Reply to Reviewer #1's comments

This work addresses an important aspect of polar clouds based on the recent THERMODCLDPHASE VAP, elucidates aspects of its algorithm, and creatively investigates how to extend it using machine learning. I believe their product will be a useful contribution to the scientific community. My only main concern is the relatively minor role that lidar plays in the phase classifications and therefore the relatively poor skill that the models have in predicting liquid. I suggest that the authors delve further into how liquid phase predictions can be improved while removing the somewhat redundant parts of the manuscript describing the drop-out experiments described below. The results also seem to show that temperature plays a more dominant role than shown by the analysis methods. I would recommend publication of this manuscript after the authors consider the suggestions below.

Answer: We thank the reviewer for these suggestions and comments. We carefully revised the manuscript according to the reviewer's comments.

Clarifying the geographical scope of the work

- *It is unclear whether the scope of the THERMOCLDPHASE VAP is limited to Arctic (or polar) clouds. Different regions may require different tuning in the algorithm, and the authors have only focused on the Arctic (NSA and COMBLE regions) in this manuscript. If the goal is to test the generalization of the machine learning models to other regions, does this also include different cloud types? Is THERMOCLDPHASE suited for classifying the thermodynamic phase of other types of clouds? Line 57 mentions "several other ARM observatories across the world" but does not specify which. I recommend that the authors clarify this in the manuscript.*

Answer: We thank the reviewer for this valuable suggestion. We agree that algorithm tuning may need to be region-specific. As noted by Shupe (2006), the multi-sensor cloud thermodynamic phase classification method "has been specifically developed for observations of Arctic clouds." Accordingly, the THERMOCLDPHASE VAP is currently applied only at the seven ARM high-latitude observatories. We have revised the sentence in line 60 to read: "as well as six other ARM high-latitude observatories." Additionally, since the algorithm does not include the classification of hail and graupel, it has difficulties distinguishing these hydrometeor types in deep convective cloud regimes over tropical and mid-latitude regions. To improve clarity, we added the following sentence between lines 60 and 63: "It is noted that the multi-sensor cloud thermodynamic phase classification was specifically developed for observations of Arctic clouds (Shupe 2006). Since the algorithm does not include the classification of hail and graupel, it has difficulty distinguishing these hydrometeor types in deep convective cloud regimes over tropical and mid-latitude regions."

Clarifying the roles of temperature and lidar

- *The sharp cut-off along the 0˚C isotherm in Figure 1(g) where ice transitions to warm precipitation suggests to me that temperature plays a critical role in determining the phase of hydrometeors. Yet, it seems that the "feature importance analysis" and Figure 8 show that radar plays an even more important role than temperature. How do the authors reconcile this?*

Answer: It is true that temperature plays a critical role in determining hydrometeor phase within the transition region from ice to warm precipitation. However, this transition typically occurs within a relatively narrow vertical layer. Outside of this zone—particularly at temperatures above 0 °C or below –40 °C—temperature becomes less influential, and radar measurements provide more definitive information for hydrometeor phase classification.

- *Also, I am concerned about the minor role that lidar plays in the phase classification presumably due to lidar attenuation. The fact that the CNN performs the best out of the three models due to its accurate prediction of ice makes sense in light of the fact that radar and temperature play the most important roles in the prediction --- radar can better observe the ice particles that are larger in size, and ice crystals that freeze homogeneously are easier to identify. I suggest that the authors separately show cases that are dominated by single-layer thin liquid clouds to check whether lidar plays a more significant role and whether the models might also show high fidelity for the liquid classifications.*

Answer: We appreciate the reviewer's concern and thoughtful suggestions. As noted in previous studies (Shupe et al., 2011; Zhang et al., 2017), lidar observations at the NSA site are often fully attenuated above approximately 1 km due to persistent low-level clouds. This limitation contributes to the generally low feature importance of lidar data in thermodynamic phase classification across the full dataset.

However, for liquid-phase identification specifically, lidar backscatter shows notable importance in the CNN model. While lidar backscatter and depolarization ratio offer direct and reliable indicators of liquid-phase presence, radar-based variables—such as reflectivity, mean Doppler velocity, and spectral width—can also contain useful signatures of liquid-phase clouds (Luke et al., 2010; Yu et al., 2014; Kalogeras et al., 2021; Schimmel et al., 2022). This is consistent with our feature importance results presented in Figure 8.

We have added several sentences in line between 439 and 442 to further clarify this point. As suggested, a dedicated analysis of single-layer thin liquid clouds could help better assess the role of lidar in these specific conditions. Developing a refined training dataset focused on such cloud regimes and retraining the models would be a valuable extension of this work, and we consider this a promising direction for future research.

References:

Kalogeras, P., Battaglia, A., and Kollias, P.: Supercooled Liquid Water Detection Capabilities from Ka-Band Doppler Profiling Radars: Moment-Based Algorithm Formulation and Assessment, Remote Sens., 13, 2891, https://doi.org/10.3390/rs13152891, 2021.

Luke, E. P., Kollias, P., and Shupe, M. D.: Detection of supercooled liquid in mixed-phase clouds using radar Doppler spectra, J. Geophys. Res. Atmos., 115, D19201, https://doi.org/10.1029/2009JD012884, 2010.

Schimmel, W., Kalesse-Los, H., Maahn, M., Vogl, T., Foth, A., Garfias, P. S., and Seifert, P.: Identifying cloud droplets beyond lidar attenuation from vertically pointing cloud radar observations using artificial neural networks, Atmos. Meas. Tech., 15, 5343–5366, https://doi.org/10.5194/amt-15-5343-2022, 2022.

Shupe, M. D., Walden, V. P., Eloranta, E., Uttal, T., Campbell, J. R., Starkweather, S. M., and Shiobara, M.: Clouds at Arctic Atmospheric Observatories. Part I: Occurrence and Macrophysical Properties, J. Appl. Meteorol. Clim., 50, 626–644, https://doi.org/10.1175/2010JAMC2467.1, 2011.

Yu, G., Verlinde, J., Clothiaux, E. E., and Chen, Y.-S.: Mixed-phase cloud phase partitioning using millimeter wavelength cloud radar Doppler velocity spectra, J. Geophys. Res.-Atmos., 119, 7556–7576, https://doi.org/10.1002/2013JD021182, 2014.

Suggestions with regards to writing:

Abstract:

- *The Abstract does not mention what the results for COMBLE are.*

Answer: We added a sentence in line 26 to read: "The models demonstrated similar performance to that observed at the NSA site."

- *Similarly, the results of the ML models' response to simulated instrument outages and signal degradation are also not summarized in the Abstract.*

Answer: We added a sentence in line 27-28: "and show that CNN U-NET model with input channel dropouts during training performs better when input fields are missing".

Introduction:

- *Lines 31-33 require references. Suggestions: for ice particle production (via the WBF process ---Storlevmo & Tan 2015), precipitation formation (Mulmenstadt et al. 2015), the evolution of the cloud life cycle (Pithan et al. 2014), and also the response of clouds to global warming (Tan et al. 2025).*

Answer: We are grateful to the reviewer for suggesting these valuable references; they have been included in the revised manuscript.

- *Lines 36-38: satellite remote sensing could also be included here, e.g. MODIS cloud retrievals as detailed in Platnick et al. (2016).*

Answer: We agree with the reviewer that satellite-based remote sensing of clouds is important and have added the suggested reference accordingly.

Concerns regarding redundancy:

- *Section 4 essentially shows what the feature importance analysis did earlier in the manuscript regarding the importance of radar and the temperature soundings for the phase predictions. The authors might want to consider removing this section and replacing it with more detailed analysis on the limited role of lidar and how the classification of liquid pixels can be improved.*

Answer: We appreciate the reviewer's thoughtful suggestion. While we agree that Section 4 reinforces conclusions drawn from the earlier feature importance analysis, it also provides additional value by quantitatively assessing the impact of missing observational inputs on ML model performance. This section offers a more detailed examination of how model predictions degrade in the absence of specific sensors and evaluates the relative resilience of different ML models to missing data. We consider this analysis to be a key contribution of the study, particularly in the context of real-world applications where data gaps are common. Therefore, we have chosen to retain Section 4 in the manuscript.

Minor/typographical suggestions:

- *Please clarify what is meant by "pixel" and "voxel" and also be consistent with the terminology throughout.*

Answer: We changed "voxel" to "pixel" for consistency throughout the manuscript.

- *Line 41: "imagers" not "images"?*

Answer: We retained the term "images" since the context refers to using the captured data for identifying cloud phase. To enhance clarity, we added the word "captured" before "particle images."

- *Line 104: no dash necessary in "reads-in"*

Answer: We removed the dash for consistency and clarity.

- *Please consider including isotherms in panels (a) – (e) as well.*

Answer: We have updated the figure to include isotherms for these panels.

- *Line 158: what is the percentage of "unknown" pixels in the VAP?*

Answer: Based on one year of data from 2021 at the NSA site, 5.9% of cloud hydrometeors were classified as unknown. We added a sentence at line 168-169 to note this.

- *Table 1: Please define "clip".*

Answer: We added a note about the clip function in the caption of Table 1.

- *Lines 237-238: Why was the data limited to only 2018-2020 instead of the full record going back to 1998?*

Answer: The ARM THERMOCLDPHASE VAP was processed only for data collected after 2017 at the NSA site, focusing on more recent and reliable lidar and radar measurements (https://adc.arm.gov/discovery/#/results/s::thermocldphase). As discussed in Section 2.3, the existing training dataset is sufficiently large to support ML model development. While including additional data from the NSA site might offer slight improvements in model performance, it would significantly increase the training time.

- *Line 170: "imbalanced" in place of "imbalance"?*

Answer: We changed 'imbalance' to imbalanced'.

- *Line 459: CCN should be CNN*

Answer: We changed 'CCN' to 'CNN' in the manuscript.

- *Line 252: "In the" in front of "Future"?*

Answer: We added 'In the' in front of 'Future' as suggested.

- *Line 284: apostrophe after "models"*

Answer: We added an apostrophe after 'models' as suggested.

- *Figure 5: Setting the max of the y-axis to 80 may help enhance the visibility of the drizzle/liq_driz/rain categories.*

Answer: We appreciate the reviewer's suggestion and have adjusted the maximum value of the y-axis to 80 accordingly.