

Retrieval of cloud thermodynamic phase partitioning from multi-angle polarimetric imaging of Arctic mixed-phase clouds
Weber et al. (2025)

Executive summary:

This work discusses a new algorithm for ice fraction derivation from multi-angle polarimetric cloud measurements from the specMACS instrument during the HALO-AC3 campaign. The retrieval combines this data, the IDEFAX neural network forward model defined in parallel work (Weber et al. 2025), 3D Monte Carlo radiative transfer simulations from the MYSTIC routine, WRF cloud simulations, and ERA5 re-analysis. The paper uses both intensity (I) and polarized (Q) multi-angle cloud data in two regimes: “slope”, or the region between 60-80 in scattering angle, and the “cloudbow” the region between 135 and 165 in scattering angle in the retrieval. The paper concludes that realistic Arctic clouds, simulated in 3D, compare best with retrieved ice fraction and cloud optical thickness (COT) over the specMACS field-of-view.

This paper is well-within the scope for AMT. It is valuable for current or upcoming polarimetric missions, such as PACE, 3MI, the polarimeter on CO2M, HACP, and the DPC/POSP series. It is also excellent that the authors are upfront about detection, modeling, and interpretation uncertainties. However, I ask for a potentially major and minor revision prior to publication.

Potentially major revision:

It is unclear how above-cloud-aerosol (ACA) impacts the derivation of ice fraction at cloud top. Because this retrieval relies on a fit to Q, aerosol loading may dampen the Q-signal like ice (Alexandrov et al. 2012, section 7, figure 9). To first order, aerosols will modify the depth of I as well.

Given cloud height in the Fig. 7 and 12 domains are ~1 km at most, ACA cannot be completely ruled out – though in the Arctic, AOD is likely low. However, AOD at 0.1 and lower can have an impact on I and the depth of the primary bow signal in Q, over clouds.

Therefore, retrieved ice fraction could be overestimated relative to cloud-only simulations in the presence of ACA for “saturated” pixels. “Unsaturated” pixels may be more complex. Aerosol has a darkening effect in I over clouds in the visible, which is opposite of increasing ice fraction/COT in the paper.

The interpretation of I and Q signals is important, because ice fraction here is quantitative value, not a qualitative phase index (Reidi et al. 2010, cited in-text).

The paper makes no mention of aerosol in modeling or simulation. If this has been considered, please discuss more clearly.

If not, I recommend the following:

(Most likely) Prove that the AOD in the specMACS scenes is negligible (or in other words, not a significant component of the multi-angle I or Q signals). Check the AOD from relevant satellite overpasses during HALO-AC3 or co-incident measurements from the aircraft (if those exist). If this is true (and likely is), also add discussion on how the algorithm could be adapted to address ACA impacts on ice fraction for non-clean scenes.

(Least likely) In the rare chance that AOD is not negligible, then this is a major revision. I suggest a rescope to include AOD as a retrievable parameter in the algorithm flow. To support this, show how a range of AOD impacts ice fraction retrieval with IDEFAX for $f_{\text{ice}} = 0.2$ for unsaturated and saturated cases (since the algorithms differ). Please demonstrate with a figure.

Alexandrov, M.D., B. Cairns, C. Emde, A.S. Ackerman, and B. van Diedenhoven, 2012: Accuracy assessments of cloud droplet size retrievals from polarized reflectance measurements by the research scanning polarimeter. Remote Sens. Environ., 125, 92-111, doi:10.1016/j.rse.2012.07.012.

Minor revision:

I appreciate the attention to detail in the paper, though the many study configurations can be hard to follow at times. It will be more impactful to the reader if the authors simplify the discussion and more concisely explain:

- The cloud measurement scenarios: unsaturated vs. saturated
- The retrievals: Q-based vs. I and Q-based
- The cloud modeling schemes: plane-parallel vs. IDEFAX
- The cloud interpretation: 1D vs. 3D
- Add more details on IDEFAX instead of referring the reader to Weber et al. (2025), add a table on Volkner et al. (2024) inputs to MYSTIC

In-line comments (many related to the minor revision):

104, 120, and elsewhere

“Observation of the cloudbow indicates the presence of liquid water and absence of the cloudbow a pure ice cloud.” (104)

“If the cloudbow is geometrically possible but not visible, 120 the cloud consists of pure ice and the ice fraction equals to 1.” (120)

See major revision above - the Q signal may appear as pure ice, but contain a mix of ACA and ice (in general). This can change the interpretation of ice fraction.

121

What does it mean for the polarization signal to be “saturated”? As in the top of the detector dynamic range? Or does that mean that the cloud has a COT > ~3 and therefore, “infinite” to a photon? Please explain in-text here.

I realized later on this definition is on line 215 - far too late into the paper. Please bring this up to an earlier section.

135

How does the Kolling et al. algorithm treat cloud sides/edges? A bit more discussion about this would be great.

159

How robust is the minimum checking on Q to instrument measurement noise?

176 (and following paragraph)

I am concerned that manual cloudbow labeling does not accurately represent the true uncertainty of the cloudbow detection, and confuses the interpretation of 3D effects and other errors in the applications later in the paper.

For example, the specMACS Q uncertainty between 3.5-6% given in Weber et al. (2024) could bury weak cloudbows in noise and add error in human interpretation. This could be where the 23.4% false detection metric is coming from.

There is evidence from this and other work (van Diedenhoven et al. (2012), and unpublished from Xu et al. on PACE/HARP2) that the ice/water detection is straightforward with multi-angle polarization statistics. As noted, the high 4% false positive metrics is likely human error as well.

Instead, I recommend a more statistical approach using Qual and RMSE metrics from Portge et al. (2023) to verify the cloudbow detection. Simple thresholds on both could differentiate real cloudbows from noise or false positives. Since the cloudbow Q fit is already part of the flow, aren't these metrics part of the calculation?

It is also valuable to have an extra category “unknown” for cloudbow cases that are ambiguous. There is a precedent for “unknown” in other cloud phase indices (esp. Reidi et al. 2010) and may clarify the results that pass RMSE minimization.

Pörtge, V., Kölling, T., Weber, A., Volkmer, L., Emde, C., Zinner, T., Forster, L., and Mayer, B.: High-spatial-resolution retrieval of cloud droplet size distribution from polarized observations of the cloudbow, *Atmos. Meas. Tech.*, 16, 645–667, <https://doi.org/10.5194/amt-16-645-2023>, 2023.

van Diedenhoven, B., A. M. Fridlind, A. S. Ackerman, and B. Cairns, 2012: Evaluation of Hydrometeor Phase and Ice Properties in Cloud-Resolving Model Simulations of Tropical Deep Convection Using Radiance and Polarization Measurements. *J. Atmos. Sci.*, 69, 3290–3314, <https://doi.org/10.1175/JAS-D-11-0314.1>.

Weber, A., Kölling, T., Pörtge, V., Baumgartner, A., Rammeloo, C., Zinner, T., and Mayer, B.: Polarization upgrade of specMACS: calibration and characterization of the 2D RGB polarization-resolving cameras, *Atmos. Meas. Tech.*, 17, 1419–1439, <https://doi.org/10.5194/amt-17-1419-2024>, 2024.

194

Also aerosol optical thickness (see major revision).

214

Add to the end “since ice clouds are brighter than liquid clouds, in our simulated cases.”

221 and elsewhere through the paper

All mentions of “reflectivity” should be “reflectance”.

220

Figures 1 and 3 show that the change in Q at different COD is nowhere near the same magnitude as the change in I, but ice fraction changes to Q happen almost independently to COD.

The consequence of a combined, equally weighted RMSE for I and Q in unsaturated cases is that the “winning solution” for ice fraction may overemphasize a good I comparison over Q, where the distinct information content is.

This may explain why biases in measured vs. modeled ice fraction persist in the Figure 10f histograms for in the cloudbow range retrieval - and also why the COD retrieval compares well on 11f.

I recommend considering an error-normalized metric instead, such as:

$$\chi = (1 - w_Q) \frac{I_{meas} - I_{model}}{RMSE_{I,meas-model}} + w_Q \frac{Q_{meas} - Q_{model}}{RMSE_{Q,meas-model}}, \quad (1)$$

where w_Q is an empirical weight on Q. This form allows Q to directly compensate for measurement-model differences in I. w_Q may be effective at 0.5, but may need fine tuning to emphasize the independent information content in Q relative to ice fraction.

250

Of the two cases shown in Figure 4, neither is labeled as “homogeneously mixed”. Do you mean “linearly distributed”?

Figure 8

The terms “cloudbow” and “slope” for the third column histograms were not immediately obvious. Please describe this more explicitly like:

“(c, f) Histogram of the differences between retrieved and model ice fractions with mean and standard deviation calculated from analysis performed in the cloudbow scattering angle range (blue) and forward scattered slope range (orange)”

And also please harmonize other figures that may have similar discussion.

350

I strongly suggest adding 2-panel figure that shows spatially, over the specMACS domain:

- The cloud pixels that correspond to the slope range retrieval, and which ones to the cloudbow range retrieval
- The cloud pixels that undergo the saturated retrieval (Q only) and which ones go through the unsaturated retrieval (I and Q).

I am curious if these distributions can help explain some of the spatial variation in the 3D study row of Figure 8 (d,e,f). This will also support discussion on errors (line 354 - 385).

Summary section

Given that the realistic 3D cloud simulations compare the best against specMACS data - of the four retrieval combinations: unsaturated slope, unsaturated cloudbow, saturated cloudbow, saturated slope - which are the most valuable and which are least effective? It is clear from Figure 13 that they may create different results and it would be excellent to summarize under what conditions they succeed and aren't as useful.