Reviewers Comments:

The manuscript describes the correction of ASTER GDEM data with ICESat-2 data to create a product with a lower RMSE (according to a validation dataset). The ASTER GDEM is a widely used dataset and improving the accuracy is a useful exercise. The journal is the right venue to publish an open-source dataset.

The manuscript is mostly written in correct English and grammar – but please refrain from using etc. The abstract alone contains three occasions using etc. and several others throughout the manuscript. Either it is important enough to spell out – then list the additional points. If it is not important, there is no etc. needed. The term adds unnecessary ambiguity.

I can mostly follow the manuscript and reasoning, but have some comments. I understand that this article has seen previous reviews. I suggest that some of these are added as caveats or critical thoughts.

Response:

Thanks for your affirmation about our research and paper. We have checked the term "etc." in the whole paper and revised the related descriptions. According to your suggestions, we have revised the manuscript item by item and given a response to each comment.

1) **Dataset description. A more detailed description of the ASTER GDEM is necessary. What is the time frame of acquisition? Is it reasonable to use an ICESat2 dataset to correct the data (ICESat2 likely postdates some of the scenes used in the generation of ASTER GDEM). Same with the validation dataset: The ICESat data likely predates the ASTER GDEM scenes. While it is not likely that the large number of validation points have changed and I don' t think there is an impact on the statistics – but it will be useful to give these relevant information and a word of caution. The years of the lidar DEMs are listed.**

Response:

Thank you for your suggestion. We have added the information of the relevant data.

In the Section 4.1,

Table 2 Characteristics of input and output data

Data Type	***Data name***	***Resolution***	***Description***
DEM data	*ASTER GDEM V3*	*~30 m (grid)*	*Input data, main source data, survey date: 2000 - 2013*
Satellite laser altimeter data	*ICESat-2 ATL08*	*~100 m (along the ground track)*	*Input data, main source data, survey date: 2018 - 2022*
Landcover data	*FROM-GLC10*	*~10 m (grid)*	*Input data, auxiliary data, survey date: 2017*
Vegetation cover index data	*GFCC30TC*	*~30 m (grid)*	*Input data, auxiliary data, survey date: 2015*
DEM data	*IC2-GDEM*	*~30 m (grid)*	*Output data*

has been added.

In the Section 4.2,

"*By using the HAGECPD data, we validated and then contrasted the accuracy of the original and corrected ASTER GDEM elevation…*"

has been revised to

"*By using the HAGECPD data (survey date: 2003-2009), we validated and then contrasted the accuracy of the original and corrected ASTER GDEM elevation…*"

2) **While the training and validation dataset are somewhat independent (training: ICESat2, validation: ICESat about 10-15 years earlier), they both exhibit the same data characteristic. I have no concerns about the data quality, but it is not an unbiased validation. The geographic location points are not the same, but both data are point measurements (albeit taken with different instruments).**

Response:

Thank you for your affirmation about our data quality. Our validation is not an absolute unbiased validation and it is difficult to achieve an absolute unbiased validation using high-accuracy data at a global scale. To reduce this impact, we adopted a comprehensive strategy, i.e., in addition to ICESat laser altimeter data, we also used LDEM to analyze ASTER GDEM before and after elevation correction. To describe clearer, we have revised the relevant descriptions.

In the 1st paragraph of Section 2.4,

After *"Two different kinds of validation data were used in this study, according to the type of survey platform (i.e., satellite and airborne platforms)."*

"We adopted this strategy for validating ASTER GDEM before and after elevation correction more comprehensively, in order to reduce the impact of biased validation."

has been added.

3) **I may have missed it, but how are the border effects of the individually-adjusted tiles treated? Each 1-degree tile is calibrated (or trained) individually and the adjustment parameters may be different than the neighboring parameters. This may cause (or not) a small offset at the boundaries of the tile. Initially, I thought there is a feathering approach used with a buffer (equations 1 to 5), but I am not certain that this point is clearly illustrated or explained. This is section 3.4.**

Response:

Thank you for your suggestions. Yes, we used a buffer strategy to reduce the impact of the boundary, as shown in equations 1 to 5. The strategy we adopted is to expand the boundaries of the central DEM (the processing DEM) by using data from neighboring DEMs. The strategy described in Section 3.4 is to expand the area for selecting training samples in order to ensure that there are sufficient samples when the number of training samples is too low. Perhaps our description was not clear enough. To describe clearer, we have revised the relevant descriptions.

The 1st paragraph before equations 1,

"To address these challenges, it was necessary to expand the area around the DEM file to be corrected. As shown in part A of Figure 3, the expanded area includes two types…"

has been revised to

"To address these challenges, it was necessary to expand the area around the DEM file to be corrected. We adopted a strategy that is to expand the boundaries of the central DEM (the processing DEM) by using data from neighboring DEMs. As shown in part A of Figure 3, the expanded area includes two types…"

In the Section 3.4,

"*Moreover, the training model had a minimum requirement for the seed points' number within the modeling area.*"

has been revised to

"*Moreover, when the number of training samples is too low, it needs to expand the area (i.e., the modeling area) for selecting training samples in order to ensure that there are sufficient training samples.*"

4) **Where is the list of attributes that are trained with RF? Is there an attribute importance list that describe the usefulness of these parameters. I see the description of the "GDEM Elevation Evaluation Attribute Set" in 3.2 and that is useful. It is not clear what is contained in the elevation correction model (e.g., is this using Nuth and Kaeaeb to make horizontal adjustment or is this just a z component?)**

Response:

Thank you for your suggestion.

In our previous study, we gave the details of the evaluation attributes in the regression model. The evaluation attribute set in the model includes topography, surface coverage, spatial distribution, and data-source quality. Meanwhile, we also evaluated the performance of each attribute. Our model just focuses on correcting DEM elevation, i.e., z component. Maybe we described the model as not clear. We have added some descriptions.

In the last paragraph of Section 3.2,

"The evaluation attribute set in the model includes topography, surface coverage, spatial distribution, and data-source quality (Li et al., 2023a)."

has been added.

In the last paragraph of Section 3.4,

"*After obtaining the model, all the DEM elevations were corrected by the model, and then a new DEM file was generated with the same format.*"

has been revised to

"*After obtaining the model, all the DEM elevations (i.e., z component) were corrected by the model, and then a new DEM file was generated with the same format.*"

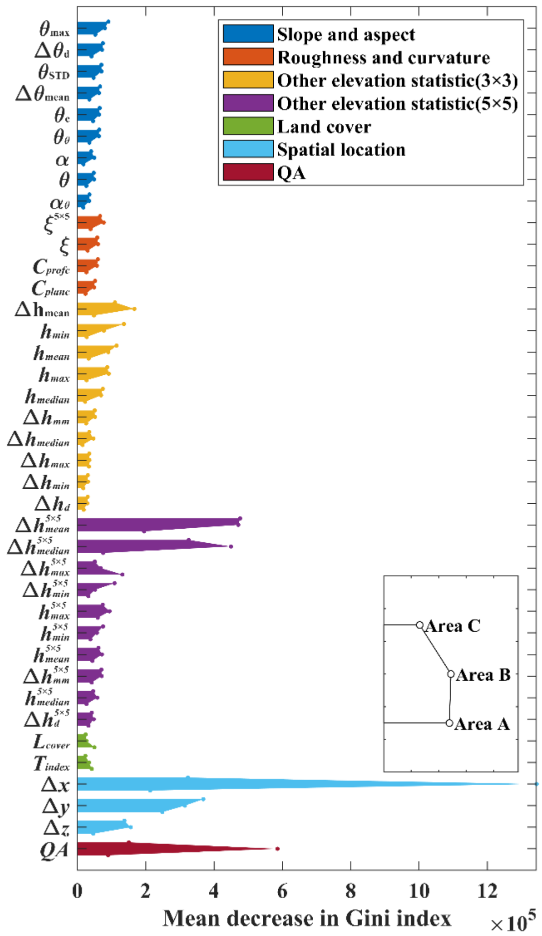
In our previous study:

Li, B., Xie, H., Tong, X., Tang, H., and Liu, S.: A Global-Scale DEM Elevation Correction Model Using ICESat-2 Laser Altimetry Data, IEEE Trans. Geosci. Remote Sens., 61, 1-15.

Table Evaluation attributes in the regression model

NO.	**Factor**	**Symbol**	**Description**
1	*Topography (Slope and aspect features)*	α	*Aspect*
2		θ	*Slope*
3		θ_c	*CV of slope*
4		θ_θ	*Slope of slope*
5		α_θ	*Aspect of slope*
6		θ_{max}	*Maximum slope*
7		θ_{STD}	*STD of slope*
8		$\Delta\theta_{mean}$	$\theta - \theta_{mean}$
9		$\Delta\theta_d$	$\theta_{mean} - \theta_{min}$
10	*Topography (Roughness and curvature)*	ξ	*Roughness is calculated by the DEM heights within a 3×3 window*
11		$\xi^{5\times5}$	*Roughness is calculated by the DEM heights within a 5×5 window*
12		C_{profc}	*Profile curvature*
13		C_{planc}	*Planar curvature*
14-15	*Topography (Other statistical elevation)*	h_{mean} $(h_{mean}^{5\times5})$	*Mean of the DEM heights within a 3×3 (5×5) window*
16-17		h_{medain} $(h_{medain}^{5\times5})$	*Median of the DEM heights within a 3×3 (5×5) window*
18-19		h_{min} $(h_{min}^{5\times5})$	*Min of the DEM heights within a 3×3 (5×5) window*
20-21		h_{max} $(h_{max}^{5\times5})$	*Max of the DEM heights within a 3×3 (5×5) window*
22-23		Δh_{mean} $(\Delta h_{mean}^{5\times5})$	$h - h_{mean}$ *within a 3×3 (5×5) window*
24-25		Δh_{medain} $(\Delta h_{median}^{5\times5})$	$h - h_{median}$ *within a 3×3 (5×5) window*
26-27		Δh_{min} $(\Delta h_{min}^{5\times5})$	$h - h_{min}$ *within a 3×3 (5×5) window*
28-29		Δh_{max} $(\Delta h_{max}^{5\times5})$	$h - h_{max}$ *within a 3×3 (5×5) window*
30-31		Δh_{mm} $(\Delta h_{mm}^{5\times5})$	$h_{max} - h_{min}$ *within a 3×3 (5×5) window*
32-33		Δh_d $(\Delta h_d^{5\times5})$	$h_{mean} - h_{min}$ *within a 3×3 (5×5) window*
34	*Earth surface*	L_{cover}	*Land cover*
35		T_{index}	*Vegetation cover index*

36	*Spatial Distribution and Data-source Quality*	Δx	*X-direction geographic location (relative value)*
37		Δy	*Y-direction geographic location (relative value)*
38		Δz	*Elevation (relative value)*
39		QA	*Data-source quality*



Performance analysis of features evaluated in correction GDEM.

5) **Along the same lines: Individually training each tile is useful and will allow to correct for local problems. The current description of the random forest training is a black-box approach – I did not see the parameters (or attributes) that are used for height adjustment (or I have missed them). Is this just the dH? In any case, it will be useful to include the adjustment parameter or vertical offset in a separate dataset. This will allow the user to see how much each tile has been adjusted. Most other global DEM datasets have additional quality data (e.g., Copernicus has the number of measurement or TanDEM-X pairs). The averaged adjusted dH value for each tile is a useful assessment metric.**

Response:

Thank you for your suggestion. It's a beneficial suggestion. We have calculated the averaged adjusted dH value for each tile. We also upload the calculation information to the National Tibetan Plateau Data Center. Please see the Q7, figure 5 shows the distribution of the averaged adjusted dH value for each tile at a global scale.

6) **I am wondering about the improved product. I can imagine that a post-processed ASTER GDEM has an reduced RMSE. But the core problem with optical data is the inherent noise (see also Purinton and Bookhagen, 2021). Even in vegetation-free areas, the correlation problems with optical data lead to lower quality DEMs. Is the noise level reduced by the random-forest filtering? A zoom in area of a characteristic area with sufficient detail would be useful (before and after correction). I point out that several studies now have shown that the Copernicus DEM is currently the best available DEM. It is difficult to compare optical and radar-based DEMs, because they measure different things. Different story again with a Lidar DEM.**

Response:

Thank you for your suggestion.

Maybe we described the model as not clear. We analyzed the comparison of the original and corrected ASTER GDEM elevation error in vegetation-free areas, as shown in Figure 10 of the original paper. About the Copernicus DEM, your view is correct. Compared with the other global DEM data, the Copernicus DEM has the best quality in general. We have revised and added the relevant descriptions.

In the Introduction,

"Owing to the noise and anomalies resulting from the limitations inherent in optical imaging, the elevation quality of the ASTER GDEM is typically deemed to be lower than that of the other GDEMs for which the source is radar data (Meadows et al., 2024; Del Rosario González-Moradas et al., 2023)"

has been revised to

"Owing to the noise and anomalies resulting from the limitations inherent in optical imaging, the elevation quality of the ASTER GDEM is typically deemed to be lower than that of the other GDEMs for which the source is radar data (Meadows et al., 2024; Del Rosario González-Moradas et al., 2023; Purinton and Bookhagen, 2021)"

"For instance, the data sources for the void filling of the Copernicus GLO-30 DEM consist of other DEMs…"

has been revised to

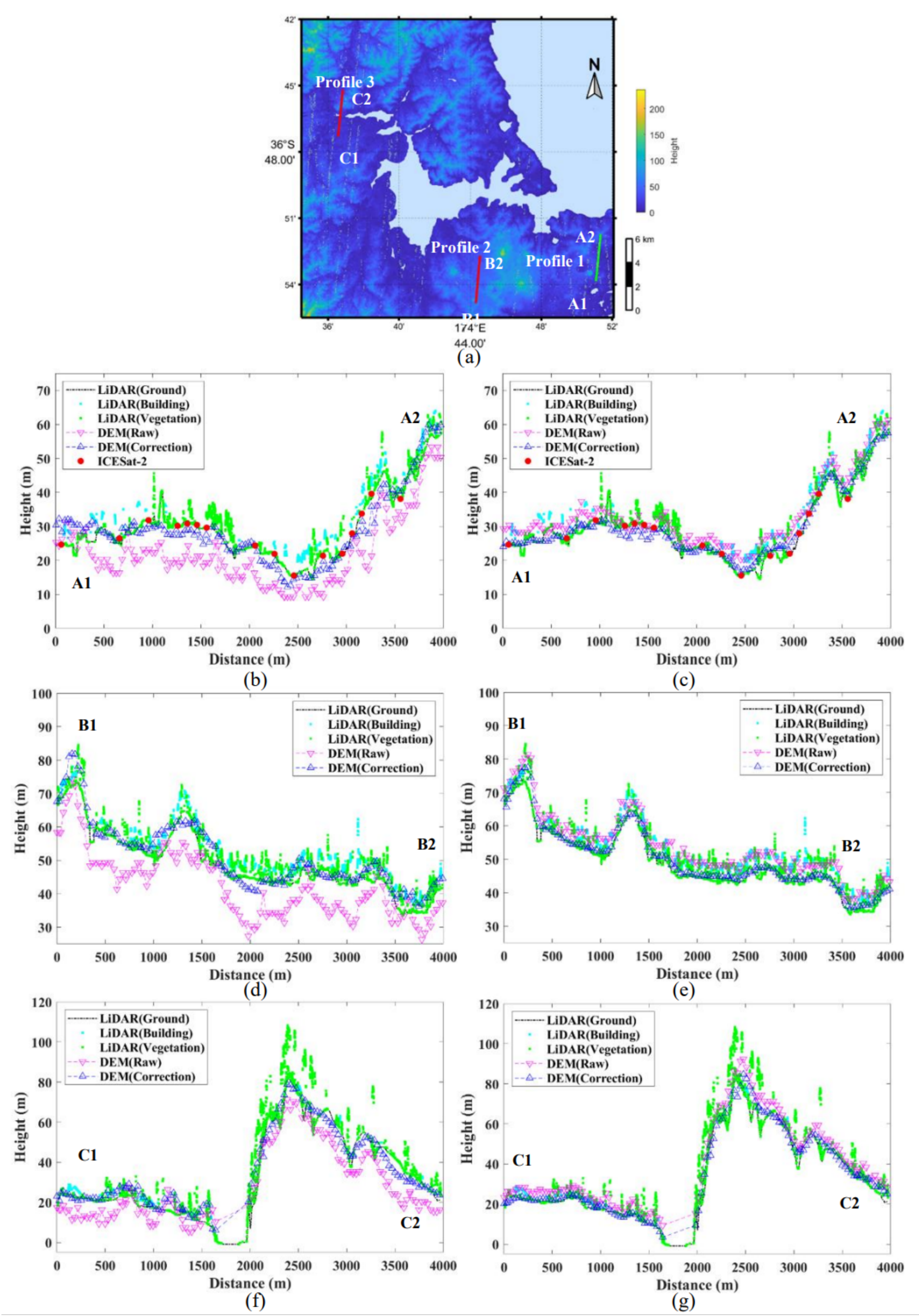
"For instance, the data sources for the void filling of the Copernicus GLO-30 DEM (currently the best available global DEM) consist of other DEMs…"

In reference,

"Purinton, B. and Bookhagen, B.: Beyond Vertical Point Accuracy: Assessing Inter-pixel Consistency in 30 m Global DEMs for the Arid Central Andes, Front. Earth Sci., 9, 2021." has been added.

In the 4th paragraph of the Section 4.3, After "*For the different land covers, the corrected GDEM elevation in the bare land areas shows the best elevation correction.*", "*This means that the DEM elevation correction model can reduce the impact caused by the inherent noise of optical data.*" has been added

In our previous study, we focused on the performance of the elevation correction model and presented some examples of the detail about quality improvement of ASTER GDEM elevation before and after, as the following figure.



Samples of DEM correction results.
(a) shows the geolocation of two profiles and the geolocation of filtered elevation points of ICESat-2 (light green scatter); (b)(c) represent the DEM elevation

correction results of profile 1 (it overlaps with the filtered elevation points of ICESat-2); (d)(e)(f)(g) represent the DEM elevation correction results of profile 2 and 3 (they do not overlap the filtered elevation points of ICESat-2); (b)(d)(f) and (c)(e)(g) represent the GDEM and SRTM elevation correction results, respectively.

7) I am not certain if Figure 4 is useful – at this scale you are not seeing any difference. WHAT would be useful is to show the elevation adjustments that have been done (in a divergent color scale).

Response:

Thank you for your suggestion. We have added a figure to show the elevation correction of IC2-GDEM relative to ASTER GDEM.

In the Section 4.1,

"*Figure 5 shows the average of elevation corrections of IC2-GDEM relative to ASTER GDEM within the global 1° ×1° grid.*"

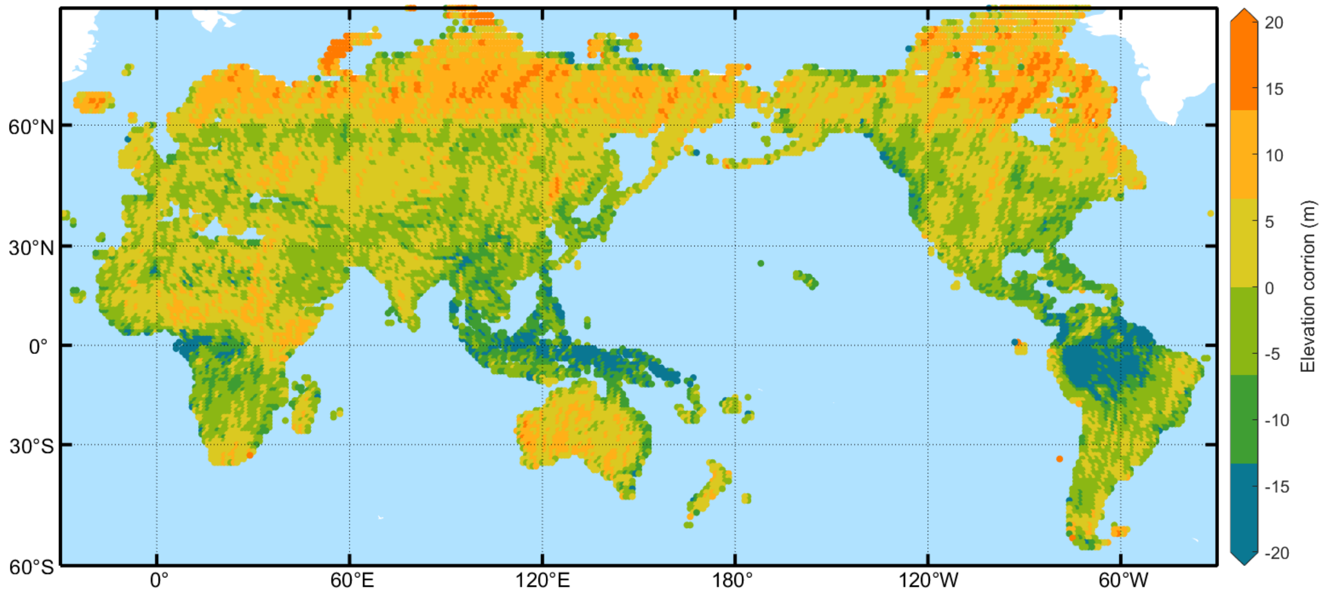


Figure 5: Average of elevation corrections of IC2-GDEM relative to ASTER GDEM within the global 1° ×1° grid.

has been added.