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?

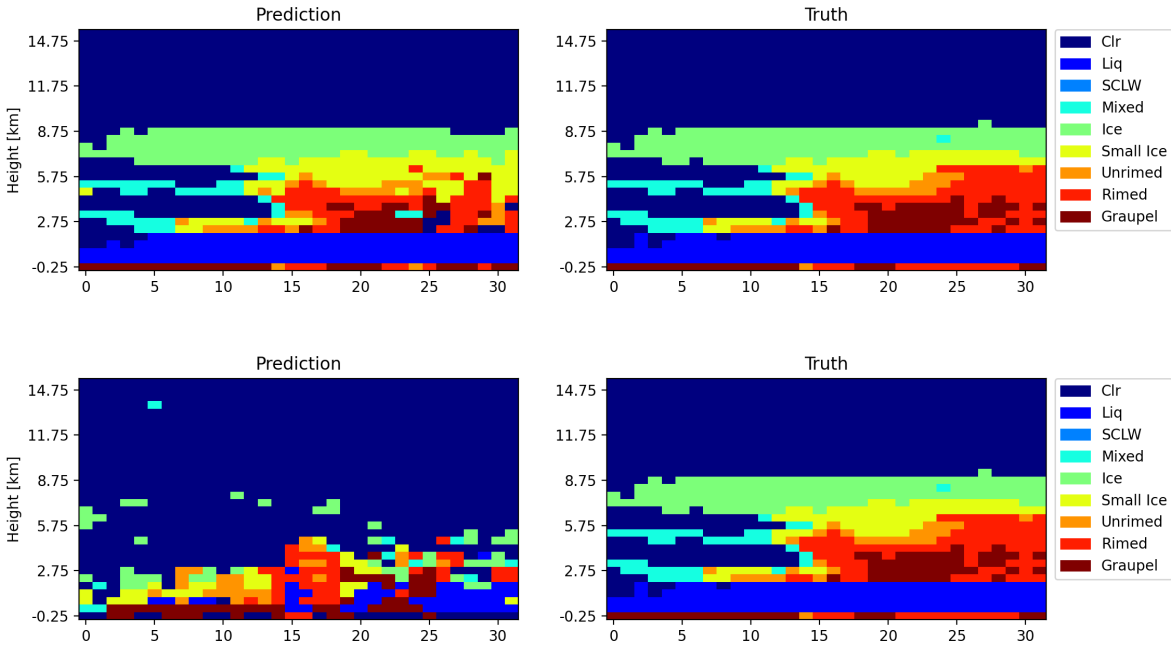
We intentionally designed to train on a typical winter storm case (Feb. 05) and predict on an atypical case (Jan. 15) to test the robustness of our hypothesis that CoSSIR-PD signals contain useful information about the vertical hydrometeor phase distribution. As most of the IMPACTS cases are typical 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.

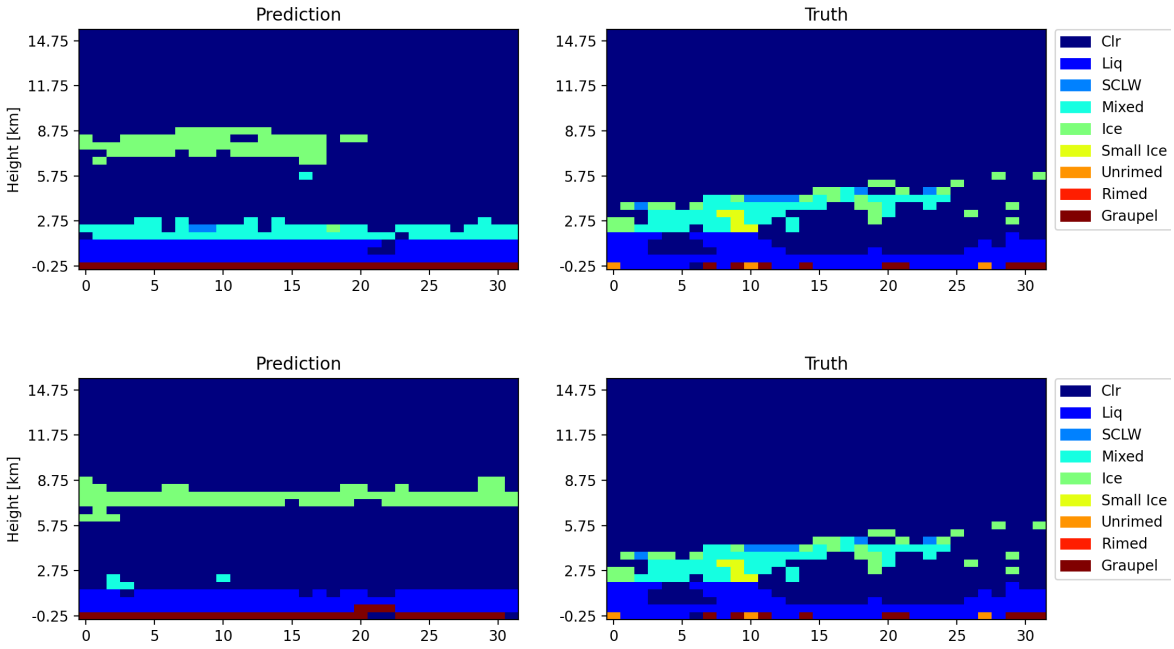
Line 370 – 395 have been modified significantly to compile in all these great suggestions.

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).



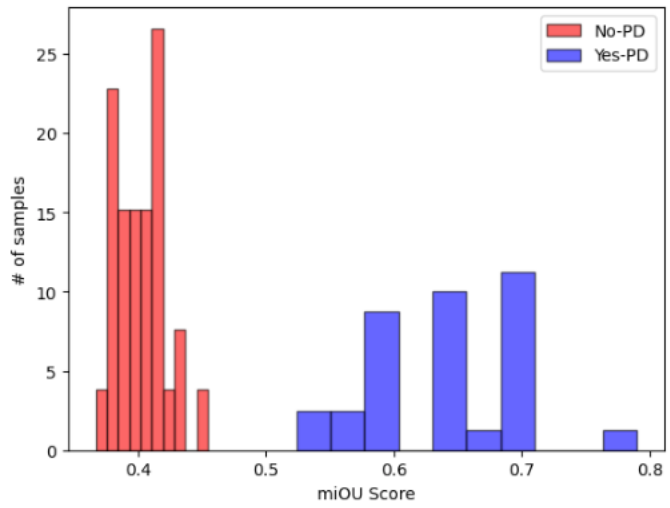
Same with modified Fig. 11, except for a PBL two-layer cloud case.

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:

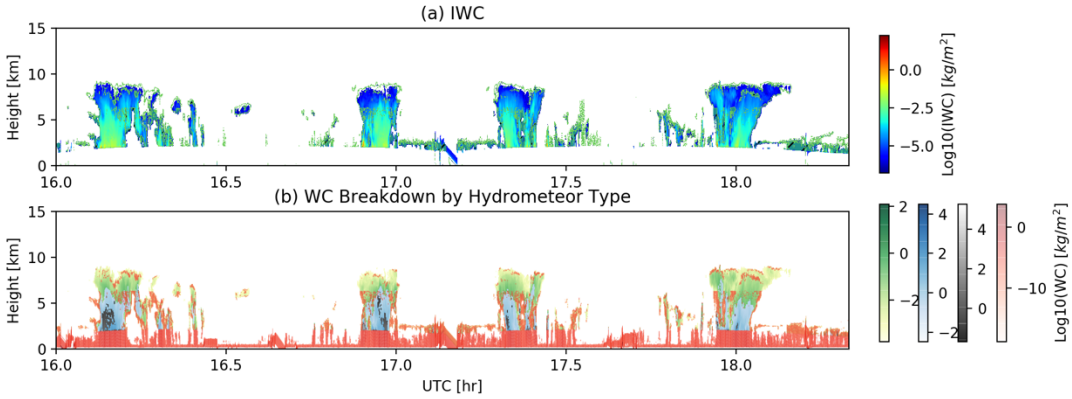


Fig. 5 plus hatched area in panel (a) to outline the mixed-phase regime where LWC and IWC co-exist.

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.

We are grateful to both reviewers for their acknowledgement of the novelty and importance of this work, as well as their insightful comments and suggestions. 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.

General remarks:

In general, it is very interesting to see a closure study combining collocated passive and active remote sensing as well as in situ measurements, investigating also the impact of the assumed vertical hydrometeor profiles. This is important for retrieval development and validation. In addition, the potential of polarimetric radiometer measurements is an important topic, in particular, also for upcoming spaceborne missions. However, currently I am missing a concise storyline and the paper appears as many different things have been stacked together. First, the closure study is presented, then a new vertical classification method is shown, and then a retrieval of ice microphysics is described and compared again to in situ measurements. The title refers rather to the second part of the paper highlighting the potential of polarimeter-radiometers, whereas the major part of the paper focuses on the closure study between different measurements. I would suggest to either think about splitting the paper into two parts or slightly reframing / restructuring the current paper.

Thanks for your acknowledgement of the value of this work. We thought about splitting into two companion papers. However, there is a simple answer to your question: budget constraint. This project has reached its end, and we only have budgets to pay for publication fees for two papers (this one and Liu et al., 2026) instead of 3.

As results generated from the closure study are essentially the inputs for Objectives #2 and #3, we feel it is justified to combine the polarimetry merit study with the closure study together as a comprehensive case study. However, we do agree with you that this paper has studied two separated topics, which makes it also challenging to come up with a proper title (as you also pointed out).

We have now changed the paper title to "A Cross-Instrument Closure and Polarimetric Study for Improving Sub-millimeter Polarimeter-Radiometer Ice Cloud Microphysics Retrievals". Hopefully this is a comprehensive title. We also add some justifications in Section 2.4 "Framework for Closure Study" and add "polarimetric" in the Section 2.4 title.

See also the specific comments below.

In section 2.4 you nicely explain the framework. However, it would have been really helpful to briefly refer to the objectives already earlier (in the introduction). After reading the introduction, the purpose and objectives of the paper were not really clear to me.

This is a great suggestion. I can see a logic-flow gap in the introduction section, and a paragraph is added now to brief the scope of this work. This paragraph reads as:

"This paper executes a comprehensive, multi-faceted research framework organized around three inter-linked objectives. The first objective stresses the criticality of cross-instrument consistency about ice microphysics properties, and the other two objectives further explore the scientific utility of multi-channel sub-mm polarimetry in better retrieving the vertical hydrometeor phase distribution and ice particle size. By linking a strict active/passive/in-situ closure study with targeted polarimetric applications, this work aims at establishing concrete baseline protocols for optimizing the operational/research science products of upcoming spaceborne sub-mm radiometer missions."

Every plot shows another time range and different time labelling (Fig. 3: 16:35-17:10, Fig. 6: 17.2 – 17.6, Fig. 7 and 8: 17.0-17.8, Fig. 9: 16.2 – 17.4). Please unify this and show the same time range and use the same labelling. This makes it much easier to relate e.g. hydrometeor types to observed differences in the radar reflectivity etc.

Thanks! We've updated the horizontal axes of Fig. 1, 3, 5, 6, 7, 8. Now the time stamp is uniformly presented in decimals in all figures and text. The only exception is the top panel of Fig. 1 and related context, because these are screenshots from the IMPACTS campaign website, and are associated with the weather reports in McMurdie et al. (2023). It's easier to cross-reference using the HH:MM format.

The companion paper by Liu et al. (2026) introducing the new vertical hydrometeor classification is mentioned. In the current form, I am not sure if section 3.2 is really needed

and adds new information if there is an additional paper introducing the vertical classification properly? That polarization information adds additional information content is, in principle, already visible from the input features. In addition, is the machine learning approach really valid? Training was performed using the data from Feb. 5, but the classification was then applied to measurements of Jan. 15, which is according to you quite different. Did the training data then cover the entire parameter space and how reliable are the potentially extrapolated results? Moreover, you trained the model including polarization data and performed some hyperparameter tuning. Then, you used this optimal setup again for input data without polarization information and show that the performance is worse. I would not expect that the same configuration of hyperparameters works equally well in both cases and therefore, part of the improvement due to polarization information could also be explained by the architecture and hyperparameter settings. As a consistency check, you could, for example, tune the hyperparameters again for the non-PD case and then apply this model also to the case including PD and compare the differences. This would help quantify the effect of the model tuning and settings.

Liu et al. (2026) has been published in its pre-print format at:

Liu, Y., Gong, J., Adams, I. S., Kroodsma, R. A., Chen, R., Wu, D. L., Bennartz, R., and Braun, S. A.: A Hybrid Ice Hydrometeor Retrieval Algorithm for (Sub)millimeter-Wave Radiometers in Support of the PolSIR and PMM Missions, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2026-1348, 2026.

Sorry about some misunderstanding here. Liu et al. (2026) describes the operational algorithm for the two upcoming sub-mm radiometer missions, which assumes one ice microphysical habit through the whole column for each pixel, and Liu et al. (2026) evaluated the impacts on uncertainty of retrieval products resulted from different ice habit assumptions. Our operational products will only include vertically integrated or averaged quantities (ice water path and mass-weighted ice particle size), while the profiling capability remains in the research realm and Liu et al. (2026) focuses on retrieving vertical ice mass distribution using a uniform ice habit assumption. This paper takes a distinctively different route for the purpose of eluding some potential new research capability/product for sub-mm radiometers. Basically section 3.2 has no overlap with the content in Liu et al. (2026).

We now add this sentence to clarify in the revised paper:

“Note that While Liu et al. (2026) builds its algorithm upon the heritage Bayesian theorem-based retrievals, it assumes one ice habit at a time for the whole column to retrieve the

vertical mass distribution, while this work focuses on exploring potential research product enabled by additional polarimetric measurements."

Why was January 15 chosen for the case study if it is an atypical case? The flight on February 5 is mentioned in Sect. 2.1, but at this point it was not clear to me that the paper focuses on the case study of January 15 and that the data from Feb. 5 is only used for the training in the machine learning approach. Why did you not directly use February 5?

A similar question was raised by the other reviewer. Basically, IMPACTS is a winter storm campaign and hence most of the flight legs sample winter storm cases. So if we train on multiple winter storm cases and run prediction for another winter storm case, we cannot assess whether part of the good performance is attributed to the similarity between training and testing datasets, rather than because of real physical signals embedded in the measurements. We now add a paragraph at the bottom of Section 3.2:

"Lastly, it is worth emphasizing that we intentionally designed to train on a typical winter storm case (Feb. 05) and to 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."

Specific comments:

Fig.1: Please add a color bar for the radar reflectivity.

Fig. 1 updated to use calibrated L1 data together with a colorbar (previously from screenshots of quicklook images).

l. 108: Please add a reference to Sect. 2.2.3 where more details about RICE are provided.

A reference for RICE is added:

RICE reference: Cober, S. G., Isaac, G. A., & Korolev, A. V. (2001). Assessing the Rosemount Icing Detector with In Situ Measurements. Journal of Atmospheric and Oceanic Technology, 18(4), 515–528. https://doi.org/10.1175/1520-0426(2001)018<0515:ATRIDW>2.0.CO;2

l. 161: related to the general comment above: it would be helpful to make clear that the case study uses data from January 15. As far as I understood the surface emissivity model is only used for the forward model for the case study and not applied to Feb. 5, so the information that Feb. 5 was over land is not needed here, which could help avoid confusion.

You are absolutely correct that Jan. 15 is mostly over ocean, and Feb. 5 case is mostly over land (some part over the great lake area). We employed the TESSEM2 emissivity model over ocean/lake, and TESLEM2 emissivity model over land. Both emissivity models had been validated previously at CoSSIR frequencies (at least up to 300 GHz). These details had been described in Section 2.2.2. Therefore, we do not worry too much about emissivity-induced errors/uncertainties for this work.

l. 164: Have you analyzed the sensitivity of the simulation results to the selected representation of the retrieved habits in ARTS? Here for example, you selected column aggregates for ice, later on you use bullet rosettes?

Nope. This is a great suggestion, but unfortunately we are out of our bandwidth of running deeper into the weeds without further funding support.

l. 180: Why do you expect dominantly horizontally oriented ice crystals? As far as I know, the fraction of oriented crystals is typically small.

This might be true for high cloud colder than ~ -30 degC (as some studies using CALIPSO depolarization measurements suggest), but horizontal orientation is prevailing at -10 to -20 degC, especially in the stratiform layer. For example, Westbrook et al. (2010) identified up to 85% chances of horizontally oriented ice in the stratiform cloud from a ground lidar measurement. For larger ice particles that sub-mm/MW spectrum is sensitive to, their orientation is primarily subject to micro-environment of aerodynamics, where this paper has a nice section explaining the mechanism.

Westbrook, C. D., Illingworth, A. J., O'Connor, E. J., and Hogan, R. J.: Doppler lidar measurements of oriented planar ice crystals falling from supercooled and glaciated layer clouds, Q. J. Roy. Meteor. Soc., 136, 260–276, https://doi.org/10.1002/qj.528, 2010.

l. 202: You use a 1D radiative transfer solver and apply the independent column approximation? Are 3D radiative effects negligible? Which vertical resolution did you use? Have you analyzed the influence of the applied vertical resolution of the IWC etc.? There might also be other work on that, which could be cited?

Yes, we used 1D solver, independent beam approximation (IBA) and plane-parallel atmosphere assumption. 3D radiative effect and IBA-induced error has been studied by Barlakas and Eriksson (2020). This paper concludes that IBA induced bias is significantly smaller for a small footprint size versus a typical PMW footprint size. As this is an airborne campaign where CoSSIR footprint size is ~ 1 km, the IBA-induced bias should be rather negligible. For a spaceborne sub-mm radiometer, this bias is not negligible. We used native CRS radar vertical resolution for all RT simulation, which is ~ 25 m with a total of 609

vertical layers. We've now add a few sentences when we introduce ARTS simulation set-ups.

"We assumed 1D atmosphere and independent beam approximation (IBA), which would introduce a negligible low bias to simulated TBs for CoSSIR according to Barlakas and Eriksson (2020). CRS range bin resolution of 25 m is employed for RT model vertical resolution with a total of 609 vertical layers."

Barlakas, V.; Eriksson, P. Three Dimensional Radiative Effects in Passive Millimeter/Sub-Millimeter All-sky Observations. Remote Sens. ***2020****, 12, 531. https://doi.org/10.3390/rs12030531*

l. 205: You use the radar-derived IWC and LWC in the forward operator. How is the effective diameter determined? To characterize the radiative properties of a cloud, the particle size is needed in addition to IWC/LWC and particle shape. Consider adding a small subsection describing the forward operator and corresponding setup in more detail.

We employed Fields et al. (2007) one-moment scheme, meaning that one IWC/LWC corresponds to one mass-weighted effective diameter (Dme), which is implicitly computed within the ARTS model. One can choose two-moment scheme to let Dme and IWC/LWC disentangle from each other, which would make sense for triple-frequency radar approach. This is described in the last sentence in Section 2.2.2.

l. 212: If the assumption of 100% oriented crystals is unrealistic, why is it assumed then?

Because currently we can only simulate two scenarios: 100% randomly oriented or 100% horizontally orientation. In reality, the orientation of these large ice aggregates tends to follow a distribution with the mean at the canting angle, the latter of which is determined by the in-cloud wind shear and vertical velocity distribution. Brath et al. (2020) provides a post-mixing solution by assuming a Gaussian distribution with mean orientation directly at horizontal plane, which is closer to reality, but we couldn't adopt due to computational cost. However, at 52.8 deg viewing angle, there's no mixing of U and V Stokes into the interpretation of PD, and PD = I-Q according to Barlakas et al. (2021). Hence we choose to use only use observations and simulations at this view angle.

l. 218: Why did you choose bullet rosettes here?

Because the RT4 solver for the oriented ice simulation only provides 4 types of ice habit instead of 34 habits that are available to randomly oriented ice database. These 4 ice habits are column, hexagon-plate, rosette and sphere (as a baseline).

Fig. 5: The tick labels of the color bars overlap. Maybe consider adding an additional panel showing only the LWC, such that there is one panel for general IWC, one for LWC, and one

for the detailed IWC of the different hydrometer types? You could avoid overlapping of the liquid and ice classes then.

The colorbar label overlap issue has been fixed in the revised figure. The second request is also suggested by the other reviewer, who asks for hatch the mixed-phase region. We've tried to implement that, and feel it looks too busy. However, if we plot the LWC separately, it's harder to visually locate where they are. Please take a look at the figure below and let us know if you think it's better to use this version.

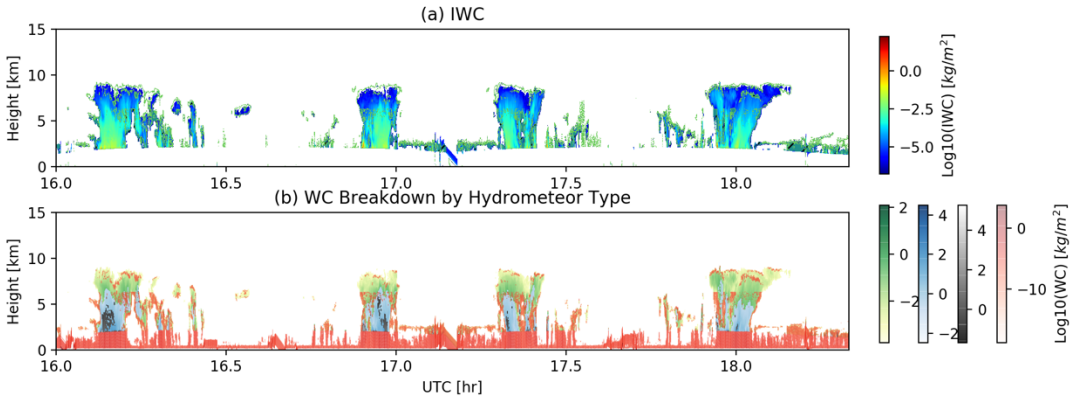


Fig. 5 (another option): same with the Fig. 5 in the revised manuscript, except the mixed-phase region where LWC and IWC co-exist are hatched in Panel (a).

l. 275: To me, this is not so clear. For W-band, the pure graupel assumption seems to improve the results, for Ka-band there is not much difference. Maybe quantify the difference by providing the average absolute difference values (for altitudes above the melting layer)?

This is a great suggestion. We've now added a 4th row to show the mean difference comparison as a function of height (above melting layer and below cloud top) for different bands and referring the mean difference panel in the text.

Fig. 9: Consider adding regression lines and R value to panel a to quantify the agreement.

Done.

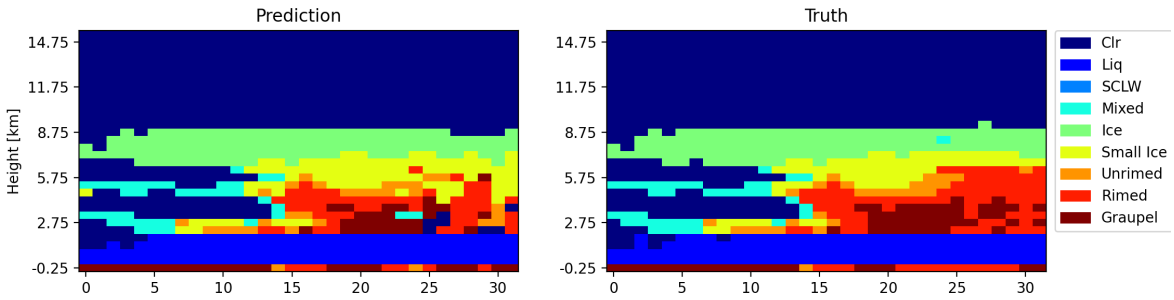
Fig. 10c: This panel is extremely hard to read and understand. The y labels are super small and in fact you only discuss the BT and PD for different channels. Instead of the (at least for me) confusing setup of panel c, you could simply display BT for all channels in a panel c and PD for all channels in a panel d. In addition, there is currently no color bar given. I assume red is high and blue is low, but please add this information.

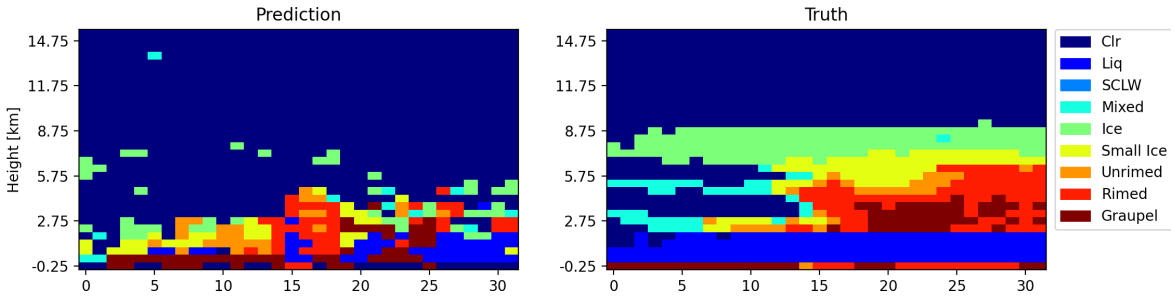
Fig. 10c is a standard way of displaying feature variation in a ML-ready dataset, and the color range is 0-1 as every line is linearly normalized by its min and max values, respectively (also standard ML-preparation). Therefore, bigger color contrast in one horizontal bar means that this specific feature has more variations, and if the variations corresponding well to cloud regimes, they are likely to be associated with higher weights in the training process. This is the reason for circling regime A, B, C for comparison with PD signals. We have now also increased the font size, and reiterated in more detail about feature names in the figure 10 caption.

Fig. 11 and 12: Should not the Full-deck reference and the No-PD reference be the same, since these are based on radar observations and do not apply PD? It seems like b and d are shifted in horizontal direction by a few bins. Why? Could you not show the same time range and then plot only panels a to c? In addition, the x label is missing. The same applies to both figures.

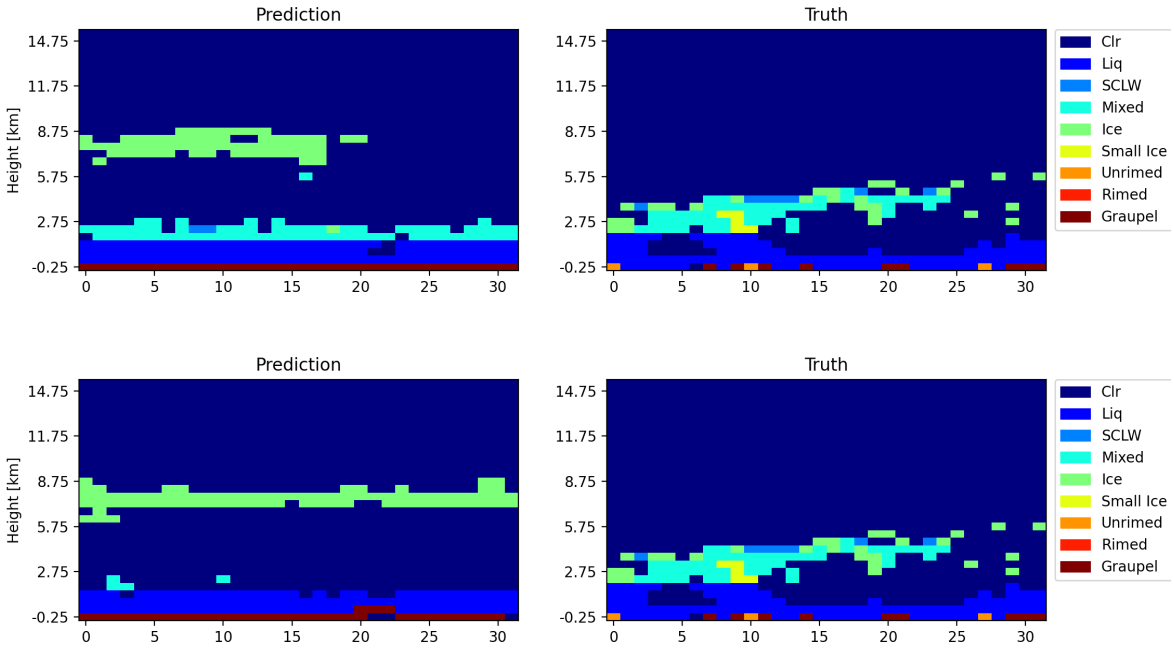
Same question was raised by the other reviewer. The independent testing dataset contains 432 images, and they are randomly grouped into 29 batches with each batch containing 16 images for prediction. Therefore, when we call the ML model each time to generate the prediction, the image order is randomized, and I had to visually identify the ones with the same "truths". 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.

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. X-label is not needed as it's just pixel number without any physical meaning.





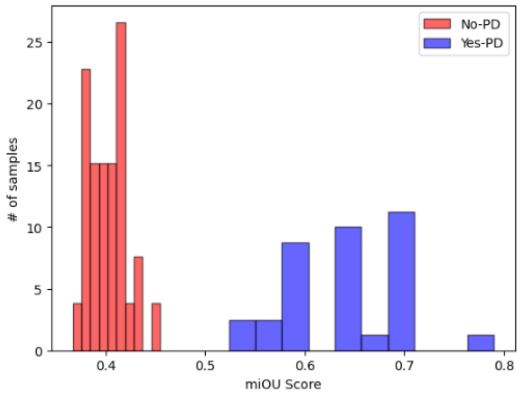
Modified Fig. 11. Top panel is from yes-PD fullset model prediction, while the bottom panel is from no-PD subset model prediction (in the revised version, panel #4 is removed due to redundancy).



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

l. 371: On the one hand, this is trivial due to the increased information content with PD. On the other hand, (part of) the improvement is likely also due to the choice of the same hyperparameters in both cases, see general comment above.

I don't think so. Excluding PD with the same hyperparameters in both cases serves as a standard ablation test to demonstrate the criticality of including PD. See the mIoU score statistic comparison below. We also identified a bug in computing the mean mIoU score for the no-PD cases. It should be 0.4024. The fullset mean mIoU score remains to be 0.642.



l. 372: Is there an explanation why there is a false cirrus cloud predicted?

No idea. We feel the artificial cirrus cloud tends to be predicted when the scene only contains low-level clouds. Our training dataset only comes from one day. Should we include multiple days the prediction might become better.

l. 380ff: This section was a bit hard to follow for me. Maybe add more explanation/motivation at the beginning. You basically compare effective diameters derived from CoSSIR with collocated in situ measurements. For the retrieval of the ice mean effective diameter you first apply PD measurements in form of "bell-curves" to choose a habit. Then you derive the diameter with a look-up table method using PD and TB measurements, taking the respective look-up table for the previously chosen habit.

That is correct. Consider this paragraph starting from Line 380 as a literature review of previous theoretical simulation works. The last sentence of this paragraph states the advancement/novelty of this work. The paragraph starting from Line 395 describes this 3-step retrieval procedure.

l. 414: In fact, you assess the influence of different vertical profiles of hydrometeor classes/habits in contrast to the assumption of a uniform habit. Vertical profiles in general could also include effects of the vertical resolution, or the accuracy of the retrieved IWC and Deff.

You are correct. Our description was not very accurate. Now it's replaced with "vertical profiles of hydrometeor phase and habits to reach agreement …"

l. 431: If Liu et al. (2026) provides a strict retrieval algorithm, why did you include Sect. 3.2 here as well? Would not a reference to Liu et al. be sufficient? Is Sect. 3.2 really new compared to Liu et al. (2026)? Unfortunately, this reference is not publicly available to compare.

Please see our response above in Page 2.

l. 444: This sentence sounds really technical and timely delivery of data products is not really the scope of ACP? How could the findings of the closure study, for example, help to improve retrievals?

You are correct. This sentence was not clearly crafted. The operational science product and algorithm are detailed in Liu et al. (2026). This paper elucidates some potential research products that are beyond mission requirement and are lower maturity to qualify for operational products. We may deliver some of these products upon available funding in the future.

l. 469: As far as I know the Jacobian and the weighting function are typically not the same. Please define more precisely what you show in Fig. A2.

"Weighting function" has now been changed to "channel response function (CRF)" whenever it is meant to say the Jacobian to IWC/LWC.

Technical corrections:

l. 76: "anfd…" something went wrong here

Corrected.

l. 129: Please add the date to the time.

Done.

l. 134: I think there is an "in" or similar missing here.

Added.

l. 154+157: "incoorporate" -> "incorporate"

Corrected.

l. 206: The abbreviation TB has not been introduced.

Added "brightness temperatures"

l. 306: WF has not been introduced so far.

Added.

Fig. 13: Color bar and y labels overlap and are therefore hardly readable.

Corrected. Now the colorbar doesn't overlap with letters.

l. 433: "?" The citation is missing.

Citation error corrected.

l. 444: “, This” -> “, this”

Corrected.

Fig. A1: The caption refers to panels a t

Corrected.