Response to Referee #1

September 26, 2025

We thank the reviewer for their thoughtful feedback and for raising several important topics. Each point is addressed below.

Specific comments

In the Introduction, the authors provide the rationale for such a method and for retrieving the vertical profile of ice water content instead of estimating the ice water path in the column. One of the arguments they present is the radiative effects of cloud ice, and I agree with this in general. It is clear that the same mass of ice can be distributed within the cloud limits in a number of ways, and the radiative transfer and radiative effects for these distributions will not be the same. For example, the emission depends on temperature, and a cloud with top-to-bottom IWC(z) falloff will not be equivalent to a bottom-to-top IWC(z) falloff, despite the fact that their IWPs are the same. However, it has already been shown using the same DARDAR dataset and DISORT calculations that the absolute differences for short-wave and long-wave fluxes estimated with and without knowledge of IWC(z) shape do not exceed 2 W/m2 at the top of the atmosphere, 2.7 W/m2 at the surface, and 4 W/m2 in the atmosphere. If these results are cloud amount weighted, these values reduce to 0.5 W/m2, 0.5 W/m2, and 1 W/m2, respectively. From this point of view, it would be useful to provide an example of a real physical situation for which the error of using constant IWC instead of a real IWC(z) profile would lead to misinterpretation of a physical phenomenon or model validation.

- The results mentioned are new to us, and so we thank the reviewer for bringing them to our attention; they are certainly interesting and relevant to our study.

  However, it is not clear to us from the results cited whether the small net flux differences are independent of the cloud height, or whether they were obtained only by comparing to a constant IWC for clouds placed at the same height. We assume that the results would vary depending on cloud-top and cloud-base heights, since the longwave emission will depend on temperature. If this is the case, this information is still important. Unlike passive thermal infrared measurements, ICI's radiances are not directly related to the physical cloud top temperature, and must be retrieved. Therefore, retrievals of IWC remain important because they provide estimates of the cloud-top and -base heights, which cannot be inferred otherwise.

  Additionally, knowledge of the IWC profile is needed to constrain radiative heating rates inside the cloud. Hartmann and Berry (2017) shows that TOA radiation balance is similar for two anvil clouds with the same IWP but different shapes, which is consistent with the findings that the reviewer reports. However, heating rates displayed distinctly different behaviour for the two cloud types. Likewise, **?** finds that while surface and TOA fluxes are relatively insensitive to changes in IWC or particle size, the averaging heating rate can vary by 10-20% between 8 and 14 km. Radiative heating is a driver of vertical motion, and impacts convective cloud systems and large-scale circulation. As such, knowledge of the real IWC(z) profile is needed for its accurate calculation.

- **Changes to the manuscript:** The following sentences have been added to the manuscript: "The vertical distribution of ice has also been shown to impact radiative heating rates (Mather et al., 2007; Hartmann and Berry, 2017), which drive convective systems and large-scale circulation.", "Unlike passive thermal infrared measurements, microwave radiances are not directly related to the physical cloud top temperature, which is needed for long-wave flux estimates, and must be estimated from retrieved vertical information."

Section 3.2. Retrieval model implementation
I am not an expert in neural network training, but my experience with forward and inverse problems tells me that adding noise to the simulated radiance increases the chances of retrieving an incorrect original profile, especially in the case of an ill-posed problem. Indeed, this is a typical self-consistency test in any method, for which the input data are passed through the forward simulator, then modified by realistic noise, and then passed through the retrieval procedure to compare with the reference data. However, I'm not sure that the training dataset should be modified by noise. It's true that the real data will be noisy, but the training process using noise-free data should yield similar weights to the neural network's nodes and paths as a noise-perturbed one, but this training will take less time/data and be more physical. Later on, its accuracy will be reduced by using noisy data, but the neural network itself will be "cleaner". Could you please comment on this? Will one still require 9.4 million cases to train the neural network (line 216), or can one achieve the same results using 10 times fewer profiles, but without noise?

- To clarify: we did not add noise to the simulations beforehand, but rather generated a new noise value each time a training sample passed through the network. In this way, although the same training sample is seen by the model multiple times during training, each time there is a slight variation, and therefore this acts as data augmentation. Seeing slightly different variants of the same base sample during training helps the model to generalise better to unseen, noisy inputs.

  It may sound counter-intuitive to train on noisy inputs, but our choices are motivated by numerous studies showing that repeatedly adding new noise *during* training improves the robustness of the model and prevents overfitting (Zur et al., 2009; Piotrowski and Napiorkowski, 2013).

  If it were the case that, as you suggest, fewer cases would be needed to train the model in the absence of noise, this does not replace the need to cover the variability of the state space. The requirement for millions of cases is also motivated by the need to cover all possible conditions, e.g. very high-IWC cases or specific surface types. For this reason, we would still expect an equally high number of simulations to be necessary in order to fully capture real-world variability.

- **Changes to the manuscript:** To clarify our approach, we have added a short explanation in Section 3.2 describing how noise is added during training, and motivating its use as a form of data augmentation.

Lines 199-200: Indeed, the retrieval of Dm and Zm does not make sense if ice water path equals zero, but this is somewhat evident. Could you please rephrase these sentences?

- **Changes to the manuscript:** The sentence has been updated.

5.1. Retrieval Ice Water Content
In this section, the authors spend considerable time explaining the effects related to averaging kernels, but they do not mention them explicitly, despite the fact that they show them in Fig. 12 and Fig. 13. I would say that the text of this section could be made much more compact and understandable for the reader if the authors moved these figures here.

- We thank the reviewer for this suggestion. Our intention with Section 5.1 is to present a broad overview of retrieval performance that is accessible to a range of readers, while Section 5.4 acts as a more in-depth exploration on why we see the results presented earlier. While it is true that Figs. 12 and 13 would help to explain some of the results in Section 5.1, moving these figures earlier would pre-empt the derivations, motivations, and assumptions made in Section 5.4, and may risk confusing readers who are less familiar with averaging kernels (particularly since these are *approximations* of averaging kernels). Alternatively, moving the entire section earlier would risk breaking the logical progression from *what* the retrieval achieves to *why* it behaves that way. That said, we agree that it may help some readers to see the connection sooner.

- **Changes to the manuscript:** In several places in Section 5.1 we now note that some retrieval behaviours are consistent with the averaging kernel results, and refer the reader to Section 5.4.

Fig. 3, 7, 9, 10: It would be interesting to see the differences between the reference and test panels either in absolute or relative values in a third (added) panel. I am somewhat concerned about the striping mentioned in this section. Wouldn't it be better to smooth/denoise the input data to avoid this effect? Perhaps one could run the retrieval twice – once for original profiles and once for smoother ones – and if the results differ strongly, then use the second solution.

- Smoothing the input data is unlikely to have an effect on the striping artefacts, since they arise from observation noise rather than structures in the input data. Because our retrievals treat neighbouring profiles independently, and because the measurement noise is uncorrelated between observations, adjacent retrievals can fluctuate slightly up or down, producing the visual striping. As evidence for this, we have added a new panel to Figs. 3 and 9 to show the same retrievals, but performed using noise-free radiances as input. The striping disappears.

  Although the striping is only a noise artefact, we agree that it is undesirable. We thank the reviewer for their suggestion of smoothing, since this could still be an effective approach post-retrieval. However, the trade-off—-whether smoothing were to be applied pre-training or post-retrieval—-is a deterioration of the horizontal resolution. Since the horizontal resolution of ICI is not particularly high to begin with, smoothing would lead to a relatively poor resolution. Alternatively, a future avenue for this study would be to perform more advanced retrievals of the entire swath, which would likely also remove the striping.

  Our aim in this study is to present the retrieval capabilities of ICI and to identify where it succeeds and where it struggles. As such, we prefer to present unsmoothed retrievals, rather than 'hide' its occurrence from the reader. If an operational product were to be produced, we believe that it would be best left to the users to smooth the data, since this would be straightforward to do. In this case, the users can decide which has a higher priority based on the use case — lower noise/less striping or higher spatial resolution.

- **Changes to the manuscript:** A panel has been added to Figs. 3, 7, and 9, showing the difference between the retrieval and the truth, relative to the truth. The discussion has been updated to briefly discuss these panels.

  Additionally, Figs. 3, 7, and 9 have a new panel showing the same retrievals but performed on noise-free input data, in order to more clearly show the impact of the noise on the retrievals and thus explain the striping effect.

  Multiple new panels have now been added to three of the figures, due to both this suggestion and the suggestions of the other reviewer. To avoid the manuscript becoming too lengthy as a result, we have decided to remove Fig. 10 from the manuscript. Fig. 10 did not add any new information that Fig. 9 did not already provide.

References

Hartmann, D. L. and Berry, S. E.: The balanced radiative effect of tropical anvil clouds, Journal of Geophysical Research: Atmospheres, 122, 5003–5020, https://doi.org/10.1002/2017JD026460, 2017.

Mather, J. H., McFarlane, S. A., Miller, M. A., and Johnson, K. L.: Cloud properties and associated radiative heating rates in the tropical western Pacific, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2006JD007555, 2007.

Piotrowski, A. P. and Napiorkowski, J. J.: A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling, Journal of Hydrology, 476, 97–111, https://doi.org/10.1016/j.jhydrol.2012.10.019, 2013.

Zur, R. M., Jiang, Y., Pesce, L. L., and Drukker, K.: Noise injection for training artificial neural networks: A comparison with weight decay and early stopping, Medical Physics, 36, 4810–4818, https://doi.org/10.1118/1.3213517, 2009.