

Review of “The Prevalence of Arctic Multilayer Clouds and their Observed and Modelled Characteristics”

by Wallentin et al.

General comments:

In this study, the authors analyze Arctic multilayer clouds (MLCs) using the ICON model and compare their occurrence, as well as their microphysical and macrophysical properties, with observations from the MOSAiC campaign. To account for the effects of locally emitted ice-nucleating particles, they additionally implemented and evaluated an immersion freezing parameterization within the model. The comparison between model results and observations indicates that ICON generally captures the occurrence of MLCs, although liquid and ice water paths are substantially underestimated.

The manuscript is logically structured and well written. I have some concerns regarding the comparison between the model and observations, particularly with respect to the occurrence of MLC and the occurrence of seeding. Nevertheless, this manuscript merits publication provided that the following comments are addressed.

Specific comments:

- My major concern with this study concerns the intercomparison of MLC occurrence and cloud properties between the model and the observations. The observationally based occurrence is derived from radiosondes (RS) and further supplemented by cloud radar observations (RS+Radar), whereas MLC detection in the model is based on cloud mass. The authors clearly demonstrate the sensitivity of MLC occurrence to the chosen cloud mass threshold. Ultimately, they select a threshold of $10^{-9} \text{ kg kg}^{-1}$ for comparison with RS detections. However, the rationale for using RS-only detection rather than RS+Radar, which reduces the detection of spurious cloud layers, is unclear. Furthermore, it is not evident why the $10^{-9} \text{ kg kg}^{-1}$ threshold was chosen, given that a threshold of $10^{-8} \text{ kg kg}^{-1}$ appears to more closely match RS detections. Or is it because using a low threshold would ultimately lead to RS-like definition as only the saturation criterion is considered? In such a case, it would be important to have a consistent definition of saturation (see also bullet point 3). Similar concerns apply to the comparison of seeding occurrence (Tab. 1). One potential way to avoid the need to arbitrarily select a cloud mass threshold would be to employ a radar forward operator to generate radar reflectivity, enabling a more consistent comparison with the Achtert et al. (2025) detection algorithm.
- P7, L160: It took some time to realize that two distinct thresholds are used in this study: a cloud mass threshold and a seeding mass threshold, which share the same numerical values. It would be helpful to clearly distinguish these thresholds throughout the manuscript, for example, by using separate mathematical symbols, to avoid confusion. Additionally, it is unclear why the cloud mass threshold changes between sections ($10^{-6} \text{ kg kg}^{-1}$ in Section 5.1 vs. $10^{-9} \text{ kg kg}^{-1}$ for MLC detection). While exploring the sensitivity of MLC properties to this threshold is valuable, once a threshold is chosen, it would be advisable to report all other microphysical properties

(cloud droplet number concentration, ...) using the same threshold, unless there is a compelling reason not to do so.

- P7, L161-162: How do the authors decide with respect to which phase (liquid or ice) saturation is determined? Is this approach consistent with the method used to calculate saturation for cloud cover in ICON or with Achtert et al. (2025)?
- In this context, it would also be helpful to clarify whether the model employs a fractional or grid-scale cloud cover scheme, as this is not explicitly indicated in the model description.
- P9, L219-241: The reported values for mean cloud droplet number concentration (N_d) appear unusually high for the Arctic, even exceeding the number of observed cloud condensation nuclei that could potentially be activated (see Fig. 2). Could this be due to the mean being influenced by outliers, as seems to be the case for mean ice crystal number concentrations (see Fig. 3)? A similar concern applies to the reported cloud ice masses and number concentrations. I would suggest that reporting median values may provide a more robust representation of these quantities. Do the results differ when evaluating the medians instead of the means?
- P10, Fig. 3: In the caption, you state that values outside the interquartile range (IQR) are excluded, yet these values still appear to be included when calculating the means shown in the figure and subsequently reported in the manuscript. This also reinforces my earlier point: reporting median values would make the reported statistics less sensitive to outliers, potentially eliminating the need to filter out extreme values in the first place.
- P13, Fig. 5: As stated by the authors, liquid water content is not given in CloudNet if liquid-containing clouds have liquid-phased precipitation. I wonder how the median liquid water path has been derived for the model and for ShupeTurner. Were time steps with liquid-phase precipitation excluded from the comparison? If not, this may lead to a definition-inconsistent intercomparison, as the rainwater path is included in the model output and in Shupe–Turner, but not in CloudNet.
- P22, L463-464: Could you give more information about the physical pathway of this increase in geometrical cloud thickness?

Minor Remarks:

- P2, L39-41: While seeding can indeed initiate glaciation, neither riming nor secondary ice production can initiate it, since both processes require pre-existing cloud ice. I would therefore describe these processes as enhancing glaciation rather than initiating it. Similarly, the current phrasing suggests that the Wegener–Bergeron–Findeisen (WBF) process initiates glaciation, whereas it also primarily enhances glaciation once cloud ice is present. Consider rewording this part.
- P2, L50-51: Downwelling longwave radiation will only influence the lower cloud layer and not “each other”.
- P3, L72-73: “... ICON Global analysis...” Are you referring to the analysis step (0th timestep) of the global forecast here? If so, I wonder whether this analysis is produced every 3 hours, as you further down state that you employ boundary conditions with 3-hourly updates.
- P3, L74-75: Here, one might understand that radiosondes are used as the only observations during the data assimilation. I assume you refer to the fact that the

radiosonde observations during MOSAiC are assimilated, in addition to the standard global observations. Furthermore, are observations really nudged (which I consider some kind of Newtonian relaxation) or simply used during the data assimilation step?

- P7, L165-166: No need to repeat the conditions, as you are referring to them in the first part of the sentence
- P8, L188-189: "... a standard deviation of the mean". Do you mean that the standard deviation is the same magnitude as the mean?
- P20, L413-414: Or because you are in an updraft-limited regime. On this end, I assume that grid-scale vertical velocity is used for aerosol activation, which might be too low at kilometer-scale resolution for Arctic clouds, which might be turbulence-driven.
- P21, L448: "... during the aircraft campaign PS106 ..." Isn't PS106 a ship cruise?

References:

Achtert, P., Seelig, T., Wallentin, G., Ickes, L., Shupe, M. D., Hoose, C., and Tesche, M.: Occurrence of seeding multi-layer clouds in the Arctic from ground-based observations, EGU sphere, <https://doi.org/10.5194/egusphere-2025-3529>, [preprint], 2025.