

Overview

This manuscript had conflicting recommendations from the Referees - one recommended rejection, and another recommended major revisions. I agree with both Referees in that there are substantial issues with aspects of the methodology and interpretation of results, and the manuscript is unsuitable for publication in the original submission. However, the authors appear to have made a good-faith effort to address some of these comments in the response, and the methodology itself (use of CNN for atmosphere profile retrieval) is novel and worthy of publication if these issues can be addressed.

Response:

We sincerely thank the Editor and Reviewers for their valuable and constructive comments. We have carefully considered each suggestion and provide detailed responses below.

In particular, I would like the authors to justify their choice of relative humidity (RH) as the water vapor variable instead of specific humidity. I don't agree with the justification that RH is a standard field in ERA5; so is Qv. Also, this may impact the interpretation of results regarding the information content in the 118 and 183 GHz channels.

Response:

We acknowledge the Reviewer's concern. At the initial stage of our study, we selected RH mainly because it is a commonly available standard variable in most reanalysis products (including ERA5), and it aligns with traditional retrieval outputs used in meteorological analysis. Additionally, RH has a relatively smaller numerical range and more stable physical characteristics, making it technically easier and more stable to train in deep learning frameworks. This choice helped improve convergence and reduce numerical instability during model training. However, we admit that we did not fully explore the potential differences in information content between RH and Qv at that time. We appreciate the Reviewer highlighting this important point, and we plan to consider specific humidity in future work to provide a more physically consistent evaluation.

While the Jacobian analysis is analytically correct, it is inverted from the typical approach, where the physical parameter is perturbed and the change in brightness temperatures in multiple channels is found. Even though the CNN provides an inverse model, it may not be realistic to perturb a single channel at a time. Rather, I would suggest that a data denial experiment is more appropriate for evaluating information content in deep learning models: remove channels individually from the architecture and identify at what levels and parameters the rmse increases the most.

Response:

We fully agree with the Reviewer that channel-wise data denial (ablation) experiments are a more robust way to evaluate the contribution of each channel in deep learning frameworks. Given the prohibitively time-consuming nature of model training, we regrettably constrained our ablation

experiments to channels 20 (118.7503 ± 2.1 GHz) - 26 (183.31 ± 7 GHz), acknowledging that a more comprehensive assessment across all channels would be ideal. Nonetheless, we conducted dedicated channel ablation experiments as advised, and the results have been clearly presented and discussed in the revised manuscript (**Lines 223-230** for temperature retrieval, with corresponding **Figure 7**; and **Lines 301-312** for relative humidity retrieval, with corresponding **Figure 12**).

The bias difference heatmaps demonstrate that removing channels with smaller frequency offsets (e.g., 118.7503 ± 1.2 GHz and 183.31 ± 2 GHz) leads to notable changes in higher atmospheric layers, while channels with larger offsets exert stronger impacts on lower atmospheric layers. These findings are generally consistent with the established understanding of spectral contributions from different channels.

We acknowledge that the vertical response patterns in our experiments are not perfectly aligned with those from traditional physically constrained retrieval frameworks. This difference likely results from the lack of explicit physical constraints in our current CNN architecture. We humbly recognize this limitation. In future work, we plan to introduce physics-informed network designs and dynamic channel-wise weighting to better separate and capture information across different atmospheric layers, thereby improving interpretability and physical consistency.

In summary, while our current experiments support the general trends reported in the literature, we recognize that further refinement is needed to achieve more physically interpretable and robust results. We sincerely thank the Reviewer for this insightful suggestion, which provides clear and valuable guidance for our future research directions.