.8	14.4	13.8	8.3	7.7	11.6	10.7
	4	18.0	16.0	23.4	20.6	13.0	12.1	19.6	16.9
Standard deviation	1	37.7	37.9	53.5	39.4	20.3	20.3	40.1	22.8
	2	20.6	20.8	36.1	26.8	12.0	11.6	29.2	21.8
	3	14.6	14.5	23.7	22.7	10.7	9.8	20.7	19.3
	4	29.6	27.4	39.5	29.4	20.9	20.8	36.9	26.0
RMSE	1	46.9	46.3	58.6	48.0	24.3	23.7	41.2	26.2
	2	22.9	22.7	37.0	28.8	12.4	11.8	29.2	22.1
	3	14.6	14.5	23.7	22.7	11.3	10.6	20.8	19.4
	4	32.1	29.9	40.4	30.0	22.9	23.2	37.6	26.6

After ~~error correction~~removing the glacier elevation change ~~,~~using the glacier elevation change dataset covering 2000−~~2019~~ 2018 (Shean et al. ~~2019~~2020)~~.~~, ~~T~~the mean difference in Zone 1 and Zone 2 decreased sharply by ~~~13~~ 14 m and ~7 m for the SRTM based DEMs, respectively ~~. The ME of NASADEM reached as low as 0.1 m in Zone 1~~(Table 3). ~~Improvements are also obvious along the elevation profiles of the four glaciers selected across the TP shown in Fig. 11. Before correction, the ICESat-2 elevation is lower than DEMs elevation in Fig. 11a−c and higher in Fig. 11d. After correction, the ICESat-2 and NASADEM profiles nearly overlap.~~ However, similar improvements are not obvious in Zone 3 and Zone 4. This may be related to the slight elevation change in the accumulation region (Brun et al. 2017; Shean et al. 2020), and high uncertainty due to steeper slopes and higher elevations (Fig. 6b−c~~)~~). MAE, STD and RMSE all improved a lot in four regions after this adjustment.



~~**Figure 11.** Elevation profiles of the original and adjusted six DEMs and ICESat-2 along the ICESat-2 tracks on four selected glaciers across the study region. Locations of glaciers in a–d are labelled 1–4 in Fig. 1, respectively.~~[陈9]

ICESat-2 data covering the period October 2018 to October 2020 repeat every 91 days. Therefore, variations of ICESat-2 elevation data caused by glacier fluctuations will have influenced the error statistics (Fig. ~~12a~~11a). Precipitation on the TP mainly occurs in June−August (Maussion et al. 2014). Hence, after precipitation accumulation on glaciers in spring and summer, the elevation ~~have~~ increased, and the mean difference decreased. With little accumulation, the glacier melt and sublimate in autumn and winter (Li et al. 2018), meaning decreased elevation, and an increase in the mean difference. However, the magnitude of these changes is much smaller, at a level of less than 3 m (Fig. ~~12~~11a), compared with the large ME, MAE and RMSE magnitude of most of the DEMs (with the exceptions of TanDEM-X and NASADEM) (Fig. 4). When taking all points from different seasons into consideration, ICESat-2 dataset gives average elevation over the 2018-2020 period, the seasonal effects could also partly cancel each other out. If only the ICESat-2 data from February was used (Table 3), NASADEM and TanDEM-X still perform better than others. Therefore, we conclude that the seasonal fluctuations of ICESat-2 data have no influence on the ~~performance~~ assessments of the DEMs. Additionally, the atmospheric forward scattering and subsurface scattering of ICESat-2 photons, which are not quantified in ATL06 due to lack of ancillary ~~,~~data, may also lead to a biased estimate of elevation (Smith et al. 2019).

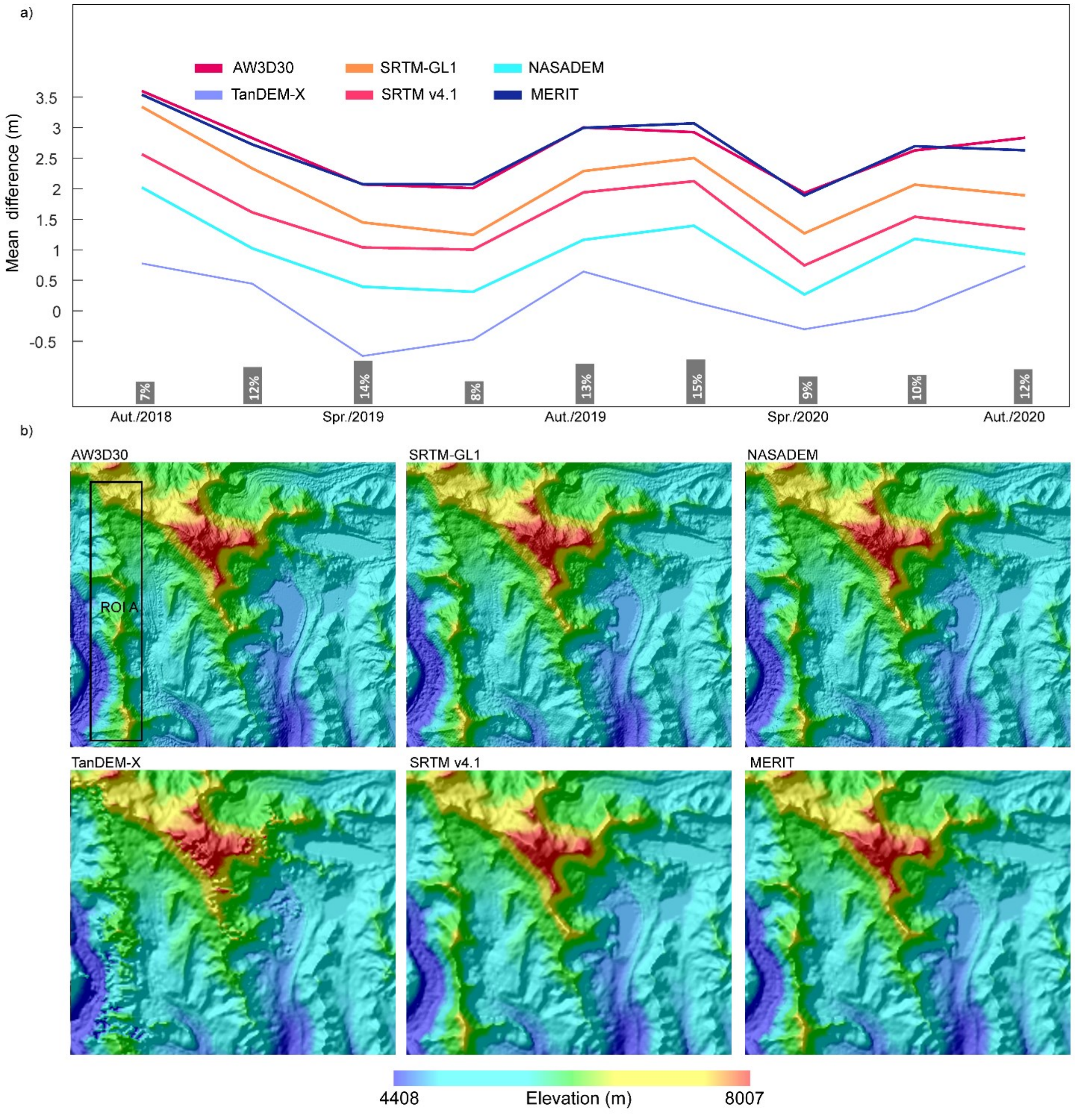


Figure ~~12~~11. Influence on elevation differences between ICESat-2 and six DEMs from glacier elevation change and terrain factors. (a) Mean absolute difference between six DEMs and ICESat-2 in different seasons during 2018−2020. The spring, summer, autumn, and winter are defined as March−May, June−August, September−November and December−February, respectively. The histogram at the bottom shows the percentage of the total number of points in each season. (b) Examples of elevation and shaded relief of six DEMs in the Shisha Pangma region. The rectangle denotes the area of interest.

4.2 Influence of terrain on the assessment of DEMs

The elevation differences have a strong dependence on terrain factors (Fig. ~~7a~~6a–b). The differences with aspect show contrasting features to the distribution of measurements in different aspects (Fig. ~~7d~~6d). The largest errors are concentrated in the north aspect, as was also reported in previous studies (Gorokhovich and Voustianiouk 2006; Shortridge and Messina 2011), in which they were attributed to the orientation of the sensor (Gdulová et al. 2020; Shortridge and Messina 2011). However, here, the data from different sensors all show this aspect dependence, and we infer that it may be related to the accordant distribution of data in different slopes with aspect. There are many more measurements with steeper slopes in the north aspect, and less measurements with flatter slopes in the south aspect (Fig. ~~13~~12). The error and spread become larger with steeper slopes (Fig. 7b), as also reported by Liu et al. (2019) and Uuemaa et al. (2020), maybe due to geometric deformation or shadow (Liu et al. 2019)~~–~~. Therefore, the error variation with aspect tends to be related to steeper slopes (Gdulová et al. 2020; Gorokhovich and Voustianiouk 2006).

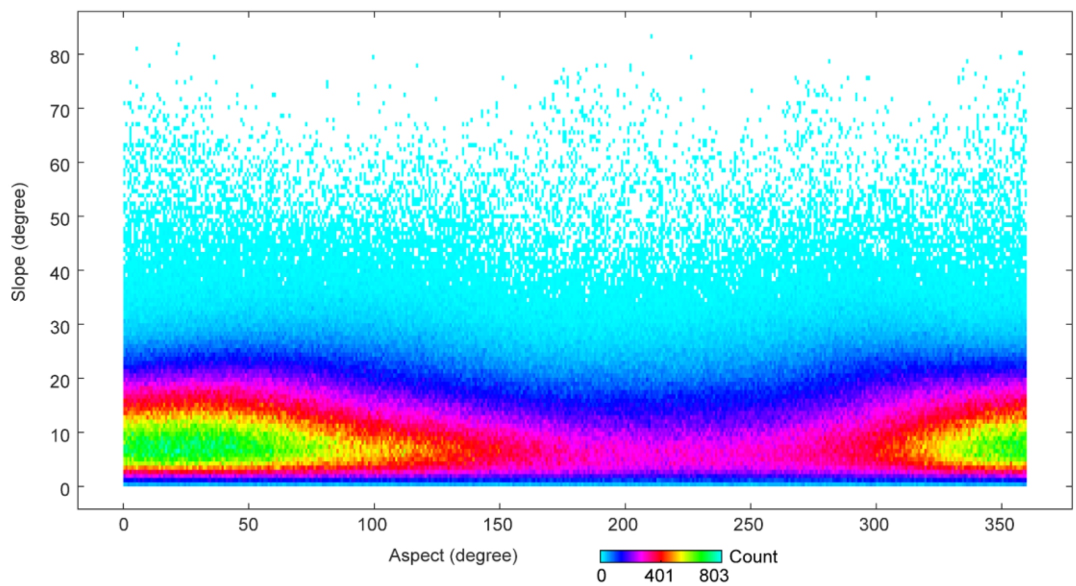


Figure ~~13~~12. Distribution of measurements in different ~~slopes~~ aspect against~~with~~ ~~aspect~~slope.

Almost half the points in the 55°–90° slope region are identified as extreme outliers (Fig. 10a). Differences also show large discrepancies for all DEMs in the steeply sloping regions where voids and large errors are frequent (Falorni 2005). Steep slopes combined with low resolution led to variations in the spread of differences in Fig. 6b. Spreads of differences were larger on steep slopes for the 90-m DEMs than those of the 30-m DEMs. Intra-pixel variation aggravates this effect in steeply sloping regions (Uuemaa et al. 2020), lower resolution or reduced pixel DEMs smooth the terrain details and lead to inaccurate elevation compared with the 20-m footprint of ICESat-2 points. The spread and the number of outliers gradually increased with the slope, especially for the TanDEM-X case (Fig. 7b). Using the terrain in the rugged Shishapangma region (Fig. 12b) as an example, we can see that the elevation from TanDEM-X suffers from serious errors along the ridge at high elevations with the output almost blurred. Even so, TanDEM-X still has overall accuracy advantages over SRTM v4.1 and MERIT, indicating the high quality of TanDEM-X in low relief regions (Fig. 7b).

4.2 Influence of misregistration on the assessment of DEMs

Pixel of different DEMs at same location may mismatch each other. This mi~~Mi~~sregistration among DEMs, which has been ignored in previous research (González-Moradas and Viveen 2020; Liu et al. 2019), is an important error source when looking at DEM differences (Hugonnet et al. 2021; Van Niel et al. 2008). This study intends to give direct insights about the quality of uncorrected DEM products, so the misregistration problem was not tackled before the evaluations were carried out. However, the influence of misregistration was evaluated. According to the sinusoidal relationship between aspect and error differences between two DEMs (Van Niel et al. 2008), using the co-registration method in Nuth and Kääb (2011) and ICESat-2 points outside the glaciers, offset pixels relative to ICESat-2 at the $1^{\circ}\times1^{\circ}$ grid scale were estimated across the TP~~ (Fig. 14)~~. Misregistration was found to be less than one pixel (Figure ~~14~~13). Offset pixel of SRTM-GL1 relative to ICESat-2 is largest; offset pixels of the other DEMs are all at less than 0.2 pixel. Considering that only the cell centre value was used, sub-pixel shift may have little influence (Van Niel et al. 2008)~~(Van Niel et al. 2008)~~. ~~A comparison of errors before and after correcting sub-pixel misregistration confirms this conclusion (Fig. 15b).~~

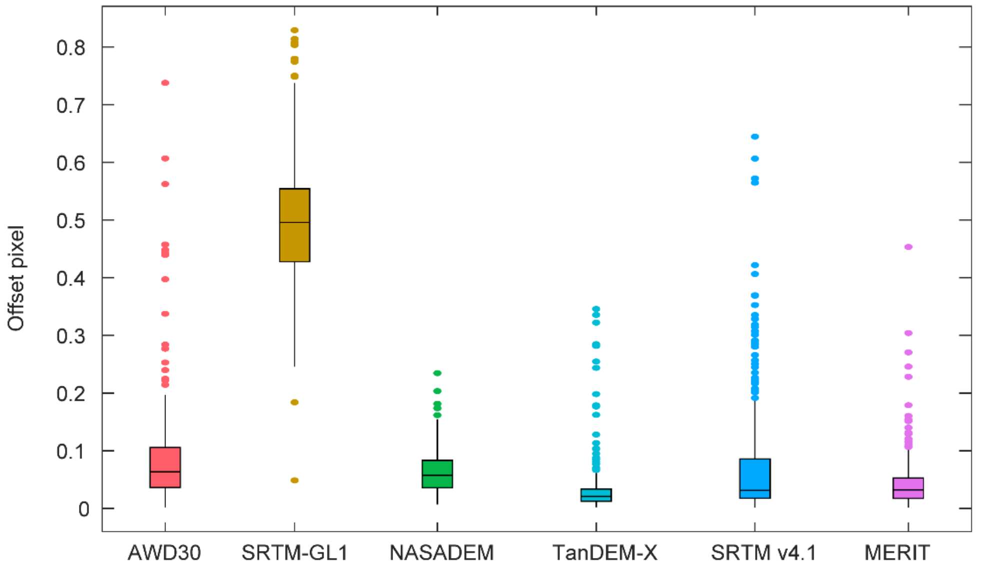


Figure ~~14~~13. Distribution of offset pixels of DEMs relative to ICESat-2 on a 1°×1° grid. Only grid squares with R^2 greater than 0.5 and number greater than 1000 are considered.

~~Misregistration was found to be worst in eastern Kunlun Mountains and Qilian Mountains where only a small number of glaciers developed (Fig. 1) and was slight in the south of the TP. All the DEMs mismatch by less than one pixel, except for AW3D30, which had the worst misregistration in the north and inner TP (Fig. 14). Considering that only the cell center value was used, sub-pixel shift may have little influence. A comparison of errors before and after correcting sub-pixel misregistration confirms this conclusion (Fig. 15b). The probability distribution of difference before and after co-registration was almost the same, as were ME, MAE, STD and RMSE (Table S2). However, examples of pixel misregistration strongly affected the probability distribution, except for AW3D30, SRTM-GL1, SRTM v4.1 and MERIT (Fig. 15a); STD changed by 1–3 m, while ME, MAE and RMSE changed by less than 1.2 m (Table S2). The probability distribution symmetry varies. Hence, we supposed that the symmetry variations of difference compensate the effect of offset. Since glaciers are distributed among the mountains with different aspects, if we shift the DEM in the x- and y-directions, the increased differences would be compensated for by decreased elevation differences (Fig. 15c). Therefore, the large errors remaining in these four DEMs should be due to systematic deviations (Han et al. 2021), rather than the influence of misregistration.~~[陈10]

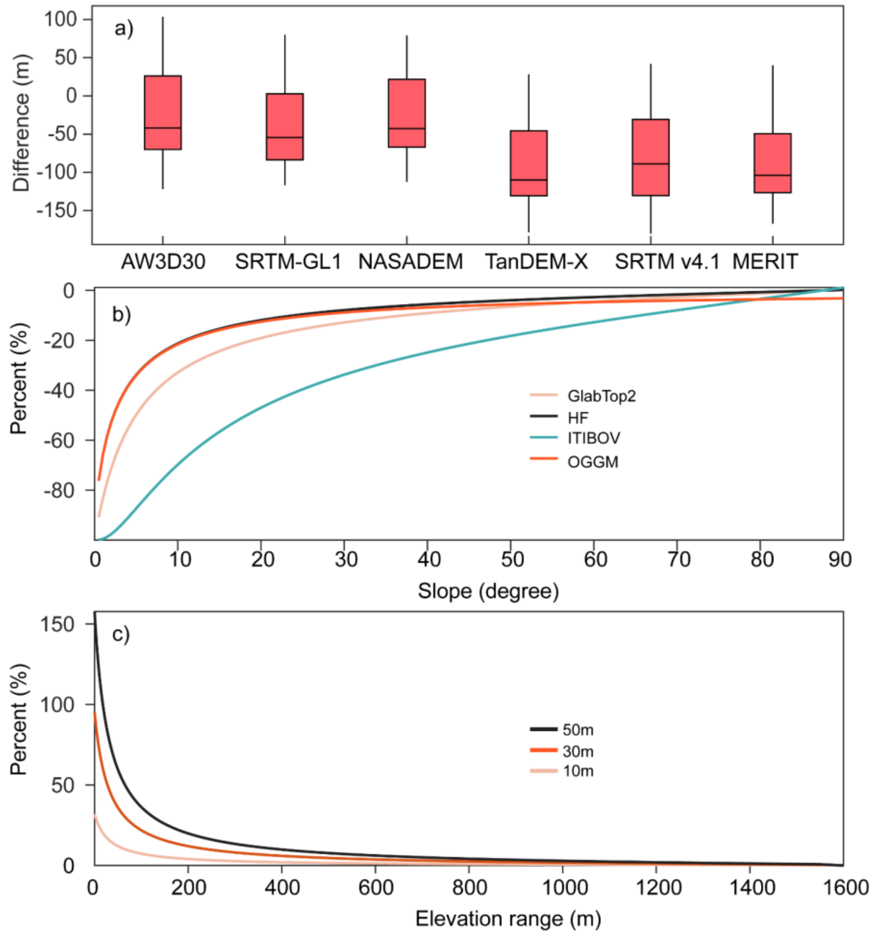
4.3 Influence of DEMs on ice thickness estimated by ITIMs

Even with the same parameters, the same model using different DEMs have different outcomes (Figure 8 and 9). The DEM indeed influence the performance of the ITIMs. Different DEMs resulted in differences in ~~ITIM~~ maximal and minimum mean ITIM ice thickness ~~with~~ at a range of ~~32–65%~~3.6-32 m (Fig. ~~10~~8).

Generally, the outcome with GlabTop2 and ITBOV using 30-m DEMs is 51% and 43% better than with the 90-m DEMs ~~,~~ ~~with~~in mean error ~~differences of 52% and 45%~~, respectively. ~~This is different from the conclusion that improving only DEM resolution, without calibrating the shape factor, did not benefit the model result in GlabTop2 (Ramsankaran et al. 2018). But when we used a calibrated shape factor of 0.66 as suggested by Ramsankaran et al. (2018), the model results from the 30-m DEMs were still better than those from the 90-m DEMs (Fig. 16a).~~ [陈11]With GlabTop2, elevation data was used to determine not only the slope, but also the shear stress (Frey et al. 2014). A +5° error in slope caused more than a –34.1 % difference in the output for slopes of less than 20°. Additionally, relative elevation errors had an enormous impact (Fig. ~~16~~ 14 ~~c~~b). For glaciers with an elevation range of less than 400 m, which accounted for 41% of the total number and 5% of the total area over the TP, +10, +30, and +50 m errors in elevation range caused more than +2%, +6% and +10% differences in output. Such errors in elevation range had greater influence (Fig. ~~19b~~14b), especially for small glaciers, which have smaller elevation ranges. These two errors propagate and lead to a much larger overall error (Table 3). Thus, GlatbTop2 with NASADEM and AW3D30 as the best quantity input achieved the best RMSE in comparison with GPR measurements. In contrast to the other ITIMs, the ITIBOV model directly estimated the ice thickness at each grid cell according to cell velocity information without interpolation. The slope sensitivity of ITIBOV is higher than that of GlabTop2, with a +5° error in slope causing more than a –71.4% difference in the output for slopes of less than 20° (Fig. ~~16b~~14a). The flatter the slope, the more sensitive ITIBOV is to the slope error (Fig. ~~16b~~14a). Although along- and across-track slope data are provided in the ICESat-2 ATL06 product, they are incompatible with the slope estimated from DEMs due to their different data formats and algorithms used (Burrough and McDonell 1998; Smith et al. 2019). Moreover, the surface terrain of glaciers changes with time due to accumulation, melting and motion (Dehecq et al. 2018; Shean et al. 2020). Nevertheless, the accuracy of the DEMs estimated here could also provide some information about slope accuracy. The better ~~relative~~ accuracy of NASADEM and AW3D30 means that ITIBOV with these DEMs as input led to the relatively best outcomes (Table 2).

For HF and OGGM, the modelled results did not show large differences when 30-m DEMs ~~were replaced~~comparing with 90-m resolution DEMs (Fig. 9): means that high spatial resolution improved the outcome little (Pelto et al. 2020). For the HF model, elevation data was used for convergence calculation of apparent mass balance and mean slope in elevation bins (Farinotti et al. 2009; Farinotti et al. 2019), whereas, for OGGM, it is used to extract flowlines, shear stress at flowlines and mass balance at an elevation (Maussion et al. 2019). ~~The relative accuracy of DEMs was more vital than absolute accuracy for these two models.~~ Although NASADEM and TanDEM-X had the large advantage of ~~absolute~~ accuracy, the output of HF and OGGM using these two DEMs did not have much advantage over that using the other DEMs (Fig. 9). The STD of RMSE values for HF and OGGM using six DEMs are ~~7.0~~6.2 and ~~6.1~~4.9 m, respectively (Table 2). STD of mean ice thickness by HF and OGGM using six DEMs are 1.1 and 1.96.0 m (Table Fig.2 8). The HF and OGGM models are not very sensitive to the DEM absolute accuracy. The performance of AW3D30 in OGGM and SRTM-GL1 in HF is even slightly better than NASADEM in these two models (Table 2). Specifically, the better performance of SRTM-GL1 should be attributed to the calculation of slope. Though the model has high sensitivity to the slope (Fig. 116a4a), the mean slope in each elevation band was used, defined as a tangent of the width and elevation difference in the elevation bin (Huss and Farinotti 2012).

The different models have various levels of robustness to the quality of the input DEMs. When the results from different models are ensembled, the influence of the input DEM manifests (Fig. 9 and Table 2). The RMSE of ITIMs from 30-m DEMs was 16.8% less than that from 90-m DEMs. Models using AW3D30 and NASADEM, equipped with higher resolution and better accuracy, achieved the best outcomes. This conclusion is of significance for ice thickness inversion models using DEMs in TP. However, it should be noted that the result may be not suitable for studies in other glacierized mountainous regions. Because various errors exit in DEMs, such as speckle noise, stripe noise and absolute bias; they behave different across the Earth (Yamazaki et al. 2017). But our method to assess the accuracy of DEMs is repeatable in different regions, combining with the recent released glacier elevation change data on Earth (Hugonnet et al. 2021).


a)
Difference (m)
100
50
0
-50
-100
-150
AW3D30
SRTM-GL1
NASADEM
TanDEM-X
SRTM v4.1
MERIT
b)
Percent (%)
0
-20
-40
-60
-80
GlabTop2
HF
ITIBOV
OGGM
0
10
20
30
40
50
60
70
80
90
Slope (degree)
c)
Percent (%)
150
100
50
0
50m
30m
10m
0
200
400
600
800
1000
1200
1400
1600
Elevation range (m)

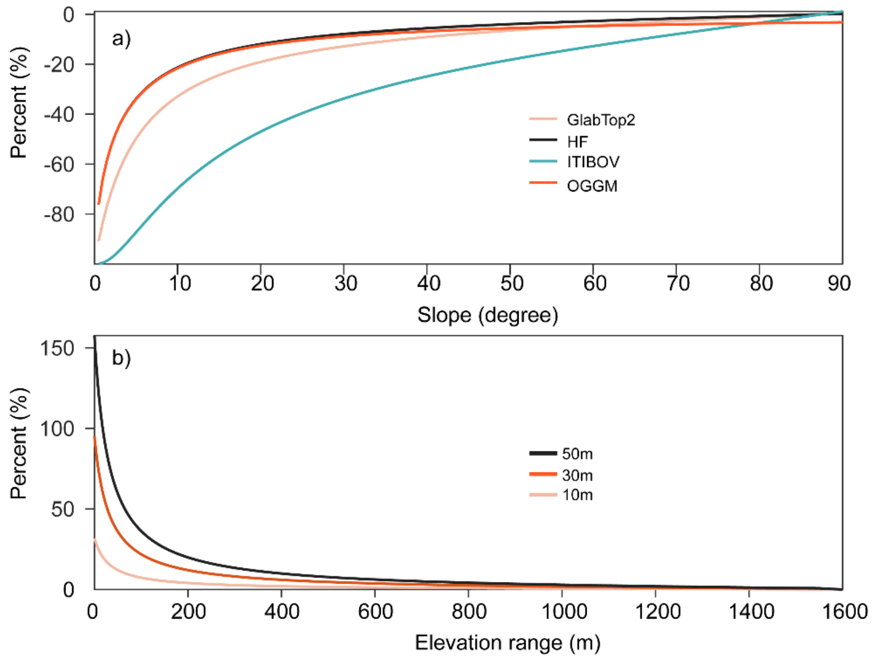


Figure ~~16~~14. Sensitivity test of~~ shape factor,~~ slope and elevation on ice-thickness inversion models. ~~(a) Difference between modelled thickness and GPR measurements when a calibrated shape factor of 0.66 was used in GlabTop2; ~~(~~b~~a) Percentage difference of modelled ice thickness from GlabTop2, HF, ITIBOV and OGGM when there is +5° slope error; (~~c~~b) Percentage difference of modelled ice thickness from GlabTop2 when the elevation range error is +10, +30 and +50 m for different elevation ranges.

~~(Fielding et al. 1994; Liu et al. 2019)~~

~~The different models have various levels of robustness to the quality of the input DEMs. When the models are comprehensively utilized, [陈12]the influence of the input DEM manifests itself (Fig. 9 and Table 2). The RMSE of ITIMs from 30-m DEMs was 16.8% less than that from 90-m DEMs. Models using AW3D30 and NASADEM, equipped with higher resolution and better relative accuracy, achieved the best outcomes. However, it should be noted that the large misregistration in AW3D30 in the northern TP may lead to the mismatch between terrain and glacier outlines. This will lead to an overestimation of slope and a consequential underestimation of ice thickness (Huss and Farinotti 2012), due to the mountain terrain being relatively steeper than the glaciers[陈13].~~

5. Conclusions

In the present study, six DEMs (that is AW3D30, SRTM-GL1, NASADEM, TanDEM-X, SRTM-GL1 and MERIT) from different sensors with different spatial resolutions were evaluated using ICESat-2 data. The influence of glacier dynamics, terrain and misregistration on the DEM accuracy were analysed. Out of the three 30-m DEMs, NASADEM was the best performer with a small ME of –1.00.9 m and a RMSE of 12.6 m. Out of the three 90-m DEMs, TanDEM-X performed best with an ME of –0.1 and a RMSE of 15.1 m. The quality of TanDEM-X was stable and unprecedented on shallow slopes, but suffered from serious problems on steep slopes, especially along the steep ridges. For AW3D30, a systematic vertical and horizontal offset existed on glacier terrain, however, it still has a similar relative accuracy to NASADEM, and even better in STD, MAE and RMSE when not considering the effect of glacier dynamics. SRTM-based DEMs (that is SRTM-GL1, SRTM v4.1 and MERIT) (~36 15 m RMSE) were inferior to NASADEM, although, when influence of glacier variations werewas excluded, all of their errors were reduced in the ablation zone. MERIT shows little improvement over SRTM v4.1 in glacierized terrain. The influence of glacier elevation change on the elevation difference is larger for DEMs acquired earlier, at low elevations and in the ablation region. However, this does not influence the conclusion that NASADEM performed the best, followed by TanDEM-X but with serious outlier in high elevation .region. For all the DEMs, the errors increased from the south-aspect slope to north-aspect slope, controlled by the increasing error with slope. Misregistration errors in the glacier-rich region are mostly within one pixel, benefiting from the 20 m footprint of ICESat-2, relative to the 30- or 90-m resolution DEMs, and only have a smalllittle influence on the evaluation.

The effect influence of DEM accuracy on ice-thickness inversion models depends on the model properties. Generally, a higher resolution DEM was helpful for better model outcomes due to the intra-pixel influence. For the widely used GlabTop2 model, which is sensitive to the accuracy of elevation and slope, using NASADEM, with the highest absolute accuracy, as the input, facilitated the best outcome. Although the OGGM and HF models are less sensitive to the quality of DEM, the use of NASADEM or AW3D30, both with a high relative accuracy, was still favorablefavourable. Among the four ice-thickness inversion models, ITIBOV was the most sensitive to slope accuracy. Ice-thickness inversion models using AW3D30 or NASADEM as input gave the best outcomes. These two DEMs also perform the best when four ice-thickness inversion results were aggregated by the minimum MAE optimization method to form an ensemble.

Considering the influence of inconsistency in data acquisition time on generating glacier terrain, we suggest that NASADEM is the best choice for ice-thickness inversion models over the TP. AW3D30 could be a good substitute if its systematic shift was correctedlimited by their mixed acquiring dates. TanDEM-X is an appropriate alternative for glaciological research focusing on the flat glacier terminus, but it requires further improvement for use in steep terrain or for ice-thickness inversion.

Code/Data availability

ICESAT-2 ATL 06 is available at https://nsidc.org/data/atl06; AW3D30 is available at https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm; SRTM-Gl1 is available at https://earthexplorer.usgs.gov/; NASADEM is available at https://search.earthdata.nasa.gov/ ; TanDEM-X 90m is available at https://download.geoservice.dlr.de/TDM90/; SRTM v4.1 is available at https://drive.google.com/drive/folders/0B_J08t5spvd8RWRmYmtFa2puZEE; MERIT is available at http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/; Glacier elevation change data is available at https://zenodo.org/record/3600624. The code for processing the ICESat-2 data are available at https://github.com/cnugis/ProcessICESat-2. The database containing elevation, slope and aspect and some other fields in MATLAB mat format is available at https://zenodo.org/record/5267309.

Author contribution

W.F Chen, T.D Yao and GQ. Zhang designed the outline of this study. W.F Chen processed the data and make all the figures. Fei Li estimate the ice thickness using OGGM. All authors contributed to writing the paper.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study was supported by grants from the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0201), the Strategic Priority Research Program (A) of the Chinese academy of Sciences (XDA20060201) and the Natural Science Foundation of China (41871056).

References

Allen, S.K., Zhang, G., Wang, W., Yao, T., & Bolch, T.: Potentially dangerous glacial lakes across the Tibetan Plateau revealed using a large-scale automated assessment approach. Sci. Bull., 64, 435-445, https://doi.org/10.1016/j.scib.2019.03.011, 2019

Altunel, A.O.: Evaluation of TanDEM-X 90 m Digital Elevation Model. Int. J. Remote Sens., 40, 2841-2854, https://doi.org/10.1080/01431161.2019.1585593, 2019

Bachmann, M., Kraus, T., Bojarski, A., Schandri, M., Boer, J., Busche, T., Bueso Bello, J.L., Grigorov, C., Steinbrecher, U., Buckreuss, S., Krieger, G., & Zink, M.: The TanDEM-X Mission Phases - Ten Years of Bistatic Acquisition and Formation Planning. IEEE J-STARS, 1-1, https://doi.org/10.1109/JSTARS.2021.3065446, 2021

Bhambri, R., Bolch, T., & Chaujar, R.K.: Mapping of debris-covered glaciers in the Garhwal Himalayas using ASTER DEMs and thermal data. Int. J. Remote Sens., 32, 8095-8119, https://doi.org/10.1080/01431161.2010.532821, 2011

Brun, F., Berthier, E., Wagnon, P., Kaab, A., & Treichler, D.: A spatially resolved estimate of High Mountain Asia glacier mass balances, 2000-2016. Nat. Geosci., 10, 668-673, https://doi.org/10.1038/ngeo2999, 2017

Brunt, K.M., Neumann, T.A., & Smith, B.E.: Assessment of ICESat-2 Ice Sheet Surface Heights, Based on Comparisons Over the Interior of the Antarctic Ice Sheet. Geophys. Res. Lett., 46, 13072-13078, https://doi.org/10.1029/2019GL084886, 2019

Burrough, P.A., & McDonell, R.A.: Principles of Geographical Information Systems. Oxford University Press, New York, 190 pp., 1988

Chen, W., Yao, T., Zhang, G., Li, S., & Zheng, G.: Accelerated glacier mass loss in the largest river and lake source regions of the Tibetan Plateau and its links with local water balance over 1976–2017. J. Glaciol., 1-15, https://doi.org/10.1017/jog.2021.9, 2021

Crippen, R., Buckley, S., Belz, E., Gurrola, E., Hensley, S., Kobrick, M., Lavalle, M., Martin, J., Neumann, M., Nguyen, Q., Rosen, P., Shimada, J., Simard, M., & Tung, W.: NASADEM global elevation model: methods and progress. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B4, 125–128, https://doi.org/10.5194/isprs-archives-XLI-B4-125-2016, 2016.

Cuffey, K., & Paterson, W.S.B.: The Physics of Glaciers 4th Edition. Academic Press, 2010

Dehecq, A., Gourmelen, N., Gardner, A.S., Brun, F., Goldberg, D., Nienow, P.W., Berthier, E., Vincent, C., Wagnon, P., & Trouvé, E.: Twenty-first century glacier slowdown driven by mass loss in High Mountain Asia. Nat. Geosci., 12, 22–27, https://doi.org/10.1038/s41561-018-0271-9, 2019

Fahnestock, M., Scambos, T., Moon, T., Gardner, A., Haran, T., & Klinger, M.: Rapid large-area mapping of ice flow using Landsat 8. Remote Sens. Environ., 185, 84-94, https://doi.org/10.1016/j.rse.2015.11.023, 2016

Falorni, G.: Analysis and characterization of the vertical accuracy of digital elevation models from the Shuttle Radar Topography Mission. J. Geophys. Res.-Earth, 110, https://doi.org/10.1029/2003JF000113, 2005

Farinotti, D., Brinkerhoff, D.J., Clarke, G.K.C., Fürst, J.J., Frey, H., Gantayat, P., Gillet-Chaulet, F., Girard, C., Huss, M., Leclercq, P.W., Linsbauer, A., Machguth, H., Martin, C., Maussion, F., Morlighem, M., Mosbeux, C., Pandit, A., Portmann, A., Rabatel, A., Ramsankaran, R., Reerink, T.J., Sanchez, O., Stentoft, P.A., Singh Kumari, S., van Pelt, W.J.J., Anderson, B., Benham, T., Binder, D., Dowdeswell, J.A., Fischer, A., Helfricht, K., Kutuzov, S., Lavrentiev, I., McNabb, R., Gudmundsson, G.H., Li, H., & Andreassen, L.M.: How accurate are estimates of glacier ice thickness? Results from ITMIX, the Ice Thickness Models Intercomparison eXperiment. The Cryosphere, 11, 949-970, https://doi.org/10.5194/tc-11-949-2017, 2017

Farinotti, D., Brinkerhoff, D.J., Fürst, J.J., Gantayat, P., Gillet-Chaulet, F., Huss, M., Leclercq, P.W., Maurer, H., Morlighem, M., Pandit, A., Rabatel, A., Ramsankaran, R., Reerink, T.J., Robo, E., Rouges, E., Tamre, E., van Pelt, W.J.J., Werder, M.A., Azam, M.F., Li, H., & Andreassen, L.M.: Results from the Ice Thickness Models Intercomparison eXperiment Phase 2 (ITMIX2). Front. Earth. Sci., 8, https:// doi: 10.3389/feart.2020.571923, 2021

Farinotti, D., Huss, M., Bauder, A., Funk, M., & Truffer, M.: A method to estimate the ice volume and ice-thickness distribution of alpine glaciers. J. Glaciol., 55, 422-430, https://doi.org/10.3189/002214309788816759, 2009

Farinotti, D., Huss, M., Furst, J.J., Landmann, J., Machguth, H., Maussion, F., & Pandit, A.: A consensus estimate for the ice thickness distribution of all glaciers on Earth. Nat. Geosci., 12, 168-173, https://doi.org/10.1038/s41561-019-0300-3 ,2019

Fielding, E., Isacks, B., Barazangi, M., & Duncan, C.: How flat is Tibet? Geology, 22, 163-167, https://doi.org/10.1130/0091-7613(1994)022<0163:HFIT>2.3.CO;2, 1994

Frey, H., Machguth, H., Huss, M., Huggel, C., Bajracharya, S., Bolch, T., Kulkarni, A., Linsbauer, A., Salzmann, N., & Stoffel, M.: Estimating the volume of glaciers in the Himalayan–Karakoram region using different methods. The Cryosphere, 8, 2313-2333, https://doi.org/10.5194/tc-8-2313-2014, 2014

Frey, H., & Paul, F.: On the suitability of the SRTM DEM and ASTER GDEM for the compilation of topographic parameters in glacier inventories. Int. J. Appl. Earth. Obs., 18, 480-490, https://doi.org/10.1016/j.jag.2011.09.020, 2012

Frey, H., Paul, F., & Strozzi, T.: Compilation of a glacier inventory for the western Himalayas from satellite data: methods, challenges, and results. Remote Sens. Environ., 124, 832-843, https://doi.org/10.1016/j.rse.2012.06.020, 2012

Fujita, K., Suzuki, R., Nuimura, T., & Sakai, A.: Performance of ASTER and SRTM DEMs, and their potential for assessing glacial lakes in the Lunana region, Bhutan Himalaya. J. Glaciol., 54, 220-228, https://doi.org/10.3189/002214308784886162, 2017

Furian, W., Loibl, D., & Schneider, C.: Future glacial lakes in High Mountain Asia: an inventory and assessment of hazard potential from surrounding slopes. J. Glaciol., 1-18, https://doi.org/10.1017/jog.2021.18, 2021

Gantayat, P., Kulkarni, A.V., & Srinivasan, J.: Estimation of ice thickness using surface velocities and slope: case study at Gangotri Glacier, India. J. Glaciol., 60, 277-282, https://doi.org/10.3189/2014JoG13J078 , 2014

Gardner, A.S., M. A. Fahnestock, and T. A. Scambos.: ITS_LIVE Regional Glacier and Ice Sheet Surface Velocities. Data archived at National Snow and Ice Data Center, 2019

Gdulová, K., Marešová, J., & Moudrý, V.: Accuracy assessment of the global TanDEM-X digital elevation model in a mountain environment. Remote Sens. Environ., 241, https://doi.org/10.1016/j.rse.2020.111724, 2020

Glen, J.W.: The Creep of Polycrystalline Ice. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 228, 519-538, https://doi.org/10.1098/rspa.1955.0066, 1955

González-Moradas, M.d.R., & Viveen, W.: Evaluation of ASTER GDEM2, SRTMv3.0, ALOS AW3D30 and TanDEM-X DEMs for the Peruvian Andes against highly accurate GNSS ground control points and geomorphological-hydrological metrics. Remote Sens. Environ., 237, https://doi.org/10.1016/j.rse.2019.111509, 2020

González, P.J., & Fernández, J.: Error estimation in multitemporal InSAR deformation time series, with application to Lanzarote, Canary Islands. J. Geophys. Res.-Sol. Ea., 116, https://doi.org/10.1029/2011JB008412, 2011

Gorokhovich, Y., & Voustianiouk, A.:. Accuracy assessment of the processed SRTM-based elevation data by CGIAR using field data from USA and Thailand and its relation to the terrain characteristics. Remote Sens. Environ., 104, 409-415, 2006

Grohmann, C.H.: Evaluation of TanDEM-X DEMs on selected Brazilian sites: Comparison with SRTM, ASTER GDEM and ALOS AW3D30. Remote Sens. Environ., 212, 121-133, https://doi.org/10.1016/j.rse.2006.05.012, 2018

Han, H., Zeng, Q., & Jiao, J.: Quality Assessment of TanDEM-X DEMs, SRTM and ASTER GDEM on Selected Chinese Sites. Remote Sen.-Basel, 13, https://doi.org/10.3390/rs13071304, 2021

Hawker, L., Neal, J., & Bates, P.: Accuracy assessment of the TanDEM-X 90 Digital Elevation Model for selected floodplain sites. Remote Sens. Environ., 232, 111724-111738, https://doi.org/10.1016/j.rse.2019.111319, 2019

Höhle, J., & Höhle, M.: Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS J. Photogramm., 64, 398-406, https://doi.org/10.1016/j.isprsjprs.2009.02.003, 2009

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Farinotti, D., Huss, M., Dussaillant, I., Brun, F., & Kääb, A.: Accelerated global glacier mass loss in the early twenty-first century. Nature, 592, 726-731, https://doi.org/10.1038/s41586-021-03436-z, 2021

Huss, M.: Density assumptions for converting geodetic glacier volume change to mass change. The Cryosphere, 7, 877-887, Density assumptions for converting geodetic glacier volume change to mass change, 2013

Huss, M., & Farinotti, D.: Distributed ice thickness and volume of all glaciers around the globe. J. Geophys. Res.-Sol. Ea., 117, 1-10, https://doi.org/10.1029/2012JF002523, 2012

Huss, M., & Hock, R.: Global-scale hydrological response to future glacier mass loss. Nat. Clim. Change., 8, 135-140, https://doi.org/10.1038/s41558-017-0049-x, 2018

Immerzeel, W.W., Lutz, A.F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B.J., Elmore, A.C., Emmer, A., Feng, M., Fernandez, A., Haritashya, U., Kargel, J.S., Koppes, M., Kraaijenbrink, P.D.A., Kulkarni, A.V., Mayewski, P.A., Nepal, S., Pacheco, P., Painter, T.H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A.B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., & Baillie, J.E.M.: Importance and vulnerability of the world's water towers. Nature, 577, 364-369, https://doi.org/10.1038/s41586-019-1822-y, 2020

Kääb, A.: Combination of SRTM3 and repeat ASTER data for deriving alpine glacier flow velocities in the Bhutan Himalaya. Remote Sens. Environ., 94, 463-474, https://doi.org/10.1016/j.rse.2004.11.003, 2005

Kääb, A., Leinss, S., Gilbert, A., Bühler, Y., Gascoin, S., Evans, S.G., Bartelt, P., Berthier, E., Brun, F., Chao, W.-A., Farinotti, D., Gimbert, F., Guo, W., Huggel, C., Kargel, J.S., Leonard, G.J., Tian, L., Treichler, D., & Yao, T.: Massive collapse of two glaciers in western Tibet in 2016 after surge-like instability. Nat. Geosci., 11, 114-120, https://doi.org/10.1038/s41561-017-0039-7, 2018

Kaser, G., Grosshauser, M., & Marzeion, B.: Contribution potential of glaciers to water availability in different climate regimes. Proc. Natl. Acad. Sci. USA, 107, 20223-20227, https://doi.org/10.1073/pnas.1008162107, 2010

Ke, L., Ding, X., Zhang, L.E.I., Hu, J.U.N., Shum, C.K., & Lu, Z.: Compiling a new glacier inventory for southeastern Qinghai–Tibet Plateau from Landsat and PALSAR data. J. Glaciol., 62, 579-592, https://doi.org/10.1017/jog.2016.58, 2016

Kienholz, C., Rich, J.L., Arendt, A.A., & Hock, R.: A new method for deriving glacier centerlines applied to glaciers in Alaska and northwest Canada. The Cryosphere, 8, 503-519, https://doi.org/10.5194/tc-8-503-2014, 2014

Koldtoft, I., Grinsted, A., Vinther, B., & Hvidberg, C.: Ice thickness and volume of the Renland Ice Cap, East Greenland. J. Glaciol., 67(264), 714-726. doi:10.1017/jog.2021.11, 2021

Kraaijenbrink, P.D.A., Bierkens, M.F.P., Lutz, A.F., & Immerzeel, W.W.: Impact of a global temperature rise of 1.5 degrees Celsius on Asia's glaciers. Nature, 549, 257-260, https://doi.org/10.1038/nature23878, 2017

Li, G., & Lin, H.: Recent decadal glacier mass balances over the Western Nyainqentanglha Mountains and the increase in their melting contribution to Nam Co Lake measured by differential bistatic SAR interferometry. Global Planet. Change, 149, 177-190, https://doi.org/10.1016/j.gloplacha.2016.12.018, 2017

Li, S., Yao, T., Yang, W., Yu, W., & Zhu, M.: Glacier Energy and Mass Balance in the Inland Tibetan Plateau: Seasonal and Interannual Variability in Relation to Atmospheric Changes. J. Geophys. Res.-Atmos., 123, 6390-6409, https://doi.org/10.1029/2017JD028120, 2018

Linsbauer, A., Frey, H., Haeberli, W., Machguth, H., Azam, M.F., & Allen, S.: Modelling glacier-bed overdeepenings and possible future lakes for the glaciers in the Himalaya—Karakoram region. Ann. Glaciol., 57, 119-130, https://doi.org/10.3189/2016AoG71A627, 2016

Linsbauer, A., Paul, F., & Haeberli, W.: Modeling glacier thickness distribution and bed topography over entire mountain ranges with GlabTop: Application of a fast and robust approach. J. Geophys. Res.-Sol. Ea., 117, 1-17, https://doi.org/10.1029/2011JF002313, https://doi.org/10.1016/j.geomorph.2019.04.012, 2012

Liu, K., Song, C., Ke, L., Jiang, L., Pan, Y., & Ma, R.: Global open-access DEM performances in Earth's most rugged region High Mountain Asia: A multi-level assessment. Geomorphology, 338, 16-26, https://doi.org/10.1016/j.geomorph.2019.04.012, 2019

Maurer, J.M., Schaefer, J.M., Rupper, S., & Corley, A. Acceleration of ice loss across the Himalayas over the past 40 years. Sci. Adv., 5, eaav7266, https://doi.org/10.1126/sciadv.aav7266, 2019

Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., Gregor, P., Jarosch, A.H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C.T., & Marzeion, B.: The Open Global Glacier Model (OGGM) v1.1. Geosci. Model. Dev., 12, 909-931, https://doi.org/10.5194/gmd-12-909-2019, 2019

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J., & Finkelnburg, R. Precipitation Seasonality and Variability over the Tibetan Plateau as Resolved by the High Asia Reanalysis. J. Climate, 27, 1910-1927, https://doi.org/10.1175/JCLI-D-13-00282.1, 2014

Mölg, N., Bolch, T., Rastner, P., Strozzi, T., & Paul, F.: A consistent glacier inventory for Karakoram and Pamir derived from Landsat data: distribution of debris cover and mapping challenges. Earth Syst. Sci. Data, 10, 1807-1827, https://doi.org/10.5194/essd-10-1807-2018, 2018

Mukherjee, S., Joshi, P.K., Mukherjee, S., Ghosh, A., Garg, R.D., & Mukhopadhyay, A. Evaluation of vertical accuracy of open source Digital Elevation Model (DEM). Int. J. Appl. Earth. Obs., 21, 205-217, https://doi.org/10.1016/j.jag.2012.09.004, 2013

Nuth, C., & 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

Paul, F., & Linsbauer, A.: Modeling of glacier bed topography from glacier outlines, central branch lines, and a DEM. Int. J. Geogr. Inf. Sci., 26, 1173-1190, https://doi.org/10.1080/13658816.2011.627859, 2012

Pelto, B.M., Maussion, F., Menounos, B., Radić, V., & Zeuner, M.: Bias-corrected estimates of glacier thickness in the Columbia River Basin, Canada. J. Glaciol., 66, 1051-1063, https://doi.org/10.1017/jog.2020.75, 2020

Ramsankaran, R., Pandit, A., & Azam, M.: Spatially distributed ice-thickness modelling for Chhota Shigri Glacier in western Himalayas, India. Int. J. Remote Sens., 39, 3320-3343, https://doi.org/10.1080/01431161.2018.1441563, 2018

Rankl, M., & Braun, M.: Glacier elevation and mass changes over the central Karakoram region estimated from TanDEM-X and SRTM/X-SAR digital elevation models. Ann. Glaciol., 57, 273-281, https://doi.org/10.3189/2016AoG71A024, 2016

Reuter, H.I., Nelson, A., & Jarvis, A.: An evaluation of void-filling interpolation methods for SRTM data. Int. J. Geogr. Inf. Sci., 21, 983-1008, https://doi.org/10.1080/13658810601169899, 2007

Scambos, T., M. Fahnestock, T. Moon, A. Gardner, and M. Klinger.: Global Land Ice Velocity Extraction from Landsat 8 (GoLIVE), Version 1. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center, 2016

Shean, D.E.: High Mountain Asia 8-meter DEM Mosaics Derived from Optical Imagery, Version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, 2017

Shean, D.E., Bhushan, S., Montesano, P., Rounce, D.R., Arendt, A., & Osmanoglu, B.: A Systematic, Regional Assessment of High Mountain Asia Glacier Mass Balance. Front. Earth. Sci., 7, https://doi.org/10.3389/feart.2019.00363, 2020

Shortridge, A., & Messina, J.: Spatial structure and landscape associations of SRTM error. Remote Sens. Environ., 115, 1576-1587, https://doi.org/10.1016/j.rse.2011.02.017, 2011

Smith, B., Fricker, H.A., Holschuh, N., Gardner, A.S., Adusumilli, S., Brunt, K.M., Csatho, B., Harbeck, K., Huth, A., Neumann, T., Nilsson, J., & Siegfried, M.R.: Land ice height-retrieval algorithm for NASA's ICESat-2 photon-counting laser altimeter. Remote Sens. Environ., 233, https://doi.org/10.1016/j.rse.2019.111352, 2019

Takaku, J., Tadono, T., Doutsu, M., Ohgushi, F., & Kai, H.: Updates of 'Aw3d30' Alos Global Digital Surface Model with Other Open Access Datasets. Int. Arch. Photogramm., XLIII-B4-2020, 183-189, https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-183-2020, 2020

Thompson, L.G., Yao, T., Davis, M.E., Mosley-Thompson, E., Wu, G., Porter, S.E., Xu, B., Lin, P.-N., Wang, N., Beaudon, E., Duan, K., Sierra-Hernández, M.R., & Kenny, D.V.: Ice core records of climate variability on the Third Pole with emphasis on the Guliya ice cap, western Kunlun Mountains. Quaternary. Sci. Rev., 188, 1-14, 2018

Uuemaa, E., Ahi, S., Montibeller, B., Muru, M., & Kmoch, A.: Vertical Accuracy of Freely Available Global Digital Elevation Models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). Remote Sens.-Basel, 12, https://doi.org/10.1016/j.quascirev.2018.03.003, https://doi.org/10.1016/j.rse.2007.11.003, 2020

Van Niel, T.G., McVicar, T.R., Li, L., Gallant, J.C., & Yang, Q.: The impact of misregistration on SRTM and DEM image differences. Remote Sens. Environ., 112, 2430-2442, https://doi.org/10.1016/j.rse.2007.11.003, 2008

Welty, E., Zemp, M., Navarro, F., Huss, M., Fürst, J.J., Gärtner-Roer, I., Landmann, J., Machguth, H., Naegeli, K., Andreassen, L.M., Farinotti, D., & Li, H.: Worldwide version-controlled database of glacier thickness observations. Earth System Science Data, 12, 3039-3055, https://doi.org/10.5194/essd-12-3039-2020, 2020

Wessel, B., Huber, M., Wohlfart, C., Marschalk, U., Kosmann, D., & Roth, A.: Accuracy assessment of the global TanDEM-X Digital Elevation Model with GPS data. ISPRS J. Photogramm., 139, 171-182, https://doi.org/10.1016/j.isprsjprs.2018.02.017, 2018

Wu, K., Liu, S., Zhu, Y., Liu, Q., & Jiang, Z.: Dynamics of glacier surface velocity and ice thickness for maritime glaciers in the southeastern Tibetan Plateau. J. Hydrol., 590, https://doi.org/10.1016/j.jhydrol.2020.125527, 2020

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S., & Bates, P.D.: A high-accuracy map of global terrain elevations. Geophys. Res. Lett., 44, 5844-5853, https://doi.org/10.1002/2017GL072874, 2017

Yao, T., Thompson, L.G., Mosbrugger, V., Zhang, F., Ma, Y., Luo, T., Xu, B., Yang, X., Joswiak, D.R., Wang, W., Joswiak, M.E., Devkota, L.P., Tayal, S., Jilani, R., & Fayziev, R.: Third Pole Environment (TPE). Environ. Dev., 3, 52-64, 2012

Zhang, G., Bolch, T., Allen, S., Linsbauer, A., Chen, W., & Wang, W.: Glacial lake evolution and glacier–lake interactions in the Poiqu River basin, central Himalaya, 1964–2017. J. Glaciol., 65, 347-365, https://doi.org/10.1016/j.envdev.2012.04.002, 2019

Zhang, G., Yao, T., Xie, H., Kang, S., & Lei, Y.: Increased mass over the Tibetan Plateau: From lakes or glaciers? Geophys. Res. Lett., 40, 2125-2130, https://doi.org/10.1002/grl.50462, 2013

Zhang, Y., Pang, Y., Cui, D., Ma, Y., & Chen, L.: Accuracy Assessment of the ICESat-2/ATL06 Product in the Qilian Mountains Based on CORS and UAV Data. IEEE J-STARS, 14,1558-1571, https://doi.org/10.1109/JSTARS.2020.3044463, 2020

Zheng, G., Allen, S.K., Bao A. M., Juan A. B., Matthias H., Zhang G. Q., Li J. L., Ye Y., Jiang L. L., Yu T., Chen W.F., & Stoffel, M.: Increasing risk of glacial lake outburst floods from future Third Pole deglaciation. Nat. Clim. Change., 11, 411–417, https://doi.org/10.1038/s41558-021-01028-3, 2021

Zhou, Y., Li, Z., Li, J., Zhao, R., & Ding, X.: Glacier mass balance in the Qinghai–Tibet Plateau and its surroundings from the mid-1970s to 2000 based on Hexagon KH-9 and SRTM DEMs. Remote Sens. Environ., 210, 96-112, https://doi.org/10.1016/j.rse.2018.03.020, 2018