Comments to the author:
Dear Dr. Wenfeng Chen,

Both reviewers have expressed concerns on central aspects of your manuscript submitted to The Cryosphere that demand a major revision of the work to be reconsidered for publication. Hence, if you decide to revise your work, it will be subjected to another round of reviewing.

Best wishes,

Arjen Stroeven

Response: Many thanks to the Editor Arjen Stroeven for the professional process of our manuscripts. We have carefully addressed all the comments from Review #1, Review #2 and Community. Our point-by-point responses are attached below in blue, while the original reviewers' comments are in black.

Review #1

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 consistent acquisition dates, NASADEM is the best choice for ice-thickness estimates over the whole 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 high vertical accuracy and resolution (Brunt et al., 2019, 2021; Li et al., 2021), 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 in Tibet plateau and one more study in the *Antarctica.*
"The ATL06 product has better than 5 cm height accuracy and better than 20 cm surface measurement precision in the Antarctic (Brunt et al., 2019,2020; Li et al., 2021) and Qilian Shan (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.
Li, R., Li, H., Hao, T., Qiao, G., Cui, H., He, Y., Hai, G., Xie, H., Cheng, Y., and Li, B.: Assessment of ICESat-2 ice surface elevations over the Chinese Antarctic Research Expedition (CHINARE) route, East Antarctica, based on coordinated multi-sensor observations, The Cryosphere, 15, 3083–3099, https://doi.org/10.5194/tc-15-3083-2021, 2021.

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 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):"

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 not only filter the differences between ICESat-2 and DEMs to exclude extreme outliers, but also keep most records in the further accuracy analysis. The ration of excluded outliers relative to the 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. Why are the RMSE values here different than in Fig. 4? How come the RMSE values here are nearly identical to the STD values in Fig. 4?

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. In fact, the RMSE in Figure 3 should be equal to the STD in Figure 4. The intercept should be equal to the ME 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 think 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. If make a composite bed estimate with only the best four ITIM-DEM combinations, it's unclear to analysis the influence from DEM, because both DEM and ITIMs are variable.

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 are 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"

"Six DEMs are produced from different sensors or by different methods. The Ppixel of different DEMs at the 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 hypothetical 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: The authors 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)."

Review #2

Comments on

"Towards ice thickness inversion: an evaluation of global DEMs by ICESat-2 in the glacierized Tibetan Plateau" by Wenfeng Chen et al.

Referee #2

Overview

This paper by Chen et al. presents a method to evaluate existing regional scale DEMs using the recently available ICESat-2 elevation product. The DEMs are then applied to model the ice thickness of glaciers in the Tibetan Plateau region. The quality of the inversion results is then analyzed to prove the effectiveness of the ICESat-2 based evaluation. The paper is generally well structured. It showed that the ICESat-2 data provided a comparison dataset for selecting an optimal DEM that can be used as input for ice thickness inversion. This work should be useful to researchers of The Cryosphere community. The paper needs to be revised according to the following major and minor comments.

The manuscript needs a serious improvement of both formal English writing and scientific meaning.

Response: We are grateful to the anonymous reviewer for the constructive comments on our manuscript. We have carefully addressed all the comments below. The English are checked by a native English speaker. We have added some content to improve the scientific meaning. Our point-by-point responses are attached below in blue, while the original reviewers' comments are in black.

"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 exist in DEMs, such as speckle noise, stripe noise and absolute bias; they behave differently across the Earth (Yamazaki et al., 2017; Takaku et al., 2020). But our method to assess the accuracy of DEMs is repeatable in different regions, combined with the recently released glacier elevation change data on Earth (Hugonnet et al., 2021). What's more, benefiting from the high accuracy and dense coverage of ICESat-2 data, the quality of DEMs can also be improved as similar as the production of MERIT (Yamazaki et al., 2017). For example, the misregistration in DEMs could be corrected and terrain-related errors could be reduced by unitizing the relation of difference against slope, aspect and elevation in Fig. 6."

Major comments

1. The subtitles in the Data and Methods sections need to be improved to reflect the actual contents and be logic (ICESat-2 elevation data referenced, DEMs evaluated, ……)

Response: The subtitles are revised and organized to reflect the actual contents.

2.1 Descriptions of ICESat-2 elevation data referenced;

2.2 Descriptions of global-scale DEMs evaluated;

2.3 Ice thickness inversion method;

2.4 Accuracy assessment method;

2. The ICESat-2 data were used to evaluate DEMs. It should be discussed how the DEMs were generated, including data sources, time periods, and uncertainties. During the time differences between the ICESat-2 data and MEDs there may be glacier surface changes that may affect the evaluation. If this is not considered, would this also affect the thickness inversion results?

Response: We add more details about how the six DEMs are generated. Detailed information is also summarized in Table1.

Yes, glacier surface change affects the evaluation. We considered it in the analysis in Section 3.3 and discussed it in section 4.1. We also added one paragraph in Section 2.4 to describe how we solve the influence from glacier surface change on evaluation.

"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."

Glacier surface elevation change could also influence the inversion of ice thickness, especially when estimating glacier thickness in regional scale. Therefore, though AW3D30 with mixing acquiring dates exhibit a good accuracy assessment, we still suggest NASADEM is a best choice for ice-thickness estimates over the whole TP.

3. ICESat-2 Level-3A land-ice ATL06 product was used in this work. There should be a good understanding of the quality of this product itself, although a systematic calval may not have been performed in the glacierized Tibetan Plateau region. I would like to see even a general discussion in that regard. I suggest to add some relevant references in the Data section:

Brunt, K. M., Neumann, T. A., and 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.

Brunt, K. M., Smith, B. E., Sutterley, T. C., Kurtz, N. T., and Neumann, T. A.: Comparisons of Satellite and Airborne Altimetry With Ground-Based Data From the Interior of the Antarctic Ice Sheet, Geophys Res Lett, 48, e2020GL090572 https://doi.org/10.1029/2020GL090572, 2021.

Li, R., Li, H., Hao, T., Qiao, G., Cui, H., He, Y., Hai, G., Xie, H., Cheng, Y., and Li, B.: Assessment of ICESat-2 ice surface elevations over the Chinese Antarctic Research Expedition (CHINARE) route, East Antarctica, based on coordinated multi-sensor observations, The Cryosphere, 15, 3083–3099, https://doi.org/10.5194/tc-15-3083-2021, 2021.

Response: We have added the above reference.

"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 20 cm surface measurement precision in the Antarctic (Brunt et al. 2019,2020; Li et al. 2021) and Qilian Shan (Zhang et al. 2020)."

Elevation differences between crossovers formed by ICESat-2 tracks may be used for the elevation accuracy evaluation.

Response: We thanks for the reviewer's suggestion. This is really a good idea for testing the performance of ICESat-2 in stable region. However, when we check the dates of point around crossover of ICESat-2 tracks, we find they are from different dates, so maybe it's not easy to know whether this elevation difference is from glacier elevation change or the uncertainty of ICESat-2.

As ATL06 product comes from ATL03 photons grouped in 40 m segments, would ATL03 data provide more terrain details? This may be outside of the scope of this paper. However, a discussion of this potential would be helpful to the readers in their future research.

Response: We added these sentences in Section 2.1.

"ICESat-2 ATL03 and ATL06 product both can be used as elevation reference. ATL03 product has a spacing of ~0.7m and can provide more terrain details than ATL06 product. In this study, considering the resolution of global dem and compute cost, we select the ICESat-2 Level-3A land-ice ATL06 product as elevation reference."

4. Based on your results, the accuracy of ICESat-2 data is better than the compared global-scale DEMs. Have you considered to use the ICESat-2 data to improve the quality of the DEMs, especially in areas where ICESat-2 along and across track data are available. Again, this may be considered as a future work

Response: Yes, that's a good point. We added this in the discussion.

"What's more, benefiting from the high accuracy and dense coverage of ICESat-2 data, quality of DEMs can also be improved as similar as the production of MERIT (Yamazaki et al. 2017). For examples, the misregistration in DEMs could be corrected and terrain-related errors could be reduced by unitizing the relation of difference against slope, aspect and elevation in Fig. 6."

Minor comments

5. Page 1, Line 11: Replace "derived" with "derived from".

Response: Corrected

6. Page 1, Lines 12–14: This sentence is awkwardly phrased. You may rephrase by separating it in to two sentences.

Response: Corrected

"However, the scarce in-situ measurements of glacier surface elevation limit the evaluation of DEM uncertainty. Hence the influence of DEM uncertainty on ice-thickness modelling remains unclear over the glacierized area of the Tibetan Plateau (TP)."

7. Page 1, Line 17: Please be clear if it is the horizontal or vertical accuracy.

Response: it's vertical accuracy. Corrected

8. Page 1, Lines 23-24: Change "one pixel" to one grid spacing.

Response: Corrected.

9.Page 1, Lines 24-28: These sentences are not communicating well. Please cut them to short and simple sentences.

Response: Corrected

Then, influence of six DEMs on four ice-thickness models: GlabTop2, Open Global Glacier Model (OGGM), Huss-Farinotti (HF), Ice Thickness Inversion Based on Velocity (ITIBOV) is intercompared. The results 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.

10. Page 2, Lines 33 & 36: Change "km2 ", "km3 " to "km2", "km3".

Response: Corrected

11. Page 3, Line 87 and Page 4, Lines 107-108: When you mention ICESat-2 in these two places, please mention the cal-val efforts and introduce the most recent accuracy assessment results by Brunt et al. (2019, 2021) and Li et al. (2021).

Response: Corrected.

"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 (Brunt et al. 2019, 2021; Li et al. 2021)"

"The ATL06 product has better than 5 cm height accuracy and better than 20 cm surface measurement precision in the Antarctic (Brunt et al. 2019,2020; Li et al. 2021) and Qilian Shan (Zhang et al. 2020)."

12. Page 4, Line 95: Replace "intersecting" with "covering".

Response: Corrected

13. Page 4, Line 105: ~17 m diameter is the design value. This value needs to be updated according to the new study: Magruder, L. A., Brunt, K. M., and Alonzo, M.: Early ICESat-2 on-orbit geolocation validation using ground-based corner cube retroreflectors, Remote Sensing, 12, 3653, https://doi.org/10.3390/rs12213653, 2020.

Response: Corrected.

"ICESat-2 ATLAS (Advanced Topographic Laser Altimeter System) emits a pulse every 0.7 m along the track covering a horizontal circular area with 0.5 m in vertical extent and ~17 m diameter. This design diameter value varied due to the photo-counting lidar technology and potentially the atmospheric conditions (Magruder et al. 2020)."

14. Page 4, Line 106: In addition to first-photon bias, transmit-pulse shape correction should be mentioned also

Response: Corrected.

"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 and transmit-pulse shape."

15. The ATL06 data provide slopes in along and across tracks. They are derived at different scales (along track with denser points and cross track with fewer points and longer separations). Have you considered the difference between these two types of slopes?

Response: Here, the slope is derived from DEMs not the ICESat-2 product. We use this slope for error analysis at slope scale and misregistration analysis. In the origin design of this research, we plan to estimate the slope accuracy derived from DEMs based on the ICESat-2 along-track slope. But we found that the algorithms they calculated slope are totally different (Burrough and McDonell, 1998; Smith et al., 2019)., so we didn't estimate this furtherly.

Burrough, P.A., & McDonell, R.A. (1998). Principles of Geographical Information Systems. Oxford University Press, New York, 190 pp.

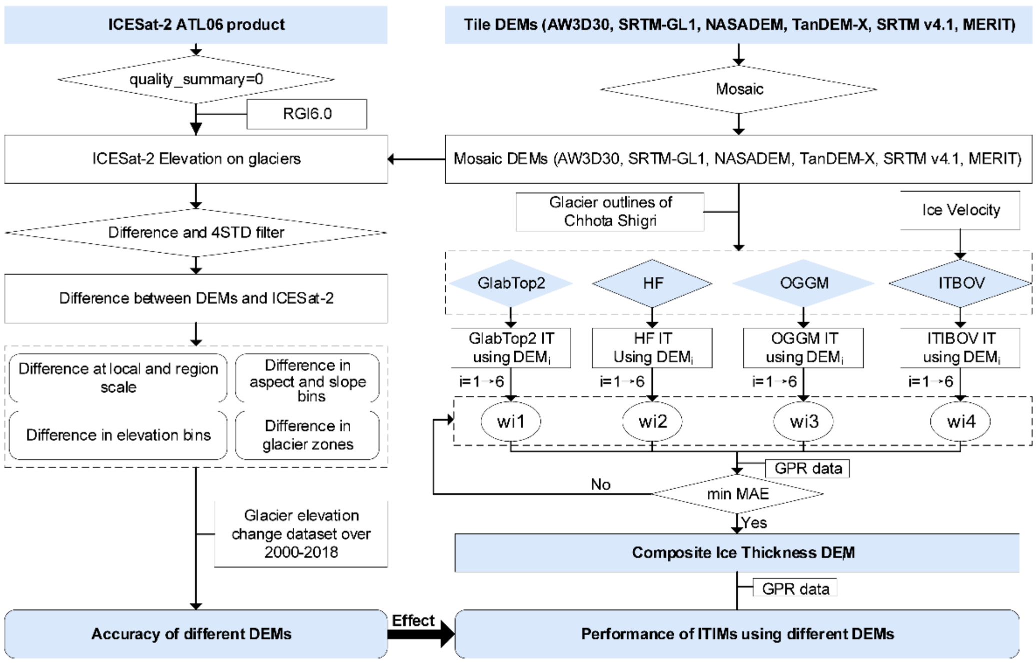
Smith, B., Fricker, H.A., Holschuh, N., Gardner, A.S., Adusumilli, S., Brunt, K.M., Csatho, B., Harbeck, K., Huth, A., Neumann, T., Nilsson, J., & Siegfried, M.R. (2019). Land ice height-retrieval algorithm for NASA's ICESat-2 photon-counting laser altimeter. Remote Sensing of Environment, 233

16. Figure 2: "Difference in aspect and slope bins".

Response: Corrected.

17. Figure 2: Make different blue boxes into just blue color.

Response: Corrected.



18. Page 6, Line 127: Please specify how severe is the gap situation and the effectiveness of gap filling

Response: We have added these sentences.

"Approximate 10 % of global land area, mainly in tropical rainforest areas and the polar areas, has voids mostly due to cloud or snow/ice covers constatation in source imageries. Data gaps are filled with SRTM, ASTER GDEM v3, ArcticDEM v3, and TanDEM-X 90 (Takaku et al., 2020). After filling gaps, the accuracies in void-filled and void-free areas are nearly consistent (Takaku et al., 2020)."

19. Page 6, Line 135: All website links need to add the last visit date to ensure their availability

Response: We have added visit date of website links in Table 1.

20. Page 8, Lines 160-162: The sentence is unclear. Please rewrite.

Response: corrected.

"Four ice-thickness inversion models (GlabTop2, HF, OGGM, ITBOV) were used to estimate the glacier thickness. The Chhota Shigri Glacier located in western Himalaya with available GPR data (Fig.1) was selected as an example to evaluate the influence of DEM uncertainty on the ITIMs."

21. Page 10, Line 209: Please justify your choice of 4-std of differences between ICESat-2 and DEMs for data filtering.

Response: We add these sentences.

"The four standard deviations (that is 4 std) was chosen to not only filter the differences between ICESat-2 and DEMs to exclude extreme outliers, but also keep most records in the further accuracy analysis. Ration of excluded outliers relative to record of each DEM is less than 1%."

22. Page 10, Lines 212 & 214: "R2"

Response: Corrected.

23. Page 11, Line 219: "R2"

Response: Corrected.

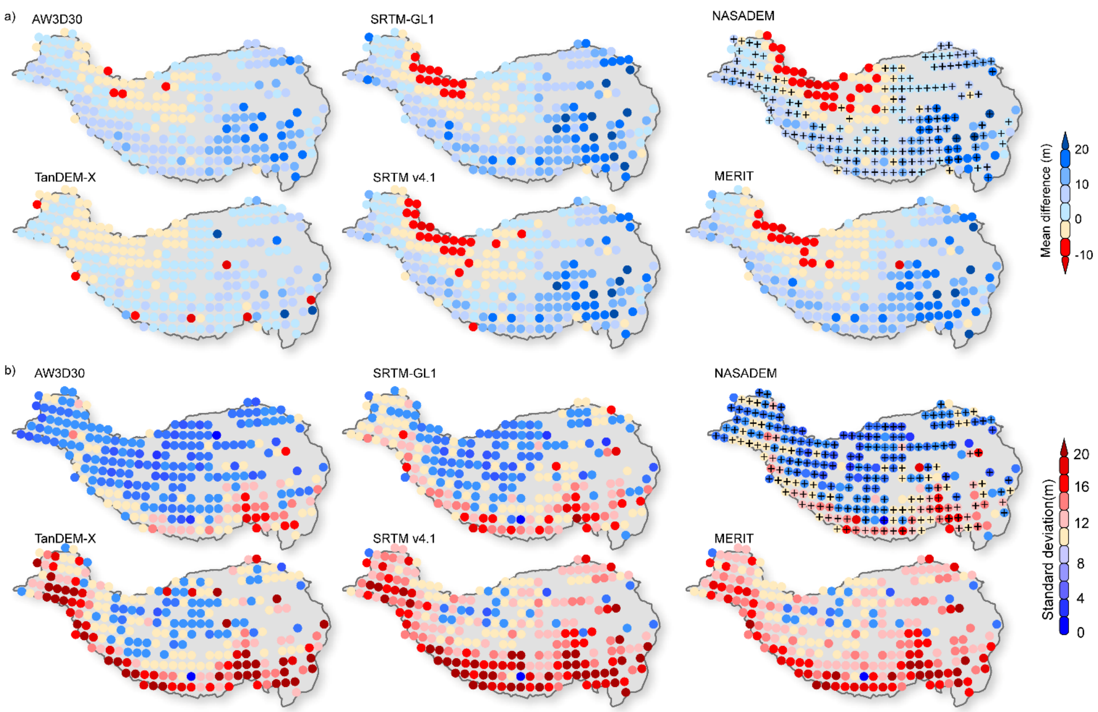
24. Figure 3 & Figure 4: Why would RMSE in both figures be different? For example, for AW3D30 it is 13.4 m in Fig 3, but it is 34.9 m in Fig 4. Please make sure the data are consistent throughout the paper.

Response: The RMSE in Figure 3 and Figure 4 are different. 'R2', 'RMSE' and 'Intercept' are fit results when the slope coefficient is set to 1. In fact, the RMSE in Figure 3 should be equal to the STD in Figure 4. The intercept should be equal to the ME in Figure 4.

25. "The ME in the Himalaya is more negative than that in southeast Tibet", but it is not clearly shown in Figure 5. Do you want to quantify it with numbers?

Response: The data in Figure 5 is updated. A small difference range is used to make the difference of DEMs more obviously. And the sentence is rewritten.

"The ME in southeast Tibet is more positive than that in the Himalaya"



26. Page 12, Line 243: Please be clear about "spatially relevant"

Response: Corrected.

" Specifically, the STD of AW3D30 and NASADEM was minimum and has similar spatial distribution."

27.Page 15, Line 278: The time of six DEMs is all earlier than the ICESat-2 data. As stated in the major comments, in addition to uncertainties would temporal changes of the glaciers also affect your evaluation efforts?

Response: Yes, the glacier elevation change could also influence the evaluation. We quantitatively estimate the influence from glacier elevation change in Section 3.3 and 4.1. Especially for glaciers in the ablation zone, the ME, MAE SD and RMSE are largely reduced for SRTM based DEMs (Table 3).

28. Figure 6: No significant information is presented in Fig 6(d). You may just deleted it.

Response: We used this Fig.6d here.

"For NASADEM and SRTM-GL1, the differences along the elevations show similar distribution and varied from –10 to 10 m over the range 4500−6500 m, where measurements are concentrated (Fig. 6d);"

"The differences with aspect show contrasting features to the distribution of measurements in different aspects (Fig. 6d)."

"Though points in 55°−90° slope region account for small fraction (Fig. 6d), almost half the points in the 55°−90° slope region are identified as extreme outliers (Fig. 10a)."

29. Page 16, Line 317: Explain the numbers of "9, 3, 4, 3 and 1".

Response: We refined this sentence and tittle of Table 2.

"Totals of 8, 7, 3 and 2 output achieved the minimum RMSE in profiles (bold number in Table 2) by different ITIMs using AW3D30, NASADEM, TanDEM-X and SRTM-GL1, respectively."

"Table 2 RMSE (m) of modelled ice thickness compared with ground penetrating radar (GPR) measurements on each profile. Location of profiles are shown in Figure 1. Bold numbers denote the best model performance on each profile using different DEMs."

30. The figure displays the proportion of outliers vs. slopes. Within the range of 55-90 degrees, the proportion is up to ~40-55% for six DEMs. This is a high percentage. Would you please discuss the reasons?

Response: We discussed in the section 4.2. We thought that the steep slope and intra-pixel effect should be attributed to this.

"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)."

31. Page 21, Line 384: "ICESat-2 elevation is … higher in Fig. 11d". This statement is not true in Fig.11d. Please describe the figure accurately and objectively

Response: Since elevation data is all converted to same reference frame, the difference between DEM and ICESat2 become small. The profiles overlap seriously, so we deleted this Figure 11. Numbers and figures are referenced to make our description more accurate and objective.

32. Page 22, Line 393: Remove "will".

Response: Corrected**.**

33. Page 22-23, Lines 401-402: Remove the sentence "Additionally, … elevation". The error of these two scatterings cannot influence the magnitude of the biased estimate of elevation you mentioned here.

Response: Corrected**.**

34. Page 26, Line 449: Explain how is this correction of sub-pixel misregistration performed.

Response: Corrected**.**

"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 in x- and y- direction at the 1°×1° grid scale were estimated by fitting method in MATLAB across the TP."

35. Page 27, Line 470: Change "A +5° error" to "An error of +5°".

Response: Corrected.

36. Page 28, Lines 495-497: I cannot get what you mean here. Rewrite it to make it clear.

Response: What I mean is that mean slope in each elevation bin is used not the pixel-based slope. It's similar as the sentence of “For the HF model, elevation data was used for convergence calculation of apparent mass balance and mean slope in elevation bins” in this paragraph. So, Lines 495-497 is deleted.

37. Page 29, Lines 527-529: It is not clear what you want to say here. It is in the conclusions section. You need to be brief and direct. Please rewrite.

Response: Corrected. Some sentences are deleted to make the conclusion section brief.

“The widely used GlabTop2 model is very sensitive to the accuracy of both elevation and slope; using NASADEM as the input, this model facilitated the best outcome.”

Community

This is a brief comment based on a quick skim of the discussion paper. It looks like a nice contribution, but I believe there is a fundamental issue with the methodology involving global DEM preparation.

I noticed what looks like a systamatic vertical datum offset in some of the results (e.g., Figure 12), with offsets of ~30 m between ICESat-2 elevation (height above ellipsoid) and a subset of the global DEM elevations (height above respective geoid model - see documentation for specific geoid used by each DEM). It's important to account for this offset before doing any analysis. The apparent 30 m bias is due to the geoid offset, and is not representative of the quality of the DEM. This bias will propagate to RMSE numbers, which will impact the conclusions.

It should be relatively straightforward to correct each of the input DEMs to provide ellipsoidal heights using available offset grids (https://cdn.proj.org/), and repeat the analysis. I can't remember if the RGI-TOPO dataset (https://rgitools.readthedocs.io/en/latest/dems.html) has already accounted for vertical datum differences between different sources. One can also use the OpenTopography GlobalDEM API (https://portal.opentopography.org/apidocs/#/Public/getGlobalDem), which provides versions of some global DEMs with ellipsoidal heights rather than the orthometric heights.

We thank Dr. David Shean for pointing out this and the information provided. These were very valuable. The offset of ~30 m between ICESat-2 elevation and the global DEM elevations is indeed due to the differences of datum reference. We ignored the reference differences between various DEMs and ICESat-2. In fact, 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). We have added this information to Table 1. We updated the data and the geoidheight function provided by MATLAB was used to calculate geoid height to unify their references. All the figures and Tables are updated thoroughly. our main conclusion doesn’t change, we still concluded that NASADEM performed best and would be a best choice for ice-thickness estimates over the TP. The details about the revisions could be found attached.