Reviewer #1

Overview

The manuscript "All-Sky Temperature and Humidity Retrieval from the MWRI-RM Onboard the FY-3G Satellite" by Minghua Liu et al. introduces a convolutional neural network algorithm for retrieving vertical profiles of temperature and relative humidity from this instrument, and includes sensitivity analysis for the products. The new dataset is interesting, the CNN methodology has potential, and this reviewer appreciates the authors' interest in performing Jacobian and sensitivity analysis to explain the behavior of the algorithm.

The retrieval apparently uses a patch of 5x5 spatial footprints from the instrument as input, and the output is a single retrieval at the center of this patch. The training target is also, apparently, a single ERA5 profile interpolated via nearest neighbor to the center footprint of this patch. In that case, what is the basis for using spatial convolution in this approach? Is there any benefit to using spatial convolution as opposed to just simply using the nearest footprint, or perhaps just spatially averaging the input patch? Also: It was unclear what the spatial resolution of the retrieval product is. In CNN terms, what is the stride used to make the product? I.e., does each 5x5 input patch contribute to only a single output profile at the center of the patch, or are there overlapping patches contributing to neighboring retrievals?

Response:

We sincerely thank the reviewer for these insightful comments. In this work, we adopted a 5×5 spatial patch as input because it provides a balanced compromise between capturing meaningful local spatial variability and keeping the model computationally efficient. Following the approach proposed by Malmgren-Hansen et al. (2019), our retrieval uses a 5×5 spatial patch centered on each footprint. In our study, the remapped input data have a spatial resolution of approximately 13 × 21 km **(lines 112-114)**, and we found that this grid size allows the model to exploit complementary information from neighboring footprints while maintaining a manageable training cost. The rationale for this design has now been explicitly stated in the revised manuscript **(lines 119-122)**.

Although the retrieval target corresponds to the center footprint, the use of convolutional operations is motivated by the need to capture spatially correlated structures that affect the radiances observed at that point. Compared with simply using the nearest footprint or applying spatial averaging, convolution enables the network to learn data-driven filters that enhance signal extraction and suppress spatially correlated noise. This design thus improves the robustness and smoothness of the retrieval, in line with the improvements reported by Malmgren-Hansen et al. (2019) for similar infrared sounding applications.

Regarding the reviewer's question on the stride and resolution, our convolutional layers are implemented with a stride of 1 and padding of 1, so that neighboring input patches overlap and the retrieved profiles remain spatially consistent. Each 5×5 patch contributes primarily to the retrieval at its central footprint, and the final retrieval product maintains the same nominal spatial resolution as the remapped input (approximately 13×21 km). The training results using this configuration met our performance expectations, so we did not further explore the effect of different patch sizes in this study. We acknowledge that this aspect could be investigated in more detail in future work to further assess the sensitivity of the retrieval to spatial context.

Also: Are any non-radiance inputs like zenith angle used as input (common practice in NN retrievals)? If not, then the spatial features encompassed within the patch (which the authors are aiming to exploit with a CNN) will be angle-dependent with no inputs to indicate this to the model.

Response:

We thank the reviewer for this thoughtful question. In the current study, we did not include non-radiance inputs such as viewing zenith angle. Our goal was primarily to investigate the relative contribution of different spectral channels to the retrieval performance, and the model already achieved satisfactory results without additional angular information. As demonstrated in previous studies (e.g., Xu et al., 2024), incorporating observation angle can further improve the retrieval accuracy by providing geometric context to the radiance measurements. In future work, we plan to extend our framework to include such auxiliary parameters to evaluate their potential benefits in combination with the spatial convolutional design.

There are few implementation details. E.g., was Pytorch used? What optimizer was used to train the algorithm? What are the kernel sizes used in the model? How were hyperparameters chosen?

Response:

We thank the reviewer for the helpful suggestion regarding implementation details. In the revised manuscript **(lines 132-145)**, we have added a more detailed description of the model configuration and training setup. Specifically, the retrieval network was implemented using the PyTorch framework and designed as an Advanced Residual Convolutional Neural Network (AR-CNN) to process 5×5 satellite brightness temperature patches with 26 input channels. Convolutional layers employ 3×3 kernels with a stride of 1 and padding of 1 to preserve spatial dimensions. The architecture integrates residual blocks, batch normalization, dropout (rate = 0.25), and fully connected layers, with Sigmoid Linear Unit (SiLU) activation functions to ensure smooth gradients. The AdamW optimizer with a learning rate of 1×10^{-5} and mean-squared error (MSE) loss function were used for training, with early stopping applied to improve generalization. These additions now clarify the model structure, hyperparameter configuration, and training procedure as requested.

More information is needed in 2.2 "Sample Construction" is unclear. Were the selected ocean samples thinned out or curated in any particular way? How was the chronological test/validation/training split performed (interleaved or consecutive)? Why was that month of data selected? 2.2 mentions "...preprocessed static parameters obtained through decoding and quality checks..." but these parameters are not defined, nor is their usage.

Response:

We appreciate the reviewer's constructive comments and have revised **Section 2.2** accordingly **(lines 109-125)** to provide a clearer explanation of the data preprocessing and sample construction process. The updated text now clarifies that MWRI-RM Level 1 brightness temperature data from October 23 to November 23, 2023 (UTC) were used in this study. This one-month dataset was selected primarily because it corresponds to the data period already processed in the BGI algorithm framework (Chen et al., 2024), which allowed us to reduce redundant data preprocessing efforts while ensuring consistency in spatial remapping. In addition, this period exhibits stable instrument performance, providing reliable data quality, and its size represents a practical balance between sample diversity and computational efficiency—using a larger temporal span would significantly increase the training time without notably improving model generalization.

As detailed in the revised text, samples were curated by excluding scan points with more than 7 km deviation between the two feedhorn rows (S1 and S2) to ensure high-quality inputs **(lines 110-112)**. Oceanic samples were then exclusively selected using the built-in land-sea mask **(line 112).** The dataset was chronologically split into 80% for training and 20% for testing, with the final 20% of the training portion used for validation, ensuring temporal independence among all subsets **(lines 123-125)**.

The Jacobian methodology needs more elaboration. The plots such as Figure 5b are weighting functions computed from a radiative transfer model. The description of "gradient backpropagation" in the manuscript, apparently, therefore means automatic differentiation of the radiative transfer model outputs with respect to test profile inputs, not, say, differentiation of the CNN model. Is this true? The sensitivity of the CNN model itself is only shown as the conv1 weights. This is not sufficient to provide the desired explainability of the retrieval methodology. With a neural network model, it is possible, and straightforward to use "autograd" in tools such as PyTorch to compute the Jacobian of the whole model's output with respect to its inputs. This would actually describe the channel usage by the model, which appears to be the desired goal.

Response:

We thank the reviewer for this helpful suggestion and apologize for the earlier ambiguity. In the revised manuscript, we explicitly distinguish between RTTOV weighting functions **(Fig. 5(b))** and model-side Jacobians obtained via PyTorch autograd. As detailed in **lines 213-224**, we de-normalize outputs, define a scalar objective (sum over temperature levels), compute input gradients with autograd, aggregate absolute gradients over batch and 5×5 spatial indices, and normalize across channels to obtain per-channel attributions **(Eqs. (3)-(4))**. Using this approach, channel importance peaks near 118.75 ± 1.2 GHz **(Fig. 5(a))**, consistent with the RTTOV sensitivities in **Fig. 5(b)**.
We have also updated **Fig. 5(a)** and **Fig. 10(a)** (labels and captions) to explicitly indicate that the results are autograd-based. We are grateful for the reviewer's guidance, which has improved the clarity and rigor of our sensitivity analysis.

Retrieving relative humidity as opposed to more common practice of retrieving q or log(q) directly is not wrong - it may actually be an interesting idea - but it is uncommon. I'm curious why this was

done - was there any benefit found to retrieving RH instead of q directly? In particular, RH is a function of temperature, which can confuse the sensitivity analysis.

Response:

We thank the reviewer for this insightful comment regarding the selection of relative humidity (RH) as the retrieval variable. We fully agree that physically independent quantities such as specific humidity or mixing ratio offer clearer thermodynamic interpretation and may reduce the explicit dependence on temperature. In the present study, however, RH was intentionally chosen for both methodological and practical reasons. From a scientific perspective, our goal was to explore the coupled retrieval of temperature and humidity, and RH provides a natural diagnostic of this interaction since it inherently reflects the joint variability of temperature and water vapor. This allows the model to learn the spectral sensitivities arising from both the 118 GHz and 183 GHz absorption bands, which are strongly linked to the temperature–moisture coupling **(see Sections 4.2 and the Conclusions)**. Consistently, the relatively high weighting of the 118 GHz channels in the humidity retrieval further supports this coupling and demonstrates the model's ability to effectively utilize the shared temperature-humidity information.

From a data-driven standpoint, RH's bounded range (0-100%) offers numerical stability and facilitates convergence during network training, whereas specific humidity varies over several orders of magnitude across atmospheric layers, which complicates normalization and regression. As this study represents an initial effort to establish a full-channel temperature-humidity retrieval framework for MWRI-RM, RH was selected as a practical and interpretable first target. Nevertheless, we acknowledge the reviewer's valuable point that retrieving specific humidity or mixing ratio would enhance the physical independence and interpretability of the results. We plan to incorporate these quantities in future work, along with error propagation analyses to explicitly decouple temperature and moisture contributions.

I may be missing something important, but I didn't see where the Appendix figure and table were referenced in the manuscript.

Response:

We thank the reviewer for this helpful observation. In the revised manuscript, we have added explicit references to the Appendix materials to improve clarity. Specifically, Table. A1 is now cited in **lines 147-148**, Fig. A1 is referenced in **lines 168-169**, and Fig. A2 is referenced in **lines 292-295**.

Reference

Chen, K., Cai, B., Han, W., and Suo, Z.: Matching of Observation Footprints in the FY-3G MWRI-RM Using BGI, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 17, 1–12, https://doi.org/10.1109/JSTARS.2024.3468437, 2024.

Malmgren-Hansen, D., Laparra, V., Nielsen, A. A., and Camps-Valls, G.: Statistical retrieval of atmospheric profiles with deep convolutional neural networks, ISPRS J. Photogramm. Remote Sens., 158, 231–240, 2019, https://doi.org/10.1016/j.isprsjprs.2019.10.002.

Xu, X., Han, W., Gao, Z., Li, J., and Yin, R.: Retrieval of atmospheric temperature profiles from FY-4A/GIIRS hyperspectral data based on TPE-MLP: analysis of retrieval accuracy and influencing factors, Remote Sens., 16, 1976, https://doi.org/10.3390/rs16111976, 2024.

Reviewer #2

Overview

This paper presents a temperature and relative humidity retrieval for MWRI-RM instrument using a convolutional neural network framework. The authors have made substantial improvements to their analysis and text from the last submission, so I appreciate their commitment to this study.

My main point of information relates to the selection of relative humidity. While the authors make many allusions to the coupling between temperature and moisture (which comes into play with contributions from both absorption bands affecting both quantities), there is not an indication why relative humidity is chosen for the moisture retrieval. After all, relative humidity is directly dependent on temperature by definition, while quantities such as absolute humidity/mixing ratio are more straightforward and independent (with still this temperature-moisture connection as they authors mention). My intuition is that this might provide better performance and behavior for the moisture retrievals as the inputs are "purer" and the dependence on temperature is more indirect than direct. But I am not trying to say this is required for publication.

Response:

We thank the reviewer for this insightful comment regarding the selection of relative humidity (RH) as the retrieval variable. We fully agree that physically independent quantities such as specific humidity or mixing ratio offer clearer thermodynamic interpretation and may reduce the explicit dependence on temperature. In the present study, however, RH was intentionally chosen for both methodological and practical reasons. From a scientific perspective, our goal was to explore the coupled retrieval of temperature and humidity, and RH provides a natural diagnostic of this interaction since it inherently reflects the joint variability of temperature and water vapor. This allows the model to learn the spectral sensitivities arising from both the 118 GHz and 183 GHz absorption bands, which are strongly linked to the temperature–moisture coupling **(see Sections 4.2 and the Conclusions)**. Consistently, the relatively high weighting of the 118 GHz channels in the humidity retrieval further supports this coupling and demonstrates the model's ability to effectively utilize the shared temperature-humidity information.

From a data-driven standpoint, RH's bounded range (0-100%) offers numerical stability and facilitates convergence during network training, whereas specific humidity varies over several orders of magnitude across atmospheric layers, which complicates normalization and regression. As this study represents an initial effort to establish a full-channel temperature-humidity retrieval framework for MWRI-RM, RH was selected as a practical and interpretable first target. Nevertheless, we acknowledge the reviewer's valuable point that retrieving specific humidity or mixing ratio would enhance the physical independence and interpretability of the results. We plan to incorporate these quantities in future work, along with error propagation analyses to explicitly decouple temperature and moisture contributions.

I would like to the authors to explain more the methodology at L96-97. What exactly are preprocessed static parameters obtained through decoding and quality checks? It's unclear to me what this actually means.

Response:

We thank the reviewer for these valuable comments and for pointing out the need for greater clarity in Section 2.2. In the revised manuscript **(lines 109-125)**, we have substantially expanded the description of the data preprocessing and sample construction procedures. As detailed in the revised text, samples were curated by excluding scan points with more than 7 km deviation between the two feedhorn rows (S1 and S2) to ensure high-quality inputs **(lines 110-112)**. Oceanic samples were then exclusively selected using the built-in land-sea mask **(line 112)**.

In every case, more minor suggestions or changes follow:
L60-68: This paragraph seems abrupt, talking first about ML but the moving back to statistics. I would consider splitting into two or separating prior work into method types.
L62: Please define the acronym PWV and GPS.
L109: with _the final_ 20%
L146: I do not understand the bias description. Bias varies vertically and is not uniform or consistent.
L148: Nevertheless, _t_he
L172: Please walk the reader through Figure 3c-h in the text.
L180: I would more explicitly state in the reanalysis, not just "input features"
Figure 3: Unit labels for a and b subfigures
L196-201: Please explain every aspect of this more, from the customization to how weights were extracted.
L274: research._ _For
Figure 8: Unit labels for a and b subfigures

Response:

We sincerely thank the reviewer for the careful reading and for the detailed line-by-line suggestions. All minor comments have been carefully addressed in the revised manuscript. Specifically, we have reorganized the paragraph previously at lines 60-68 to improve logical flow by separating the discussion of statistical and machine-learning-based retrieval methods **(lines 51-80)**. In addition, the acronyms PWV (precipitable water vapor) and GPS (Global Positioning System) have now been explicitly defined, as suggested **(line 59, line 60)**.

We have revised the text at **lines 166-168** to specify that both bias and RMSE values represent layer-averaged statistics rather than uniform vertical quantities. To enhance readability, we have also expanded the description of Figure 3(c)-(h) at **lines 174-188**, guiding the reader step-by-step through each subfigure. Unit labels have been added to Figures 3 and 8 for clarity.

In the revised manuscript, we explicitly distinguish between RTTOV weighting functions **(Fig. 5(b))** and model-side Jacobians obtained via PyTorch autograd. As detailed in **lines 212-224**, we de-normalize outputs, define a scalar objective (sum over temperature levels), compute input gradients with autograd, aggregate absolute gradients over batch and 5×5 spatial indices, and normalize across channels to obtain per-channel attributions (**Eqs. (3)-(4)**). Using this approach, channel importance peaks near 118.75 ± 1.2 GHz **(Fig. 5(a))**, consistent with the RTTOV sensitivities in **Fig. 5(b)**. We have also updated **Fig. 5(a)** and **Fig. 10(a)** (labels and captions) to explicitly indicate that the results are autograd-based.

Minor textual corrections have also been implemented throughout the manuscript.

We thank the reviewer again for these precise and constructive suggestions, which have helped improve the clarity and readability of the paper.