Response to egusphere-2026-1291:

We are deeply grateful to the two reviewers for sharing your constructive helpful insights. Below please kindly find your comments in black, our detailed response in blue, and the changes made in the revised manuscript can be found in red.

Summary of the overall improvements of the revised version:

- Both reviewers provide many detailed suggestions about improving the figure quality/presentation/caption. Fig. 1, 3, 5, 6, 7, 10, 11, 12 and 13 have now been edited following the two reviewers' suggestions.
- As this paper covers two topics through conducting three tasks that are inter-linked, we added a paragraph in the introduction section to make sure readers are clear about the logic and scopes of this study before reading Section 2.4. The manuscript title is also changed to reflect the content of both topics.
- We replace Fig. 13 with a more complicated scenario to demonstrate the criticality of PD in revealing the vertical structure of phase distribution. We add a paragraph at the end of Task #2 to reiterate the justification of training and testing dataset configuration.

Review #1:

This manuscript presents a closure study that employs active, passive, and in-situ measurements to explore and emphasise the importance of knowledge of hydrometeor types. In the study, a hydrometeor classification is performed and subsequently used for retrievals, exploring how hydrometeor type assumptions impact agreement between simulations and observations. Additionally, the study presents how sub-millimetre polarimetric measurements can benefit retrievals of vertical hydrometeor type distributions and ice particle size.

The study is thoroughly performed, well-described, and results are carefully validated. Publication following minor revisions is recommended. Specific comments and questions are as follows:

Thank you for your kind encouragement.

Section 3.1:

It is interesting to see the impact of the assumptions of hydrometeor types in Fig. 6. However, it would be helpful to also see where in the cloud the different ice classes were assumed for Figs. 6a, 6b and 6c. For example, as shown in Fig. 3e.

Thanks for this nice suggestion. Instead of adding a new figure, we've now replaced Fig. 3 with the same time period that's consistent with Fig. 6 and Fig. 7 (17-17.6 UTC). We've modified the description of Fig. 3 and added some sentences when discussing Fig. 6 to offer the context (see red letters for modified sentences).

The authors state that "Only using the detailed hydrometeor types can we reproduce cold enough TB depressions that match observations well." (lines 291-292). This appears to be true - using the 8-class hydrometeor types leads to clear improvements. However, the simulations still differ from the observations, e.g. by ~20 K at 17.4 UTC for 325.15+-11.5 GHz. Do the authors attribute this to still imperfect hydrometeor type classification? Or is this caused by another aspect of the simulations?

This is a great question. Although we don't know the exact answer, we suspect that it's likely caused by our overestimation of triple-frequency radar retrievals for liquid water content within the melting layer that subsequently brings the warm bias to the simulated TBs. Because liquid emission signal is proportional to ~1/f where f is the channel frequency, the warm bias is expected to decrease with increasing channel frequency, which is supported by what we can observe at 170.5 versus 325.15/11.5 GHz (Fig. 7a and 7b). The ignorance of potentially horizontally oriented large snow aggregates between 5-7.5 km might be another candidate to explain the warm bias, as TB_H would be excessively cold in such a scenario. We now modify the text to include this discussion, which reads as:

"Only using the detailed hydrometeor types can we reproduce cold enough T B depressions that ~~match observations well~~ can produce the closest resemblance of the observations, although they are still 10 − 20K too warm compared to the observations for the two wing bands at the deep convective cores (Fig. 7a and b). Such a warm bias might be caused by the excessive LWC retrieved at and below the melting layer, which also causes the low bias of simulated W-band radar reflectivity there (Fig. 6c). It may also be explained by ignoring the horizontally oriented snow aggregates between 5−7.5km (Fig. 3f) that would introduce a colder TB_H signal due to the enhanced scattering."

Section 3.2:

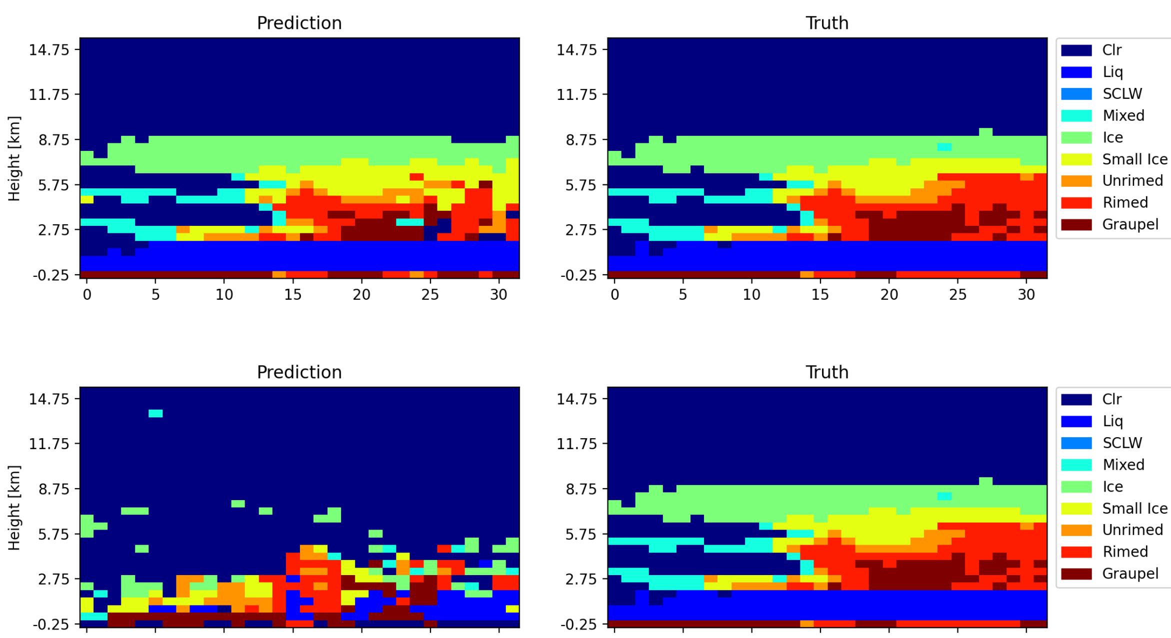
The ML model was trained on the Feb. 05 case and tested on the Jan. 15 case. This is reasonable and well-motivated. Given than the Jan. 15 case was earlier described as atypical (Sect. 2.1), could the authors discuss how this might affect the generalisability of the results shown? What might the model's performance look like for more typical cases?

We intentionally designed to train on a typical winter storm case (Feb. 05) and predict on an atypical case (Jan. 15) to test the robustness of our hypothesis that CoSSIR-PD signals contain useful information about the vertical hydrometeor phase distribution. As most of the IMPACTS cases are typical winter storm cases, making a good prediction on another winter storm case might be partially explained by the similarity of winter storm vertical structures. Unfortunately we couldn’t afford to expand this work to include more IMPACTS campaign days.

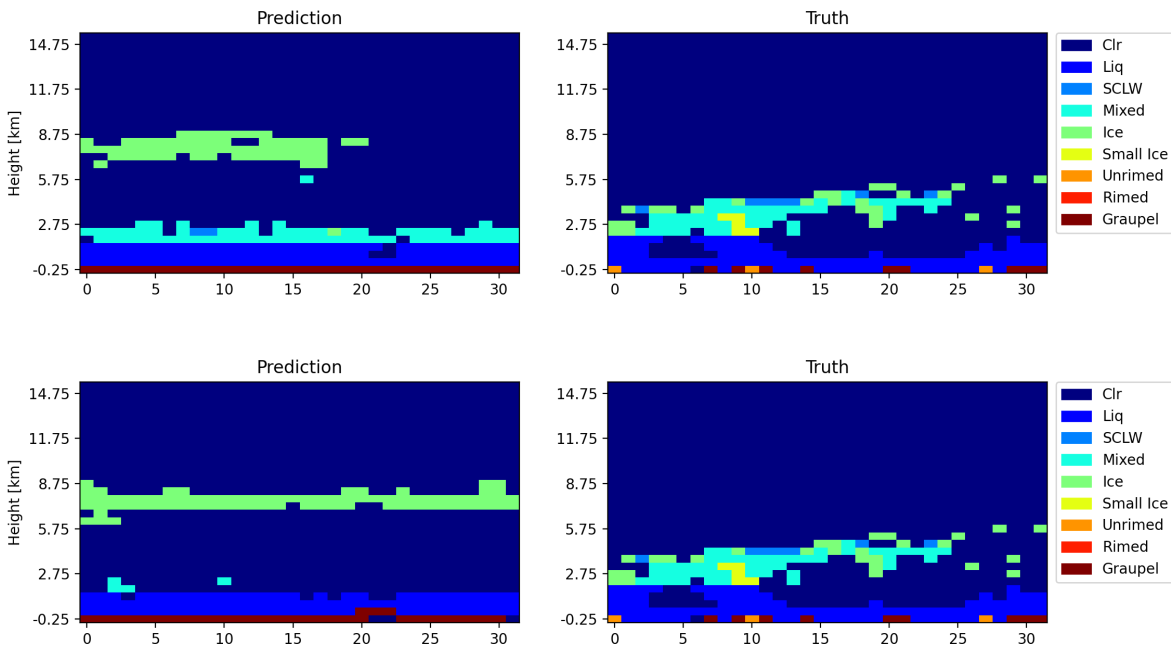
Line 370 – 395 have been modified significantly to compile in all these great suggestions.

In Fig. 11, the two panels showing the reference (11b and 11d) are not the same. Rather, it looks like panel d has a small horizontal offset compared to panel b. Were the targets not exactly the same for both models? Or is it a plotting artifact? If the latter, then I suggest plotting the same reference for both to aid comparison between models, or at least making the difference between the two clearer through the choice of variable on the x-axis. The same applies to Fig. 12.

Thank you for catching that. Now the figures in the revised manuscript have been updated to keep the “truth” identical for all cases showing (hence reduced from 2X2 panels to 3X1 panels for Fig. 11 and Fig. 12). The second planetary boundary layer (PBL) cloud case is replaced with a more complicated two-layer broken PBL cloud scenario to showcase that while both ML models predict a wrong thin ice cloud layer in the upper troposphere, the yes-PD model still makes a better prediction for the PBL mixed-phase cloud above the liquid layer than the no-PD model. The written text also changed accordingly. See below for the cases used in the modified Fig. 11 and Fig. 12, respectively.



Modified Fig. 11. Top panel is from yes-PD fullset model prediction, while the bottom panel is from no-PD subset model prediction (in the revised version, panel #4 is removed due to redundancy).



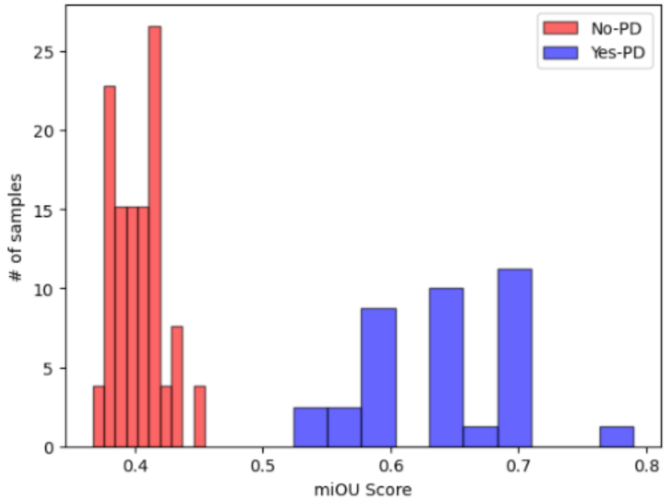
Same with modified Fig. 11, except for a PBL two-layer cloud case.

The independent testing dataset contains 432 images, and they are randomly grouped into 29 batches with each batch containing 16 images for prediction. Therefore, when we call the ML model each time to generate the prediction, the image order is randomized, and I had to visually identify the ones with the same "truth". I have now uploaded the entire prediction image sets on the zenodo link (https://doi.org/10.5281/zenodo.19161985, version 2; under "Triple_freq_paper_ML.tgz" -> "prediction_figures" folder). Please take a look if you are interested.

The improvement upon including PD measurements is very clearly illustrated in Figs. 11 and 12, and quite impressive. Although, as the authors mention, the no-PD model does appear to capture the cloud top for the multi-layer thick cloud, it does seem to struggle overall. Likewise, it gives a false cloud for the single layer cloud case. It was therefore surprising that it achieves an mIoU score of 0.545, which the authors frame earlier as a typically good result. Do the authors attribute this to the model's ability to capture the cloud top/thickness, or did it perform better in other scenes than those shown?

Thanks for helping us identifying a code bug. The mIoU score for the no-PD model was actually wrongly computed, and it should be 0.4024, while the mIoU score for the fullset yes-PD model remains to be 0.642.

Please see the mIoU score distribution amongst all 432 independent prediction samples below.



The “yes-PD” (i.e., fullset) model consistently outperforms the “no-PD” (i.e., subset) model for geometrically thick cloud cases. For thin cloud cases, they are equally bad, especially for thin boundary layer cloud scenarios, but the “yes-PD” prediction in general looks in better agreement with the “truth”. The miOU score statistics can be found on the zenodo link (https://doi.org/10.5281/zenodo.19161985; version 2) under “Triple_freq_paper_ML.tgz”-> “predictions”-> “iOU_score_prediction_statistics”.

Additional comments:

Line 433: A reference is missing.

Citation fixed.

Line 222: Framework is misspelled.

Corrected. Thanks for spotting that!

CRS, HIWRAP acronyms are not defined. Please check acronyms throughout.

CRS = cloud radar system

HIWRAP = High-altitude Imaging Wind and Rain Airborne Profiler

These acronyms were spelled out when they first appeared at Line 70-71.

Figure 1: Top panels: What do the dots signify? The stretches shown in lower panels are not clearly marked. Lower panels lack a colorbar.

Dots mark the aircraft positions at :05, :10, :15, ... The bottom panels of Fig. 1 were previously screenshots (uncalibrated L1 quicklook images) taken directly from the campaign website. Now they are replaced by our own plots of the calibrated L1B data. Figure caption has been modified.

Figure 5: The interpretation of the figure would be simplified by using log10, instead of log. The color ranges are very wide and appear poorly adjusted. For example some of the color ranges reach 10, matching 22000 kg/m2. For lWC, the lower end is at -40, matching 4e-18 kg/m2. The colorbar ticks are partially covered by the colorbars. It would also be helpful to clearly differentiate in the figure where the separate overpasses start and end, if possible to do so clearly.

Thanks for the suggestion. We have now updated the colorbar to log10 and adjusted the colorbar locations to avoid overlapping. I'm not quite sure if I understand your suggestion about "separate overpasses start and end". If you meant to suggest mark the boundaries of overlapping areas where IWC and LWC co-exists, I think it would make the figure too busy and harder to read. I've decided to not include that in the updated Fig. 5, but just to show you what it looks like in case you feel it's better than the current Fig. 5:

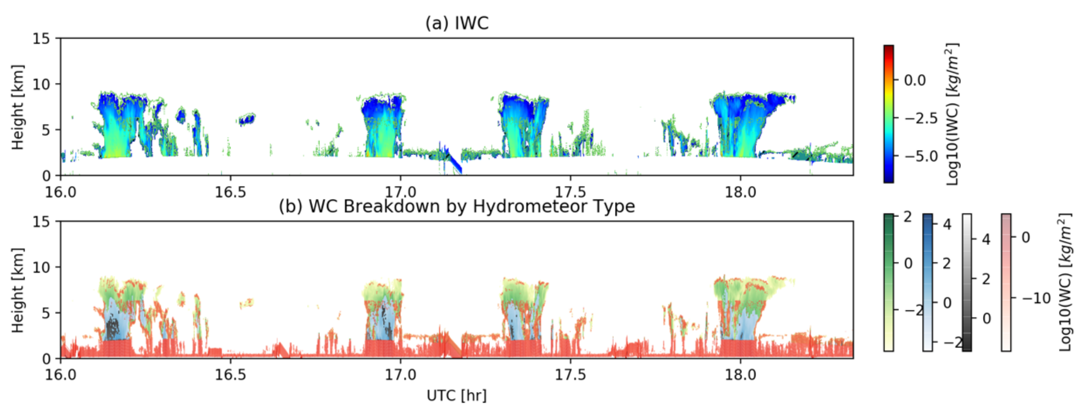


Fig. 5 plus hatched area in panel (a) to outline the mixed-phase regime where LWC and IWC co-exist.

Figure 6: Add units to colorbars. The top height could be set lower than 15 km.

Done adding units. We set all cross-section figures with the same height range of [0, 15 km] now for uniformity. 15 km is about the cruising level of ER-2 so we believe it's more appropriate to leave the blank spaces in those figures just to demonstrate the full column of active sensor measurements.

Figure 7: Unit for y-axis is missing. Label text is quite small.

Units added to the figure and label font enlarged.

Figure 10: Labels on y-axis of panel c are too small (even on screen after zooming in). The meaning of the labels must be explained. Panel c lacks a colorbar.

Since there are 36 features, enlarging font size would make the y-axis label too crowded unless we make panel C a separated large figure. We've listed the feature names in the figure caption, and now in the revised manuscript we add more details in the description. Please check the revised figure caption highlighted by red letters. Panel c does not need a colorbar as every feature is normalized to 0-1 value range for ML training purpose (this has been already described in the figure caption). The value therefore doesn't contain physical meaning.

Figure 13: The figure text should clarify if simulations or observations shown. The colorbar label overlaps with the y-axis label.

They are purely from the observations. It's now clarified in the figure caption, and the overlapping labels have been fixed.