Manuscript number: egusphere-2025-6184 (AMT)

Full title: Evaluation of ice hydrometeor retrieval using multi-band radar and millimeter-wave radiometer measurements from the IMPACTS campaign

Author(s): K. Ohara and H. Masunaga

The authors perform the validation of the ice hydrometeor retrievals from the synergistic multi-band radar and millimeter-wave radiometer measurements using an array of in-situ cloud probe measurements through tandem flights in the IMPACTS field campaign. A high-altitude flying aircraft deploys radars and radiometers to profile winter-storm clouds, and a low-altitude flying aircraft equips cloud probes to obtain the microphysical properties of the cloud along the flight track. A one-to-one comparison of ice water content, total number concentration, and median geometric diameter between the retrievals and in-situ measurements shows overall a reasonable agreement with some systematic biases in a deep convective cloud case. In addition, the impacts of a choice of radar-radiometer combinations and ice hydrometeor habits on the retrieval performance were evaluated. This manuscript is very well-written and well-organized. The sufficient details of the methodology and datasets for the retrievals were described. However, there are a couple of items that are critical but are not adequately explained in the manuscript. Please see my major comments. The topic in the present paper is suitable for *Atmospheric Measurement Techniques (AMT)*. This manuscript requires major revisions towards publication.

Major comments

First of all, winter storms contain both thermodynamic phases, namely liquid droplets and ice crystals. The present retrieval framework allows an estimation of IWC and total number concentrations of ice crystals. The treatment of liquid droplets is minimal; LWC is derived from Ku band radar reflectivity using an empirical approach, as briefly described in Lines 209-211. If the Ku-band-derived LWC is treated as a constant in the retrieval framework, two essential questions arise:

1) How are the potential biases and uncertainty of the LWC estimations incorporated into the retrieval framework, or is the LWC profile estimation treated as a truth with no uncertainty? If so, the retrieval framework may overfit the radar-radiometer measurements by compensating for potential biases in LWC by systematically biased IWC and total number concentration retrievals. This potential bias/uncertainty of LWC is particularly important to the effective use of microwave radiometer measurements, as these are sensitive to the presence of liquid droplets. The error associated with the LWC estimation propagates into the LWC and number concentration retrievals. I would like to see in the manuscript either a quantitative discussion regarding the impacts of LWC uncertainty in the retrievals or an integration of the LWC uncertainty in the retrieval system. Please see my minor comments #8 to consider performing the validation of LWC.

2) Is the use of Ku-band-derived LWC self-conflicting with the W-Ka-Ku-Tb-based retrievals? Ku band radar reflectivity can be fully explained by liquid droplets through the Ku-LWC relationship. Then, the algorithm tries to explain Ku band radar reflectivity by considering both liquid droplets and ice crystals. I might not fully understand the retrieval framework. It would be at least helpful from the readers' perspectives to describe more details about the treatment of liquid droplets in the algorithm. Also, citing Masunaga (2022) is less helpful from the reader's perspective as it is a textbook with >400 pages. Instead, I suggest that the authors briefly describe the Ku-LWC relationship-based LWC estimation.

Minor comments

1. Line 33, *"graupel and hail,"* these should be plural.

2. Line 54, *“combined radar-lidar.”* This should be “combined spaceborne radar-lidar,” as the airborne radar-lidar approach can profile in-cloud structures by flying through inner clouds, and ground-based radar-lidar can profile the lower portion of the clouds.
3. Lines 62-64: Figure 1b is the derivative of the 165.5 GHz brightness temperature with respect to a logarithmic IWC. This requires an IWC profile, but the corresponding description is lacking. Please clarify and explain how the IWC profile is obtained to compute the derivative.
4. Lines 101-102: Clouds can evolve even within 15 minutes. In particular, large hydrometeors can fall out quickly, which may skew the validation results. I would suggest adding discussions regarding the potential biases in the comparison due to the time difference between the remote sensing and in-situ measurements.
5. Line 131 *“… are averaged using a Gaussian weight.”* How did you make an average of radar reflectivity? A simple average does not work as radar reflectivity is in a dBZ unit. These quantities should be converted into radar backscattering power, averaged, and then converted back to a dBZ unit. The current manuscript was unclear whether the radar reflectivity was appropriately averaged.
6. Line 131 “GoSMIR” should be “CoSMIR.”
7. Lines 152-154: the 2D-S has a lower detection limit, and this probe misses lots of small hydrometeors, mainly liquid droplets but potentially small ice crystals that contribute to the total number concentration. Please elaborate on potential biases in in-situ measured PSDs and resultant impacts on the validation effort.
8. Table 3. I am wondering if P-3 aircraft deploy cloud probes that directly measure total water content (e.g., King probe) and both total and liquid water content (e.g., Nevzorov probe). This could provide a reliable dataset for the validation of the retrievals.
9. Lines 314-315: It seems that reading O25 is a prerequisite to reading this manuscript. I suggest the authors consider either inserting a sentence that guides the readers to read O25 in the Introduction or providing a summary of how covariance matrices are derived.
10. Section 4.1: Regarding the validation of the retrievals with in-situ measurements, did the authors make some averaging of the in-situ measured microphysical properties to match the large radiometer’s footprint, or are these treated as point measurements in the comparison?
11. Section 5.1: The terminal velocities obtained from both cloud probes and the retrievals rely on a series of empirical equations (i.e., Eqs. 4-7). Basically, this is a comparison of the mean diameters between cloud probes and the retrievals through the lens of terminal velocity, and it does not add any information from the direct *D*m comparison described in Section 4.1. Please describe any benefit in adding the terminal velocity comparison to the *D*m comparison.
12. Lines 460-461: *V*d should not be used as a data-filtering criterion for the *V*d plots. Suggest the authors use a different meteorological variable for a data-filtering criterion (e.g., temperature).
13. Figure 11f: In this histogram, I would add the original *V*d histogram, so that the readers can see how much 1) the width of the histogram is narrowed after a subtraction of estimated *V*t, and the peak of the histogram is shifted from a non-zero peak toward zero.
14. Section 5.2: The manuscript lacks a sufficient discussion of the difference between Zw+Tb and Zw+Tb+Zka+Zku. I can see it only in Lines 512-514. It seems to me that adding Zka and Zku does not much improve the retrievals. Ka band and Ku band radar reflectivities should be sensitive to large ice hydrometeors that may be beyond the Rayleigh scattering regime at W-band frequency. I just suspect that only a trivial improvement by adding the two radar frequencies may be due to a self-conflicting assumption in the Ku band (related to my major comment). Please elaborate on the reasons.
15. Figure 13: This plot is somewhat confusing. I am not certain how I can interpret the vertical axes. Is a positive value better than a negative value, or is any deviation from zero better?