We thank all three reviewers for their constructive comments and suggestions. This manuscript will be much improved by their input. We have made changes to our manuscript. In the following responses, we use "**bold**" text for reviewer's comments, "non-bold" text for our responses, and "*italic*" for changed text in the manuscript.

Referee #1

General Reviewer Comments:

This paper presents the results of a validation effort for ICESat-2 ice surface elevations in East Antarctica. The campaign was coordinated with the CHINARE Antarctic expedition from December 2019 to February 2020. The authors focused on an assessment of both the ATL03 (geolocated photons) and ATL06 (land ice elevations) along-track ICESat-2 Level 3a data products. The goal of this effort was to determine the ICESat-2 product vertical accuracy and precision using four different approaches.

The four resources for ICESat-2 validation are:

- **520 km GNSS traverse from the coastal Zhongshan Station to the inland Taishan Station**
- **Two line arrays of corner cube retro-reflectors**
- **Two retroreflective target sheets.**
- **Imagery from two UAV platforms**

The authors designed this campaign as a complementary approach to those efforts performed previously by the ICESat-2 mission team and confirm the regional accuracy to support uncertainty determination within mass balance and change rate calculations for East Antarctica. The logistics accomplished in this coordinated approach for validation are commendable and I congratulate the authors on such an effort.

My primary concern with this work is primarily with two of the validation techniques that use ground based retroreflectors. These results are very limited as there are only 7 data points evaluated for the target sheet and 1 data point evaluated for a CCR. Clearly, the traverse comparison provides a robust performance assessment for vertical accuracy based on a reasonable sample size for statistical aggregation. The same does not seem not true for the CCR and RTS approach.

We agree that the GNSS-based traverse comparison provided a robust assessment of the ICESat-2 surface elevations along the CHINARE route, which is complementary to the mission team results along the latitude of 88°S (Brunt et al., 2020b), considering its number of observations, surface types over the 520 km traverse, and achieved accuracy. Furthermore, the long GNSS traverse, areal UAV tests, and CCR and RTS retroreflectors are designed as a systematic and integrated ICESat-2 assessment approach, although they had different observation opportunities individually during the experiments. In addition to their contributions to the overall assessment approach, the CCR and RTS techniques presented in this paper demonstrate a way of using targeted and controlled photons to assess the ICESat-2 *elevations*, in complement to the result in Magruder et al. (2020) in which the *horizontal accuracy* and footprint size were estimated successfully by using CCRs. Particularly, the results of these less-costly CCR and RTS devices used in this paper were validated using the high precision GNSS technique and showed a potential for future applications in more sites with different Antarctic conditions and mass balance rates.

Based on your comments and suggestions, we revised the paper to point out a) the need for aggregation of (observation) opportunities in the future, and b) to lower the significance of the CCR and RTS results in comparison to the GNSS – traverse results.

We added the following in **Abstract**.

"*…… It should be emphasized that the results based on the CCR and RTS techniques can be improved by further aggregation of observation opportunities for a more robust assessment......*"

We added the following in the **Discussions** section.

"*…… It should be emphasized that constrained by expedition logistics the observation opportunities of one CCR overpass and seven RTS check points in this study may not be considered as a large sample size. The assessment result may vary with locations, environmental conditions, and times. Thus, there is a need for aggregated opportunities of CCR and RTS observations to achieve a validation result with variant influence factors accounted for (e.g., ATLAS attitude, solar angle, and atmosphere).*"

We added the following in the **Conclusions** section.

"*… … In addition to the different ice surface types covered from coast to inland Antarctica along the 520 km CHINARE route, the result of the CCR elevation assessment complements that of the CCR horizontal accuracy and footprint size assessment in Magruder et al. (2020). Furthermore, the RTS and UAV – DEM assessment results are, to our knowledge, reported first time for the validation of these early stage ICESat-2 data in the AIS environment. Although the UAV - DEM coverage is relatively small and the CCR and RTS observation opportunities are relatively limited in comparison to the large number of GNSS observations along the 520 km GNSS traverse, their performances and achieved results in this study pave a way for future applications with aggregated observations at more sites for a more robust assessment. Therefore, our ICESat-2 validation methodology and sensor system will be applied to carry out the continued assessment of the ICESat-2 data, especially for calibration against potential degradation of the elevation measurements during the later operation period.*"

- **The authors should include the appropriate data product citations (via NSIDC) in addition to the references to the ATL03 and ATl06 Algorithm Theoretical Basis Documents by Neumann and Smith.**

We added the data product citation of "NSIDC 2021". It is also stated in "*Data availability*".

- **The Neumann et al., 2020b should be replaced with:**
  - **Magruder, L. Brunt, K., and Alonzo, M. (2020) Early ICESat-2 on-orbit geolocation validation using ground-based corner cube retro-reflectors. Remote Sensing, 12(21), 3653; https://doi.org/10.3390/rs12213653**

We have replaced the citation with the suggested one “Magruder et al., 2020”. It is also updated in references.

- **‘two sets’ of CCRs would be better represented as two line arrays. ‘Sets’ is not applicable to individual CCRs nor does it describe the implementation effectively.**

We have replaced “*sets*” with “*line arrays*”.

- **Is an RTS a ‘set’ as well?**

Each RTS consists of a set of RTS pieces with two different kinds of coatings. We would like to see if we can keep “sets” for RTSs in this context.

- **L65: CCRs do not ‘capture’ photons, they reflect them so that the unique signature they provide to ATLAS can be analyzed relative to their known height and position.**

We replaced “*capture*” with “*reflect*”.

- **Why do the authors use 6 cm diameter CCRs for ICESat-2 when clearly for 532 nm that size optic would suffer from velocity aberration? Was this considered in the selection? Does that make a difference in the analysis if the returns are coming from the lobes of the diffraction pattern rather than the central disk?**

We add the text in the **Data** section to explain why we use 6 cm diameter CCRs: “*Under the time constraints of equipment shipment before expedition and manufacturing cycle of new products we used ten readily available CCRs of 6 cm diameter (Fig. A1a) at each site, which were designed for ground - based laser distance measurement. They were placed linearly at a 10 m interval across a nominal ICESat-2 ground track to reflect photons from ATLAS (Fig. 2a)……*”

The central disc of our 6 cm aperture CCRs has a diameter of ~10.82 m in comparison to ~40.57 m of the ~8 mm diameter CCRs in Magruder et al. (2020), according to Chang et al. (1971). Here is the Fraunhofer diffraction pattern of our CCR:

$$D = \frac{1.22 \times \lambda \times H}{d} \times 2.$$

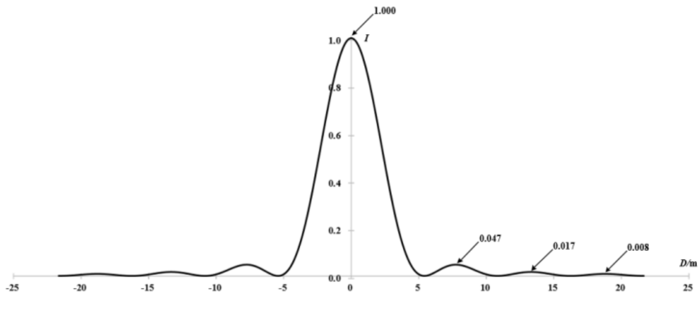

Figure of the Fraunhofer diffraction pattern of the 6 cm diameter CCR (λ=532 nm, H=500 km, and d=6 cm)

With the larger CCR aperture, in addition to the smaller central disc diameter, the total signal level (for both central disc and lobes) is also increased. Thus, it makes signals from the lobes of the Fraunhofer diffraction pattern also detected. In the following figure, at an approaching position ($t_{\mathrm{Approaching}}$) ATLAS received and accepted signals from lobes of both the nadir CCR #6 (red signal curve) and the neighboring CCR #7 (light blue signal curve), both at a lower signal level; this resulted in the reflected photons of higher elevations (green dots) of CCR #6 and those of lower elevations (green dots) of CCR #7. However, at the nadir CCR position (t_{Nadir}) ATLAS received and accepted high level signals from the central disc of the nadir CCR #6 (higher elevation), but may have rejected the lower-level signals from the neighboring CCR #7 (lower elevation) because of the much increased ratio between the signals. This allows us to determine the window size, ~9 m gap of the lower streak, to select photons of the nadir CCR (including those inside the central disc) for CCR elevation estimation.

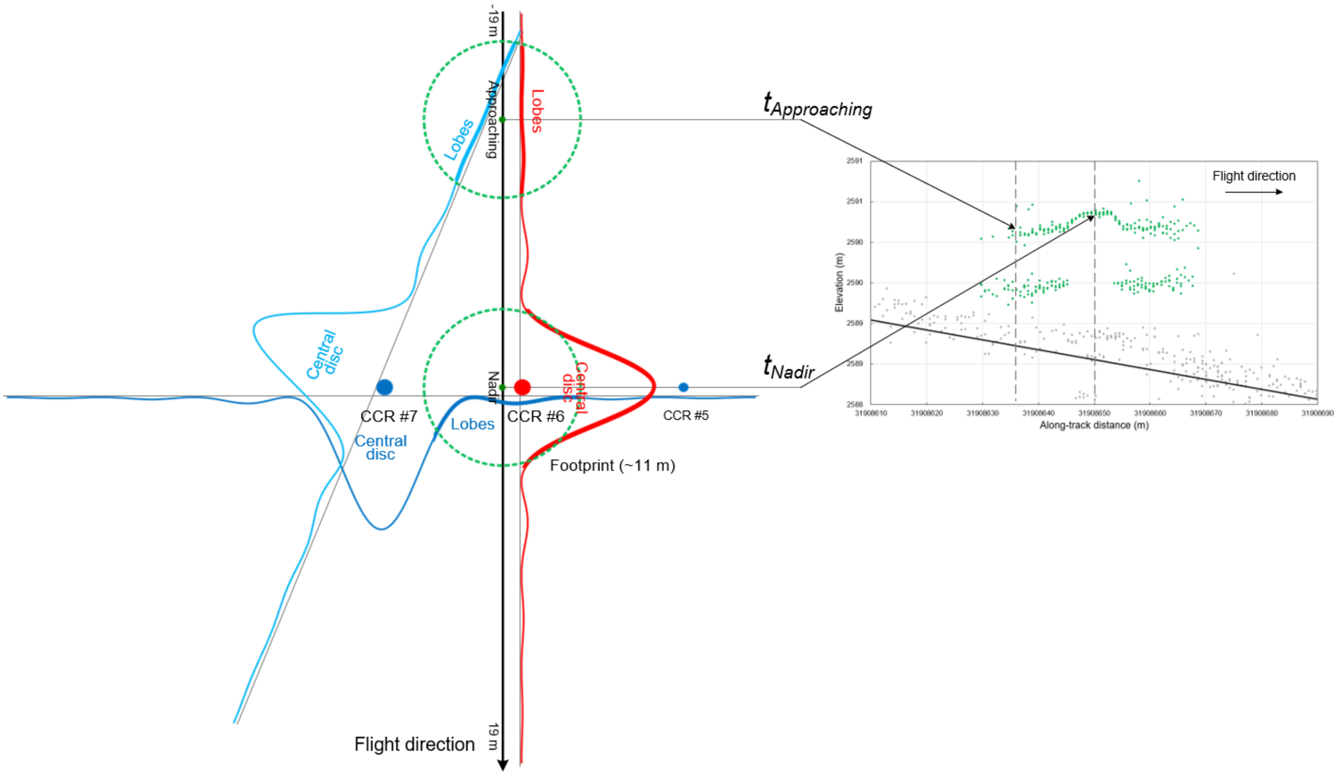

Figure for CCR signal analysis at the site near Taishan Station

We added a paragraph in the **Discussions** section to explain the impact:

"*The use of the readily available CCRs of 6 cm diameter for the 532 nm wave length of ATLAS, which is larger than 8 mm of the CCRs used in Magruder et al. (2020), is subject to velocity aberration caused by a decreased central disc and receiving signals from the outer lobes of the Fraunhofer diffraction pattern (Chang et al., 1971; Born et al., 1999; Sun et al., 2019; Magruder et al., 2020). In addition, the larger aperture of the CCR resulted in a higher level of the total signals received by ATLAS so that signals from both the smaller central disc and outer lobes are detected and used to estimate elevations in ATL03 data. This may have attributed to the creation of the long along-track streaks of ~35 m (Fig. 6a) and ~38 m (Fig. 6c) in comparison to those of ~11 m in Magruder et al. (2020). Thus, photons reflected from the lower neighboring CCR(s) in the cross-track direction (Fig. 6c) were detected for the same reason because of the symmetric Fraunhofer diffraction pattern. Similarly, the one-layer photon streak (green dots in Fig. 6a) may include those reflected from one or both neighboring CCRs because the elevations of all three neighboring CCRs (#4, #5 and #6) are within a 15* *cm range (Table C1) due to local ice surface topography and logistic constraints, although the poles were manufactured in different lengths. On the other hand, a temporal distribution of energy within a pulse is approximately Gaussian (Smith et al., 2019); the received signals in the central disc are generally of a higher level (about 84% of the total energy) than in the outer lobes given atmospheric scattering and other optical losses (Magruder et al., 2020). Correspondingly, we observe that within the window of the nadir CCR (red rectangle in Fig. 6c) the photons are densely aligned along a curve. The same curve trend appears to continue towards both ends, but diverged by potentially blended signals reflected from neighboring CCRs (Figs. 6a and 6c). Therefore, by selecting photons inside the central window of the CCR streak it ensures that high quality photons in the central disc of the Fraunhofer diffraction pattern be used to estimate the elevation of the representative photon of the nadir CCR through the fitting curve. The result is also validated by the nadir CCR position surveyed by using the high-precision GNSS RTK technique.*"

Fig. 6 is improved to reflect the changes in the text:

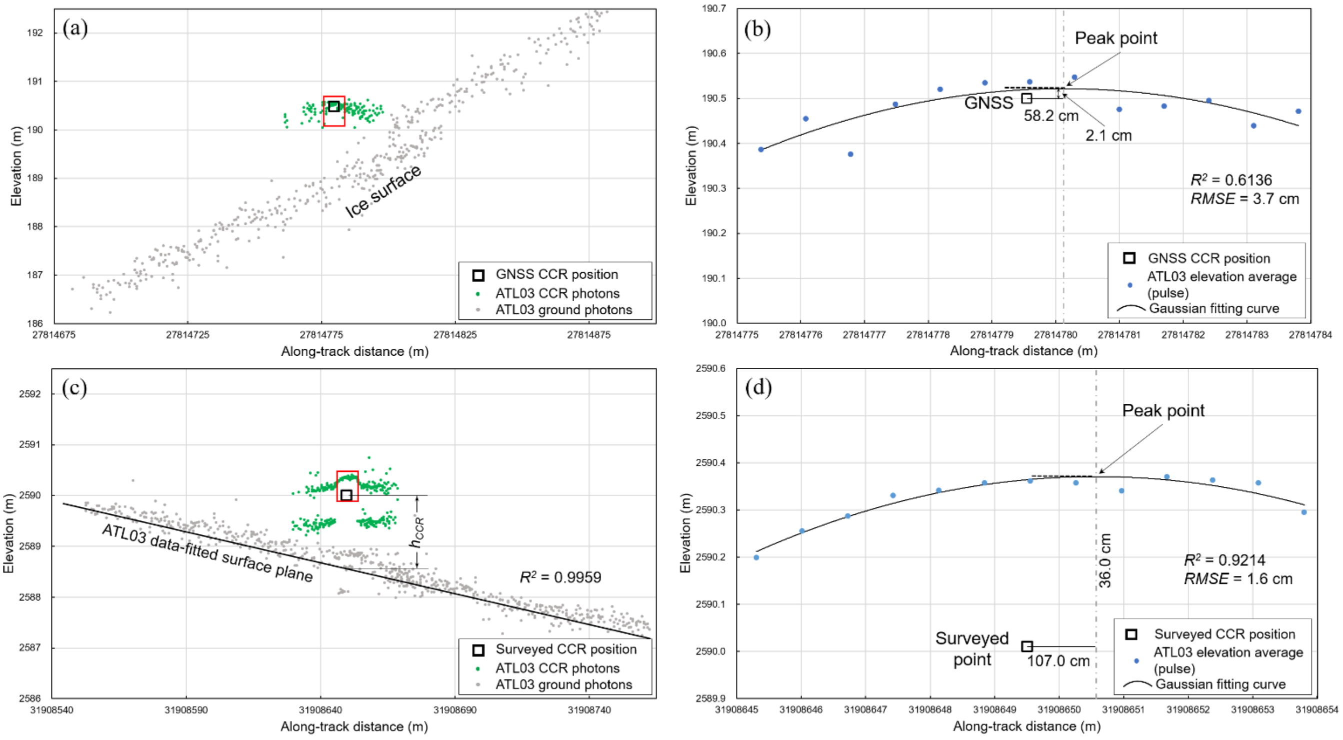

Figure 6. (a) CCR experiment near Zhongshan Station: returned photons (ATL03), GNSS-surveyed CCR position, and ice surface photons (ATL03); (b) elevations averaged in each pulse in the red rectangle in (a) to compare with the GNSS-surveyed position; (c) CCR experiment near Taishan Station: returned photons (ATL03), steel tape-surveyed CCR position, and h_{CCR} - height between CCR center and ice surface (ATL03); and (d) elevations averaged in each pulse in the red rectangle in (c) to compare with the steel tape-surveyed position.

We revised and added the text in the **Results** and **Method** sections to explain how we select the appropriate photons to analyze the data:

In the **Results** section "*……Among the CCRs, CCR #5 near Zhongshan Station and CCR #6 near Taishan Station were closest (~0.3 m and ~1.3 m) to the ground tracks and are called nadir CCRs. Given these CCR shifts from ground tracks, uncertainty of the ground tracks themselves of up to ~6.5 m (Magruder et al., 2020), and the CCR interval of ~10 m, we expected to have one to two CCRs falling into a footprint of ~11 m. Fig.*

6c clearly shows returned photons (green dots) of two layers with an elevation difference of ~50 cm. Furthermore, the photons in the upper layer within a window of ~9 m (red rectangle in Fig. 6c) around the nadir CCR (#6) were received by ATLAS, while those of the lower elevation were not present inside the window. Therefore, we only use the returned photons inside a window of ~9 m around the nadir CCR to validate ICESat-2 elevations".

In the **Method** section "……*where the photons are generally of high quality and have the confidence tag "signal_conf_ph" equal to 3 or 4 to calculate an average elevation in each pulse to reduce noises and potential atmospheric effect (blue dots in Fig. 6b). The average elevations are then further fitted to a symmetric curve, i.e., a Gaussian curve. The central point (or peak) of the curve is treated as the representative CCR photon position. An offset is then calculated between the positions estimated from the ATL03 photons and the GNSS survey*."

- **What are the elevation differences between the 10 CCRs or what is the combination of elevations in the line array?**

We added the text in the **Data** section: "*The elevations of the nominal centers of the CCR lens were designed to vary within one meter for identification of individual CCRs from which the signal was reflected. The poles were manufactured before the expedition with different lengths. The actual elevations were influenced by the ice surface topography where they were deployed (Table C1)*." See Table C1 for measured elevations of all CCRs near Zhongshan Station and all CCR heights from the ice surface near Taishan Station.

- **3.1.3 The ICESat-2 laser pairs are separated by 3.3 km in the across-track direction.**

Changed it from 3 km to 3.3 km as suggested. It is changed in Fig. 5a also.

- **L195 The authors report a laser diameter value of 11 m in the introduction and then move to estimating a ~13 m. Where does this new value come from?**

"~13 m" was an early value from website https://icesat-2.gsfc.nasa.gov/science/specs. We changed it now to "~11 m" based on Magruder et al. (2020). We also added this new citation in text.

- **Since the weak beam detectors are saturated at ~4 photons how does this contribute to your radiometry study. The ~10 photons for the strong beam is not saturated (which would be 16).**

Counting the photons nearby CCR and RTS was done within an ice surface segment of 250 m along track. This was performed in two segments on both sides of RTS near Zhongshan and Taishan stations. We found that the selected segments were outside RTS but included CCR. Thus, it resulted in ~4 photons for weak beams and ~10 photons for strong beams. Now we moved the counting segments 300 m away from RTS and CCR. Here are the new counts: "…… *ATL03 data have an average number of photons (confidence flag: low-high) of ~2 for weak beams and ~7 for strong beams per pulse on the ice surface in our study area;* ……" They are not saturated for both weak and strong beams. So, the radiometric characterization of the ice surface is not affected by "saturated" beams. This is also reflected in the radiometric analysis result presented in Tables 3 and 4.

- **Figure 6 (c) Should this height be dh$_{CCR}$? That was described as the distance between the top of the pole and the middle of the CCR, presumably not at the meter scale that is shown here?**

(Fig. 6c): Sorry, it should be h$_{CCR}$ (instead of dh$_{CCR}$) that is the height between CCR center and ice surface (ATL03). It is changed in the figure. So, it is of meter scale.

Also, in this figure why is there a bulge in the CCR signature rather than a straight line as shown in Figure 6 (a)? This seems as if the footprint illuminated multiple CCRs at different heights? This doesn't seem to be normal streak characteristics.

We added the text in the **Discussions** section to explain it:

"*…… On the other hand, the received signals in the central disc are generally of a higher level (about 84% of the total energy) than in the outer lobes given atmospheric scattering and other optical losses (Magruder et al., 2020). Correspondingly, we observe that within the window of the nadir CCR (red rectangle in Fig. 6c) the photons are densely aligned along a curve. The same curve trend appears to continue towards both ends, but diverged by potentially blended signals reflected from neighboring CCRs (also see Fig. 6a). Therefore, by selecting photons inside the central window of the CCR streak it ensures that high quality photons in the central disc of the Fraunhofer diffraction pattern be used to estimate the elevation of the representative photon of the nadir CCR through the fitting curve. The result is also validated by the nadir CCR position surveyed by using the high-precision GNSS RTK technique.*"

- **Section4.2: The ATL03 CCR streak is 40 pulses across 35 m, how is that with 0.7 cm per pulse. It should be closer to 28 m in length.**

Sorry, the empty pulses in the ATL03 CCR streak were not counted.

Now we added pulse id range for the streak in Fig. 6a: "*…… The elevations of the photons within each pulse, pulse id (/gtx/heights/ph_id_pulse) from 133 to 145, were averaged and used to fit a Gaussian function curve ……*"

We also added pulse id range for the streak in Fig. 6c: "*…… After that, the photons were averaged within each of pulse (pulse id from 19 to 31, blue dots in Fig. 6d) to fit another Gaussian function ……*"

- **How does the selection of the CCR returns impact the results, meaning if you included more of the streak's length does it negatively impact the comparison or change the quantitative values you are achieving?**

As suggested, we performed an analysis of ΔZ between the ICESat-2 CCR elevations and GNSS surveyed CCR elevations achieved using different streak lengths, including 17 m (a large window size), 13 m (ICESat-2 tech spec: https://icesat-2.gsfc.nasa.gov/science/specs), 11 m (footprint size in Magruder et al., 2020), and 9 m (this study).

Site \ ΔZ (W-size)	ΔZ±σ for 17 m (cm)	ΔZ±σ for 13 m (cm)	ΔZ±σ for 11 m (cm)	ΔZ±σ for 9 m (cm)
Zhongshan Station	2.0±7.7	2.4±5.1	1.4±3.8	2.1±3.7
Taishan Station	32.3±4.4	34.9±3.1	35.8±2.5	36.0±1.6

The CCR position near Taishan Station has an accuracy of ~1 m and the result may not be very meaningful. On the other hand, the CCR position near Zhongshan Station was surveyed by GNSS at a centimeter level accuracy. The elevation differences vary within 1 cm. Considering both elevation difference and fitting error, the smaller window sizes (9 m and 11 m) have better results. The errors with the larger windows may be attributed to the lower quality photons from the outer lobes and possibly reflected from multiple CCRs. We suggest that the window size be limited to between 9 m to 11 m (within one footprint).

- **Do you average the elevations per pulse or delta time? There is not a pulse id parameter.**

Yes, the elevations are averaged per pulse.

We added pulse ids for the streak in Fig. 6a: "*…… The elevations of the photons within each pulse, pulse id (/gtx/heights/ph_id_pulse) from 133 to 145, were averaged and used to fit a Gaussian function curve ……*"

We also added pulse ids for the streak in Fig. 6c: "*…… After that, the photons were averaged within each of pulse (pulse id from 19 to 31, blue dots in Fig. 6d) to fit another Gaussian function ……*"

- **The authors should emphasize in the paper that the CCR and RTS techniques are localized assessments with results that vary with satellite orientation and time of year (e.g. solar angle). This is particularly important when there are so few opportunities to evaluate the performance (e.g. 1 CCR overpass and 7 RTS points). These techniques cannot yet be compared to those associated with the traverse without further aggregation of opportunities for analysis.**

We accepted the suggestion and added texts in three places.

We added the following in **Abstract**.

"*…… It should be emphasized that the results based on the CCR and RTS techniques can be improved by further aggregation of observation opportunities for a more robust assessment......*"

We added the following in the **Discussions** section.

"*…… It should be emphasized that constrained by expedition logistics the observation opportunities of one CCR overpass and seven RTS check points in this study may not be considered as a large sample size. The assessment result may vary with locations, environmental conditions, and times. Thus, there is a need for aggregated opportunities of CCR and RTS observations to achieve a validation result with variant influence factors accounted for (e.g., ATLAS attitude, solar angle, and atmosphere).*"

We added the following in the **Conclusions** section.

"*… … In addition to the different ice surface types covered from coast to inland Antarctica along the 520 km CHINARE route, the result of the CCR elevation assessment complements that of the CCR horizontal accuracy* *and footprint size assessment in Magruder et al. (2020). Furthermore, the RTS and UAV – DEM assessment results are, to our knowledge, reported first time for the validation of these early stage ICESat-2 data in the AIS environment. Although the UAV - DEM coverage is relatively small and the CCR and RTS observation opportunities are relatively limited in comparison to the large number of GNSS observations along the 520 km GNSS traverse, their performances and achieved results in this study pave a way for future applications with aggregated observations at more sites for a more robust assessment. Therefore, our ICESat-2 validation methodology and sensor system will be improved to carry out the continued assessment of the ICESat-2 data, especially ATLAS performance during its later operation period.*"

- **An additional issue with using just 1 CCR overpass is the impact of the atmosphere on the signal or background noise levels to understand precision. The numbers reported here would represent the variability for one environmental scenario but doesn't capture a representative data quality variability relative to the many influences.**

We also accepted this suggestion and the changes are made in three sections as seen above.