Review of Gong et al. (2026)

General remarks:

The work by Gong et al. applies active and passive remote sensing and in situ measurements from the IMPACTS campaign. A closure study between different measurements collected during the research flight on January 15 is presented. The radar-retrieved IWC is compared to IWC retrieved from collocated in situ observations. In addition, radar reflectivity is simulated with a radar forward operator using radar-derived IWC, LWC, and hydrometeor classification as input and is compared to measured radar reflectivity. Moreover, sub-mm radiometer measurements are compared to simulated brightness temperatures (BT), which are also computed from the radar retrieved hydrometeor profiles using a forward operator. The BT and radar reflectivity simulated with the actual vertical hydrometeor profiles are compared to simulations under the assumption of a uniform, single hydrometeor type, showing that realistic vertical profiles are needed for a closure.
Besides the closure study, the potential of polarimetric radiometer measurements for constraining cloud microphysics is shown. A hydrometeor classification for a limited number of vertical layers is derived using a machine learning approach including and excluding polarization difference (PD) measurements. In addition, it is demonstrated for the example case study how PD can first constrain the ice crystal habit and then be used together with the BT in a look-up table method to derive mean effective ice crystal diameters. The retrieved crystal sizes are compared to collocated in situ measurements.

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. 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.

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.

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.

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?

Specific comments:

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

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

l. 161: related to the general comment above: it would be helpful to make clear that the case study uses data from January 5. 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.

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?

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

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?

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.

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

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

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.

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)?

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

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. 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.

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.

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

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.

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 D_{eff}.

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.

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?

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.

Technical corrections:

l. 76: “an*f*d…” something went wrong here

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

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

l. 154+157: “incoorporate“ -> “incorporate”

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

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

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

l. 433: “?” The citation is missing.

l. 444: “, This” -> “, this”

Fig. A1: The caption refers to panels a to c, but there are only two panels and no labels at all.