Egusphere-2023-1118 RC manuscript review:

The paper presents results from two methods for quantifying riming of ice particles in mixed-phase clouds using normalized rime mass, based on measurements collected during the HALO-$(AC)^3$ aircraft campaign. The performance of the two methods is compared: one method combines radar and in situ data, while the other uses only in situ data. The good agreement between these two methods allows for confident utilization of in situ data alone, especially when coincident and co-located radar data is unavailable for quantifying riming. Additionally, the authors discuss correlations between normalized rime mass, radar reflectivity, and the level of cloud liquid water content in two study cases. Considering the significance of the presented results, I believe the manuscript represents a valuable contribution to AMT. However, I have a few major comments that I suggest to be addressed before the paper is published.

General comments:

The manuscript requires improvement in terms of presentation and readability. Several sentences are unclear and confusing. The in situ method utilizes in situ data with a particle diameter gap ranging from 1mm to 1.4mm. However, it's not explicitly stated whether the same subset of data was employed in the combined method. If not, what would be the potential impact?

In Section 2.3, second paragraph, the text mentions the use of Tb to estimate LWP. On line 155, it states that Tb is measured from an 89 GHz passive channel. However, on line 157, the authors mention, 'Thereby, the retrieval for the LWP is based on Tb derived from simulations with the Passive and Active Microwave radiative TRAnsfer tool (PAMTRA, Mech et al., 2020), using profiles of nearby dropsondes and artificial LWPs as input.' I'm not sure which Tb value was used for the LWP estimation,

In section 3.2, only dendrite aggregates are used to estimate the relation of M and the complexity parameter and Dmax. Would it be sufficient enough to represent for other types of rimed particles? Would the relation change if other simulated aggregates are used? Also, the relation is estimated with pixel size of 20um, it brings the question of whether the outcomes would differ if the pixel size were set to match that of the Cloud Imaging Probe (CIP) or the Precipitation Imaging Probe (PIP).?

The discussion of the unit for Dmax, A, and P appears confusing. Examining Eq (4), if the size measurements are divided by the instrument's (either CIP or PIP) pixel size, one will get the pixel number. Is it correct? Regarding page 10, line 260, the statement "… the higher the resolution of the snowflake image, the larger the perimeter (resulting in an infinitely large perimeter for an infinitely high resolution)" is unclear. Does the author mean an 'infinitely large perimeter' in terms of pixel numbers? Also, in Eq(2), what is the unit for Dmax? Is it the same as in Eq (4) (i.e. pixel numbers)?

In section 4.1, I would present the second paragraph comparing M from the two method first before discussing about riming fraction with different thresholds of M. Furthermore, in Fig. 5, the authors may consider combining the two histogram plots into one (supercomposing) for a more effective visual comparison.

I find it challenging to follow Section 4. The current organization presents Section 4.1 as an analysis of data from collocated flights, Section 4.2 as all in situ only flights, Section 4.3 as an example of a case study for collocated flights, and Section 4.4 as an example of a case study for in situ only flights. The section structure could benefit from better organization. It might be helpful for the authors to consider adding explanatory text to clarify why this particular structure is necessary or beneficial.

In the case study 1 (collocated flight segment), it is not clear to me when the measurements are taken near cloud top, both methods are unreliable or only in situ method? Would the authors suggest which method should be used in those scenarios? In Fig. 10k, I don't see the correlation between time series of M from the combined and in situ methods.

Specific comments:

- Page 1, line 11: "… we obtain average rimed fractions of 77 % and 75 %." of what clouds or which study case?
- Page 1, line 12: "…the radar volume (about 45 m footprint diameter) …" At what distance the radar footprint (beamwidth?) is calculated? And what is the vertical resolution? (to give an ideal of a radar volume)
- Page 6, line 138: "… to be ice crystals (liquid droplets) …" I'm not sure I understand this statement.
- Page 7, line 171: "In both cases, LWC measurements were averaged to be on a regular vertical grid with a resolution of 10 m." Could the author provide more details on the setup of the measurements, specifically how they were configured to obtain vertical LWC profiles?
- Page 7, line 173: "integrated LWC": do the author mean LWC calculated from PSD? Could the authors provide insights or comments on the accuracy comparison between Liquid Water Content (LWC) derived from Particle Size Distribution (PSD) measurements and LWC obtained through Nevzorov probe readings (LWC(Nev))?
- Page 8, line 205: "… the diameter of the smallest encompassing circle…" of what?
- Page 8, line 210: the complexity parameter is not defined until section 3.2 in page 10.
- Page 8. Line 223: "… and interpolate a_m and b_m to obtain parameters for a continuous M." I find it unclear which variable the interpolation is performed with respect to. Are a_m and b_m computed for each radar volume or for each flight segment?
- Page 9, line 228: what is the definition of the state vector x?
- Page 9, line 240: "… and Sy is the corresponding measurement uncertainty of 1.5 dB." Because Sy is a matrix, please rephrase this sentence.
- Page 9, line 240: "…in.." should be "at".
- Page 10: sentences 269-270 should be placed after Eq(6).
- Page 11,line 277: what is the spatial resolution corresponding to an averaging window of 30s?
- Page 11, line 280: "… suggest shapes rounder than a sphere" I'm not sure I understand this.
- Page 11, line 283: "By comparison to the combined method in addition to manual inspection of CIP and PIP images, we find that in sum at least …" What are the criteria the authors used in this comparison?
- Page 11, line 286: there's no Fig 3.2.
- Page 11, line 290: "Unrimed particles (Fig. 4 (a), left) have much more complex shapes and therefore larger χ, than more heavily rimed ones (Fig. 4 (a), right), which are almost spherical (χ close to 1)." This statement is true for the example shown in Fig 4 but might not be true in general, please consider to rewrite it to avoid confusion.

- Page 13, line 306: are all collocated flights segments are used to generated the histogram plots and the joint distribution in Fig. 5?
- Page 14, line 313: what is the difference in flight 10 April?
- Page 14, line 323: "… where cloud bottom is the lowest signal above 150 m". Please consider to rewrite this stanement. Cloud base could be defined based on a threshold with respect to the radar noise floor. I also wondering if 150m is good enough to avoid the ground clutter sidelobes.
- Page 14, line 329: "Both methods show larger disagreement and (local) minima of riming between -10 and -15 °C." Do the authors mean local minima of both methods and the disagreement between the two methods? If so, please consider to make this sentence clear.
- Page 14, line 334-336: If the data from April 10th is deemed unreliable for the in situ method, would the authors contemplate excluding it from the analysis. The inclusion of potentially erroneous data could compromise the validity of the analysis, especially when dealing with temperatures above -10°C.
- Page 14, line 337: "There is no clear dependence of riming on LWC." How's about the dependence of riming on TWC? I would expect to see high correlation between M and TWC, that also confirm the reliability of the methods.
- Page 15, Fig 7: legend "retrieval" should be "combined". It is hard to read the value of M. Is the black dashed line M=0.01? I couldn't find the definition of the 'normalized position in cloud' parameter used in Fig. 7m.
- Page 16, line 351: which retrieval?
- Page 16, line 352-355: "However ,…in Appendix (C)." I'm not sure what are the key points in this discussion.
- Page 16, line 360: "Figure 8 analyses the …" please rewrite.
- Page 16, line 379-380: " We can conclude that the collocated measurements are representative of the complete Polar 6 data set above 150 m flight altitude." I'm not sure I understand this statement.
- Page 18, line 412: "… close to the upper edge of the radar signal…" what is the upper edge of the radar signal?
- Page 18, line 423: what do the authors mean by "MPC variability". Variability in what parameters of MPC?
- Page 18, line 427: "… low LWC". Does the author mean "low LWC at flight altitude" ? In Fig. 11c I only see some spikes in the LWC curve or is it overlapped by the TWC curve?
- Fig 10 and 11: In addition to the D32 time series plot, it would be helpful if the authors could include a Particle Size Distribution (PSD) spectrum plot and/or a total number concentration plot. This additional information would assist in identifying segments with low particle counts. Also, what does the dash-dot line labeled as "images" mean?
- Page 22, summary section, line 485-487: In the discussion concerning the performance of the combined method near cloud base, there seems to be an implication that the method might not be reliable in this context. I would suggest a case study to explore this scenario further. Could the authors provide details on the specific limitations in cloud probe detection efficiency that might have affected the method's performance? For example, is it related to particle sizes exceeding the PIP measuring range or challenges in accurately capturing high LWC with the Nevzorov probe, or contamination of ground clutter on radar signal at low altitudes.