Reply to referee #2

We thank Referee #2 for reviewing the manuscript and the valuable comments and suggestions which we address below. The responses to the referee comments are given in blue italic letters.

The authors investigate a marine cold-air outbreak event over the Norwegian Sea that was probed during a field campaign. They used multi-spectral retrievals to obtain several cloud properties and derive additional ones that indicate cloud condensate phase. Based on the quasi-Lagrangian flight pattern and variations in cloud-top heights, the authors translate observational products into vertical profiles as a function of downwind distance, which allow to infer cloud evolution. The authors infer that cloud-tops are dominated by liquid condensate, that ice production occurs at the coldest temperature, and that mixed-phase processes were largely absent. I think the authors need to better justify the tools and to better explain what assumptions were used to arrive at their conclusions. Given the many concerns listed below, I recommend returning the manuscript for resubmission.

Major concerns

Value of ice index – the index is introduced as a source of "qualitative information" (l. 106), but then largely used in this paper to make quantitative arguments, which later serve in a discussion to rule out certain processes. I'm highly skeptical that this index is valuable enough to rule out frozen hydrometeors. The authors should conduct thorough sensitivity tests (e.g., run ice index retrievals on forward-simulated signals using idealized profiles) that convince the reader that quantitative information can be gained here. The authors suggest to study phase inference using model data – I agree and think it should be done prior to this work. It would also be good to list any previous applications of ice index and introduce them in more detail. The above work will particularly useful when considering these results as benchmark for models (ll. 444-446) that may not easily mimic ice index for comparison.

The ice index introduced by Ehrlich et al. (2008) is a well-established method to determine information about the thermodynamic phase of a cloud based on spectral measurements. Detailed sensitivity studies have already been performed by Ehrlich et al. (2008) and Ehrlich et al. (2009). These studies showed that the ice index is very sensitive to cloud ice. Qualitative measure means here that no quantitative ice fraction in terms of a ratio of the IWC to the TWC or in terms of the optical thickness is obtained, but an accurate phase classification is possible and the absolute values give an indication whether a mixed-phase cloud is dominated by liquid water or ice. Based on radiative transfer simulations, it was shown by André Ehrlich that ice indices below the threshold value of 20 correspond to liquid water clouds. As all retrievals based on passive remote sensing, the retrieved ice index is of course also affected by 3D radiative effects, which introduce an additional uncertainty. This is, for example, reflected in the large variability of the observed ice indices and discussed in the new draft. The ice index has already been applied specifically to clouds during MCAOs and warm air intrusion by, e.g., Ruiz-Donoso et al. (2020). Besides the ice index, the polarimetric phase retrieval was also validated with synthetic data obtained from WRF simulations in Weber et al. (2025b).
We added more information about the retrieval to the text and further extended the discussion of the observed ice indices. Please see the latexdiff for all changes.

Value of ice fraction – the authors should provide more information here that answers the following questions: (1) how did the authors decide whether to retrieve ice optical depth for a pixel or not? (2) If a pixel is categorized as ice or liquid, does that mean all condensate is assumed to be of that phase? (3) In light of the strong mixed-phase nature of MCAOs, what are the uncertainties connected to ice fraction based on the decisions in (1) and (2) and can it be reliably used for phase portioning?

The retrieval of the ice fraction is based on multi-angle polarimetric imaging. The ice fraction is derived for all pixels respectively cloud targets, where the observation geometry allows to observe all necessary scattering angles. The cloud targets are identified by using the brightness-based cloud mask from Pörtge et al. (2023). The pixels are neither categorized as liquid water nor as ice, but a quantitative optical ice fraction, defined as the ratio of the ice optical thickness to the total cloud optical thickness under the assumption of a homogeneously mixed cloud, is derived for every pixel by fitting multi-angle polarization signals from a forward operator to measurements. For cloud targets where the polarization signal is saturated, the cloud optical thickness does not have to be derived and the polarization signal is directly sensitive to the ice fraction. Detailed information about the retrieval, including a validation based on synthetic data of low-level Arctic mixed-phase clouds with a complex 3D cloud geometry and a discussion of the retrieval assumptions and uncertainties, can be found in Weber et al. (2025b). The uncertainties of the retrieved ice fractions are large, but nevertheless the retrieval provides quantitative information about cloud thermodynamic phase partitioning for two-dimensional horizontal fields. Given the large uncertainties, the ice index and the ice fraction were both computed from spectral and polarimetric imaging and additionally compared to lidar and radar data. Passive as well as active remote sensing observations indicated the presence of a more liquid-dominated layer at cloud top, as reported by previous studies, and showed general agreement. In addition, we have also checked the in situ measurements and found general agreement about the thermodynamic phase between all measurements. However, in situ measurements were only taken at two constant flight altitudes such that no or only very limited vertical information is available.
We added more details also about this retrieval to the paper draft, please see the latexdiff as above.

A sense of contradiction – The authors introduce MCAOs as typically having a liquid cloud-top layer and then apply the Lensky/Rosenfeld method that used the range in cloud-tops and their temperature to obtain pseudo-profiles of ice fraction and index. If liquid layers are expected at every cloud-top, how come liquid tops are only found for the lower temperatures (i.e., the upper end of the cloud-top height range)?

The clouds during MCAOs initially form as liquid water clouds. Ice crystals are then formed through heterogeneous freezing of supercooled-liquid cloud droplets. These ice crystals grow and sediment downwards. An explanation for the more liquid-dominated layer at higher altitudes/lower temperatures could therefore be that the ice crystals have to reach a certain size to sediment and the clouds therefore have to be deep enough.
The specMACS (and VELOX) measurements provide two-dimensional, horizontal fields of the ice index/ice fraction/droplet effective radius and the corresponding heights respectively temperatures of the measurements. Some of the measurements correspond to larger heights, other measurements to lower heights, since the cloud pixels between the cloud center and the cloud edges as well as between different clouds differ in height. The pseudo-vertical profiles are obtained by combining these two-dimensional fields into a single vertical profile. In contrast to tropical convective clouds, the Rosenfeld and Lensky method is usually applied to, there are strong airmass transformations during MCAOs and the thermodynamic structure of the boundary shows a strong temporal and spatial evolution. For this reason, we sorted the measurements into different time bins along their temporal evolution. For every time interval, we see the more liquid-dominated cloud tops. If we would have combined all measurements into a single vertical profile, we would not see these clear differences at the cloud tops, as mentioned by you. We added this information: "Based on the backward airmass trajectories, the measurements were sorted into different time ranges above open ocean to account for the strong airmass transformations and the related temporal and spatial evolution of the thermodynamic structure of the boundary layer during MCAOs."
By using this method, it is assumed that the distribution of the considered quantity is horizontally homogeneous for a given height respectively temperature for a given time interval. As already mentioned in the discussion section, this assumption has only been proven for the droplet size whose vertical evolution is analyzed in Sect. 3.2 and not for the thermodynamic phase in Sect. 3.1. Therefore, we additionally analyzed the lidar and radar data, which were generally in agreement with the

passive remote sensing observations. In addition, our findings are in agreement with our theoretical understanding and with existing literature. Future studies should, of course, further investigate this assumption. However, performing a study similar to Lensky and Rosenfeld (2006) and including it into the paper draft is out of the scope of this study. For this, a model that realistically represents the cloud thermodynamic phase partitioning and in particular also its spatial distribution would be required, which is one of the deficits of current weather and climate models. We extended the discussion about this limitation.

Another sense of contradiction – in Fig. 4, the authors show increasing ice index and decreasing ice fraction closer to the cloud tops. While the authors argue there are penetration depth issues with the spectral retrieval, this should only result in vertical shift of the latter product, but still not explain the general discrepancy between both metrics. Since a liquid-dominated top is one of the main findings of the paper, the authors should investigate any disagreement that may cast doubt onto their finding.

Fig. 2 shows pseudo-vertical profiles of the ice index and ice fraction. The ice fraction shows a strong decrease at the coldest temperatures. This decrease is not always, but at least partly visible in the ice index with a small decrease at the top. These differences are, however, not a contradiction but can be explained by different retrieval sensitivities and penetration depths. The polarimetric retrieval is very sensitive to liquid water and can detect very small amounts of supercooled liquid water in ice clouds. In contrast, the ice index is very sensitive to small amounts of ice. In addition, both retrievals have different penetration depths. This is not an issue but allows us to gain information about the thermodynamic phase at two distinct altitudes. The polarimetric signal originates from higher altitudes than the spectral measurements. A detailed characterization of the signal location is provided in Weber et al. (2025b). Differences in the penetration depths do not only shift the profiles vertically. If the layer of increased supercooled liquid water is thinner than the penetration depth of the spectral signal, the ice fraction will be much more influenced by cloud liquid water than the spectral ice index. Therefore, the conclusion of a geometrically thin, more liquid-dominated layer at the cloud top as stated in the paper draft can be justified. This conclusion is also supported by the lidar and radar measurements shown in Fig. 4 and in agreement with previous studies. We extended the corresponding discussion to: "The spectral retrieval is sensitive to deeper altitudes within the cloud than the polarized retrieval. Therefore, the ice fraction from the polarized retrieval is more strongly affected by a thin, more liquid-dominated layer at the cloud top than the ice index from the spectral retrieval if the thickness of this layer is smaller than the penetration depths of the spectral signal. In addition, the polarized retrieval is very sensitive to liquid water, whereas the spectral retrieval is more sensitive to ice. This together can explain the observed differences."
Moreover, we always refer to this layer as a more liquid dominated layer, since the clouds still contain liquid water droplets also at lower altitudes. The observations correspond to the very initial phase of a MCAO. The clouds during MCAOs form initially as liquid water clouds. After some time, ice formation sets in and these crystals sediment downwards. This leads to increasing fractions of ice also at lower altitudes. However, we still see liquid water at the lower altitudes as these clouds are observed during the very initial phase of their evolution. We added this to the corresponding section and generally provided more explanations throughout the corresponding section. For all changes please see the latexdiff.

The convective nature of MCAOs – I'm worried the authors applied methods that require stratiform conditions. For instance, (1) passive retrievals work with plain-parallel assumptions and their use in convective-natured MCAOs (that are initially fairly broken and of smaller cell size – i.e., can we rely on all pixels?) introduced errors and (2) dropsondes may sample a mix of cloudy and clear areas in between, failing to provide a comprehensive picture and thereby affecting the parcel model application. The authors should make sure to quantify any such error sources and propagate them into Fig. 5.

The ice index and ice fraction are, of course, influenced by 3D radiative effects. This, however, applies to all retrievals based on passive remote sensing observations. The retrieval of the ice fraction is based on polarization which is dominated by single scattering and therefore less influenced by 3D radiative effects. In addition, the 3D cloud geometry is accounted for in the retrieval by applying a parameterization of 3D cloud geometry in the forward operator. Furthermore, the retrieval results were filtered and the most uncertain results with the largest RMSEs of the fits in the retrievals were excluded from the analyses. The remaining uncertainty is reflected in the larger variability and discussed in the new draft. The retrieval was validated based on synthetic data for a field of low-level Arctic mixed-phase clouds with a complex 3D geometry in Weber et al. (2025b) where a detailed discussion of uncertainties can be found.
The cloud droplet effective radii obtained from the cloudbow retrieval applied in Sect. 3.2 are only negligibly affected by 3D radiative effects. The retrieval uses observations of the angular shape of the cloudbow which is formed by single scattering on liquid cloud droplets and determines the cloud droplet size distribution by fitting polarized scattering phase functions according to Mie theory to the observations. Therefore, it does not rely on a plane-parallel assumption. The retrieval was validated using synthetic data for a field of shallow cumulus clouds with a complex 3D cloud geometry in Volkmer et al. (2024) and the cloudbow retrieval method is widely used in different groups.
The ice index is directly computed from spectral radiance measurements. These radiances are of course more strongly affected by 3D radiative effects compared to the polarimetric retrievals. For this reason, we also applied radar and lidar data and computed both, ice index and ice fraction. In addition, we did not analyze the small-scale variation of the ice index, which would be strongly affected by 3D effects, but performed a more statistical analysis such that 3D effects partly cancel out.
Concerning the dropsondes, the cloud fraction during this research flight was very high (see e.g. Fig. 3 in Weber et al. (2025a)) and there was almost no clear-sky. The dropsonde data was only used (quantitatively) to compute the condensation rate for the parcel model. Most other studies applying the parcel model assume a constant condensation rate using a typical temperature. Here, we applied the temperature and pressure measurements of the dropsondes to obtain as accurate results as possible, but in general, the condensation rate has only a small dependence on the temperature and pressure. In addition, the dropsondes were used to determine the cloud base height. For this, a combination of the relative humidity measurements of the dropsondes and the 3D cloud geometry obtained from the stereographic retrieval was applied and a large number of data points was used, such that outliers could be detected and excluded. Therefore, the influence of a broken cloud field on the derived cloud base heights is also negligible.
We added more information about the different retrievals and extended the discussion of uncertainties and limitations. For all changes, please see the latexdiff.

Abstract, discussion, and conclusions contain several statements that lack context and express a higher degree of certainty than shown (largely due to the above concerns). I provided a few examples below and think the authors should revisit similar statements throughout the paper:

- “pure liquid water clouds” (l. 422) – what is this finding based on and what are the chances that this information is incomplete?
  This finding is based on measurements of the ice index. Values below the threshold value of 20 can be associated with liquid water clouds, according to sensitivity studies based on synthetic data (Ehrlich et al., 2008, 2009). The ice index is very sensitive to small amounts of ice. Therefore, cloud (tops) with values below 20 can be considered as entirely liquid. We changed the sentence to: “The initially liquid water clouds transitioned to a mixed-phase regime during the first approximately 30 min the airmass spent above open ocean, according to the specMACS observations, as also discussed in Weber et al. (2025a).”
- Absence or irrelevance of collision-coalescence (ll. 437-438, also ll. 388-391) – this appears to be based on a single study, that rules out small droplets from riming. While larger drops have

a greater riming efficiency, there should still be a chance for smaller drops to participate in riming, too (e.g., Saleeby and Cotton, 2008).
Collision-coalescence throughout the paper draft refers to collisions of two liquid cloud droplets whereas riming involves a liquid cloud droplet and an ice crystal. Our conclusion about riming is based on two studies (Avila et al. (2009) and Wang et al. (2002)) about the collision efficiencies and the impact of riming on the droplet size distribution. In addition, it is based on radar and in situ observations of riming on the same day in the same region by Schirmacher et al. (2024) and Maherndl et al. (2024). Both observed riming, in agreement with your statement and additional reference, but the rime masses were very small. We added the additional reference suggested by you and tried to make the distinction between collision-coalescence of liquid water droplets and riming clearer throughout the paper.

Thank you very much for this comment. We have tried to be careful with too strong and general statements to not express a higher degree of certainty than justified and usually wrote "The measurements indicate" or "suggest". We revisited and adjusted all statements throughout the paper draft, please see the latexdiff.

Minor concerns

l. 35 Please add a reference here. It is my understanding that relatively quiescent Arctic clouds have such liquid layer on top and would not expect this for marine cold-air outbreaks.

We added references as suggested.

ll. 36-38 Is the WBF process typically dominant? I think past studies have mostly hinted at riming.

We added riming as a second process here.

ll. 38-39 Can MCAOs be treated like other Arctic clouds? I'm also not sure why "However," was used here.

The "However" refers to the persistence of Arctic mixed-phase clouds despite the WBF mechanism and riming, which can rapidly deplete cloud liquid water. With the changed sentence in response to the previous comment, this should be clearer.

ll. 208-209 It's good to be more specific here to clearly show the reader what the authors are referring to.

We extended the corresponding discussion to: "At later times, the ice index is still almost constant with temperature except for the coldest temperatures, where it increases with decreasing temperature and crosses the threshold value from the liquid water into the mixed-phase regime. This increase in the ice index starting from the coldest temperatures indicates that ice formation occurs preferentially at the coldest temperatures. This is reasonable as several studies based on in situ measurements have shown that the relative fraction of ice increases with decreasing temperature and the probability for freezing of supercooled liquid cloud droplets increases with decreasing temperatures (Korolev et al., 2017)."

ll. 219-220 I think the authors should turn such information into error bars in Fig. 2.

This does not have to do with uncertainties which could be turned into error bars but is due to different retrieval sensitivities and penetration depths, see our answer to major point four above. Both different sensitivities and penetration depths are explanations for the observations but they are *not classical uncertainties which could be added as error bars. We extended the discussion about the retrievals and their sensitivities, as discussed above, which should make this clearer.*

ll. 224-225 I think the authors should include this figure into the supporting information.

We added a supplement with the additional figure.

ll. 250ff. It's not clear whether mixed-phase pixels were excluded here and, if not excluded, how the authors dealt with them.

Mixed-phase pixels were not excluded in the analyses. The effective radius of liquid cloud droplets was derived using the multi-angle polarimetric cloudbow retrieval, which is based on single scattering on liquid cloud droplets. We performed sensitivity studies which showed that the cloudbow retrieval is not affected by the presence of ice and that the results are reliable for ice fractions up to 0.8. We added this information to Sect. 2.1.

Fig. 2 – is not clear where the great temperature range stems from? Do cloud-top temperatures at "0-15 min" really span -25 to -10 degC? Looking at Fig. 3, temperature between 250 and 500 m cloud-tops should span -23 to -20 degC. In fact, Fig. 3 shows that -25 degC is not measured at all in the lowest 2 km (i.e., the first 300+ min according to Fig. 4).

Thank you very much for this comment. The temperatures in Fig. 2 are brightness temperatures measured by the thermal imager VELOX on board HALO whereas Fig. 3 displays dropsonde measurements. The influence of the warm ocean for pixels with optically thin clouds at lower altitudes/warmer temperatures can lead to an overestimation of the observed brightness temperatures and is therefore an explanation for the observed larger temperature range. We have already applied a cloud mask to the data and tried to further improve the filtering. The number of data points at these warmer temperatures is comparably small, but hard to filter. This is unfortunately not directly visible from the histograms, since these were normalized for every temperature bin by the total number of measurements of the respective bin to focus on the actual evolution instead of the number of measurements. In addition, the influence of the surface increases with decreasing cloud optical thickness towards to cloud edges at lower altitudes and warmer temperatures. This explains the "stretching" of the temperature range. However, the overall structure and the main results are not changed. Furthermore, the pseudo-vertical profiles as a function of height did show similar features, as discussed in the text, and the cloud top height measurements from the stereographic retrieval are not affected by the ocean surface. We added a discussion about this to the text:
"The brightness temperatures measured by VELOX in Fig. 2 partly show warmer temperatures and a larger temperature range than the measurements of the dropsondes in Fig. 3. This can be explained by the influence of the warm ocean surface, which can lead to an overestimation of the measured brightness temperature for optically thin clouds. A cloud mask was applied to the data to filter the measurements, but some influence persisted in the warmer parts of the profiles. The number of affected data points is, however, small. The histograms in Fig. 2 are normalized by the total number of measurements for every temperature bin, leading to a false impression of many affected measurements."

Fig. 5 – The authors should use the uncertainties in their discussion (ll. 316-325) to create error bars here.

Thank you very much for this suggestion. We tried to add error bars to the figure. However, the combination of the 2D histogram and the lines with the average measurements and theoretical profiles with additional color-shaded error bars or lines made the figure in our opinion more *confusing. For this reason, we kept the figure as is. The shading of the histogram already reflects the variability of the measurements and an uncertainty of 25 or 50% can easily be imagined.*

All figures – it would be good to list the various observational products used in each Figure's caption. For example, where did the droplet number concentrations in Fig. 6 stem from?

The caption of Fig. 6 and most other figures already included this information ("Cloud droplet number concentration … from in situ measurements."), but we added more information where missing.

References

Saleeby, S. M. and Cotton, W. R.: A Binned Approach to Cloud-Droplet Riming Implemented in a Bulk Microphysics Model, J. Appl. Meteor. Clim., 47, 694–703, https://doi.org/10.1175/2007JAMC1664.1, 2008.