We thank the reviewer for their helpful comments. Our response to specific points is given below:

1. *Please color the flight legs as a function of height so its directly visible which heights the cloud information comes from.*

   The aircraft tracks shown in Figure 1 in the manuscript are only for the above-cloud remote sensing runs. We will include additional plots (fig 1) showing the aircraft tracks during the in-cloud runs as a supplement to the revised manuscript, and update the text on p4 accordingly.

2. *Other than to avoid H/V complication, why you particularly interested in only focusing on analyzing and comparing the near-nadir observations? Considering that the orientation information is most prominent at slant views, and considering that the ICI viewing geometry is designed to be conical (?, or am I have a wrong impression, I remember the viewing geometry is somewhat similar to GMI), near-nadir view story might not be so suggestive for the satellite-borne mission. Please do elaborate somewhere in the paper about what conclusions might be changed when we move to slantwise view? What are the potential difficulties that observations and RTM might face with the slantwise view, especially for other microphysical parameters that further perplex the problem. Would multi-frequency V/H observations potentially helpful on resolving some of the issue, given that ARTS can now qualitatively capture the V/H through the Monte-Carlo method (i.e., not using the single scattering database, but using the 3D radiative transfer).*

   ICI is indeed a conical scanner. As noted in the manuscript (p5, l20), our main reason for focusing on the near-nadir observations is to maintain consistency between the viewing direction of the radiometer and the nadir-pointing lidar. Although it would be possible to interpolate the lidar vertical profiles along the slant-path of the off-nadir radiometer observations the process could introduce additional uncertainties. Additionally, we would expect any impact from horizontally aligned particles would be more significant at off-nadir viewing angles, but the ARTS scattering database (and other available databases) currently only contain randomly oriented particles. Evans and Stephens [1995] discuss the effect of viewing direction on sensitivity to IWP, and show that for spherical particles it is approximately constant. We therefore expect little difference in the results for nadir and off-nadir viewing directions for randomly oriented particles.

   Multi-frequency V/H observations are indeed expected to be helpful, particularly for detecting the presence of horizontally aligned particles [as discussed by Evans and Stephens, 1995, for example], and are included for the two window channels of ICI for this reason. Polarised brightness temperature can be simulated in ARTS with the RT4 solver used in this study, and the ARTS scattering database includes all required elements of the scattering matrix. However, as noted above the scattering database does not yet include horizontally aligned particles, and this limitation is common to other available databases.

   We will include additional discussion in the conclusions of the revised manuscript about the applicability of our nadir results to the slantwise views relevant for ICI.

3. *Its not a very good assumption to assume the particle shape doesnt change over the whole flight leg, neither horizontally nor vertically. Also, its natural somewhat to me to understand why 338 and 664 GHz cannot be achieved best-match at the same time, simply because they are sensitive to different size/shape of the particles which may likely co-exist in the vertical column at different altitudes. So I think you should discuss these possibilities that may likely happen in the real world and that partially account for the failure to match RTM simulations with observations.*

   We agree that particle shapes are likely to change both horizontally and vertically within the cloud, and that this may result in different single habits giving better matches at different frequencies or at different times; this is already discussed briefly in the text (p20). Using layered cloud models with different shapes in different altitude ranges could result in better agreement at all frequencies, but we would expect the brightness temperatures for a habit mixture to lie within the range of values predicted by the single habits. Given the difference in time and location between the in-situ and remote sensing observations and the limited microphysical information available from the lidar it is difficult to determine how suitable layered models could be derived based on the available observations. We will expand the discussion on this in the revised manuscript.

4. *How do you deal with the antenna pattern (i.e., line function) for the ISMAR sensor? That may cause 1-2 K warm bias even if you have a perfect background atmosphere setting.*

   We do not account for the antenna pattern and assume an idealised pencil beam. For the relatively narrow ISMAR beamwidths (less than 4 degrees FWHM) and nadir viewing geometry, clear sky simulations suggest that the impact of the main beam width will be significantly less than 0.1K. For off-nadir views the impact of the antenna pattern is much greater (order 1K) and would need to be accounted for in the simulations. We will include this information in section 4 of the revised manuscript.

5. *B895 is really not an idea leg for this study because its reaching the lower boundary of sub-mm sensitivity, as you also pointed out in your manuscript. So I would rather not put too much effort on match the B895 result channel noise, imperfect background atmosphere, etc., all these factors can beat down the observed BT difference for this flight.*

   Although it is true that B895 is approaching the limits of sub-mm sensitivity we think that it is notable that none of the habits with sufficient ice mass for a given particle size are capable of simulating the low sensitivity of the 664GHz observations.

6. *I still dont get why the smallPlateAggregate produces the best match for IWC (Fig. 8), but not for IWP (Fig. 10)? It seems to me that the problematic SectorSnowFlake actually produces the best match for the BT difference IWP relationship as shown in Fig. 10.*

   We do not fully understand this comment. Fig. 8 is a time-series of brightness temperatures and does not refer to IWC. The smallPlateAggregate produces a good match for IWC in figure 5, which simply implies that it has a mass-dimension relationship that is consistent with the in-situ PSDs and bulk IWC measurements. The fact that it gives a poor match to the brightness temperatures in fig. 10 shows that it does not have consistent scattering properties. We do not agree that the SectorSnowFlake provides a good match to the BT difference - IWP relationship in Fig 10 except perhaps for a few outlying points at 448 and 664GHz. This is further demonstrated in Table 3, where it has the largest biases and RMS differences of all the habits considered.

7. *In the discussion or conclusion section, please elaborate with a few sentences that whether your approach can be applied to previous campaigns, e.g., OLYMPEX with multi-frequency radar, CIP and CoSMIR? Further back in TC4 campaign, we have CoSSIR that is similar to ISMAR in some sense. If possible, using previous campaign observations that carried out in different weather regimes might aid greatly on identify the sub-mm capability and RTM caveats/advantages.*

   The general approach used here of combining in-situ observations with active remote sensing to constrain the cloud properties for input to radiative transfer models could certainly be applied to other datasets, although the details will vary depending on the available instrumentation (radar, lidar etc). Indeed, similar techniques have already been applied to CoSMIR [Olson et al., 2016], although the relatively low frequency channels (up to 183GHz) make it less applicable to ICI. Revisiting the CoSSIR observations from TC4 and CRYSTAL FACE could be interesting, although the existing retrieval studies [Evans et al., 2012, 2005] are, to a certain extent, a rather indirect validation of the radiative transfer models. We will mention the possibility of using CoSSIR observations in the conclusions of our revised manuscript.

References

K. F. Evans, J. R. Wang, D. O'C Starr, G. Heymsfield, L. Li, L. Tian, R. P. Lawson, A. J. Heymsfield, and A. Bansemer. Ice hydrometeor profile retrieval algorithm for high-frequency microwave radiometers: application to the CoSSIR instrument during TC4. *Atmospheric Measurement Techniques*, 5(9):2277–2306, 2012. doi: 10.5194/amt-5-2277-2012. URL `http://www.atmos-meas-tech.net/5/2277/2012/`.

K. Franklin Evans, James R. Wang, Paul E. Racette, Gerald Heymsfield, and Lihua Li. Ice cloud retrievals and analysis with the compact scanning submillimeter imaging radiometer and the cloud radar system during crystal face. *Journal of Applied Meteorology*, 44(6):839–859, 2005. doi: 10.1175/JAM2250.1. URL `http://dx.doi.org/10.1175/JAM2250.1`.

KF Evans and GL Stephens. Microwave radiative transfer through clouds composed of realistically shaped ice crystals. part ii: remote sensing of ice clouds. *Journal of the Atmospheric Sciences*, 52(11):2058–2072, 1995. doi: 10.1175/1520-0469(1995)052¡2058:MRTTCC¿2.0.CO;2.

William S. Olson, Lin Tian, Mircea Grecu, Kwo-Sen Kuo, Benjamin T. Johnson, Andrew J. Heymsfield, Aaron Bansemer, Gerald M. Heymsfield, James R. Wang, and Robert Meneghini. The microwave radiative properties of falling snow derived from nonspherical ice particle models. part ii: Initial testing using radar, radiometer and in situ observations. *Journal of Applied Meteorology and Climatology*, 55(3):709–722, 2016. doi: 10.1175/JAMC-D-15-0131.1. URL `https://doi.org/10.1175/JAMC-D-15-0131.1`.

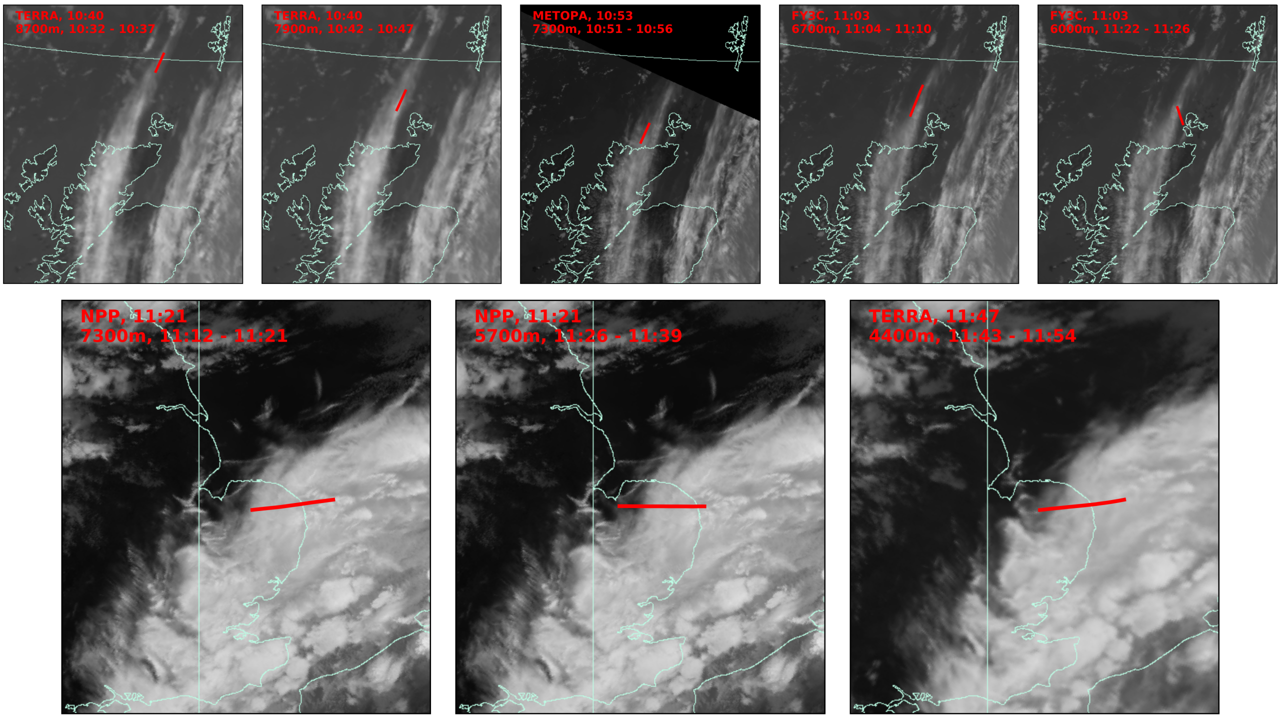

Figure 1: 10.8μm polar orbiting satellite images and aircraft tracks during in-situ sampling runs for B895 (top) and B939 (bottom). The satellite, overpass time, run altitude and run times are indicated on the plots.