Report #1

This work aims at quantifying the effect of oxygen sounding channels and polarization difference on reconstructing radar reflectivity profiles by comparing three neural networks with different inputs. The varying inputs require that each model is trained separately. The loss curve of each model needs to be stable and reach a global minimum. However, the loss curves in Figure A1 suggest that training is unstable and not equal among the three models. The baseline model (Ex14) was trained for only 40 epochs, while Ex26 and Ex35 were trained for 100 epochs. This difference is probably due to early stopping initiated at a peak in validation loss that persists across several epochs. These peaks in both validation and training loss can be found also already earlier after about 32 epochs and could be a sign of overfitting. This behavior is not discussed in the manuscript, but it will definitely affect model comparison and could explain major parts of the higher uncertainty of the Ex14 predictions compared to Ex26 and Ex35 - rather than the missing oxygen channel and PD inputs. I strongly suggest to ensure that the training becomes more stable, e.g., modify the model architecture and/or data split. Also, the influence of the random processes during model training on the model comparison should be discussed (weight initialization, data shuffling, ...). Currently, I assume that training with a different random seed would largely affect the Ex14 performance. Also, I would like to see the loss curves of each model on a log scale in the same figure instead of the stacked form in Figure A1 to see whether their training and validation losses consistently differ at a given epoch.

Response: Thank you for your valuable feedback, which has greatly improved the robustness of our analysis. We have addressed your concerns regarding training stability, random seed influence, and loss curve presentation as follows:

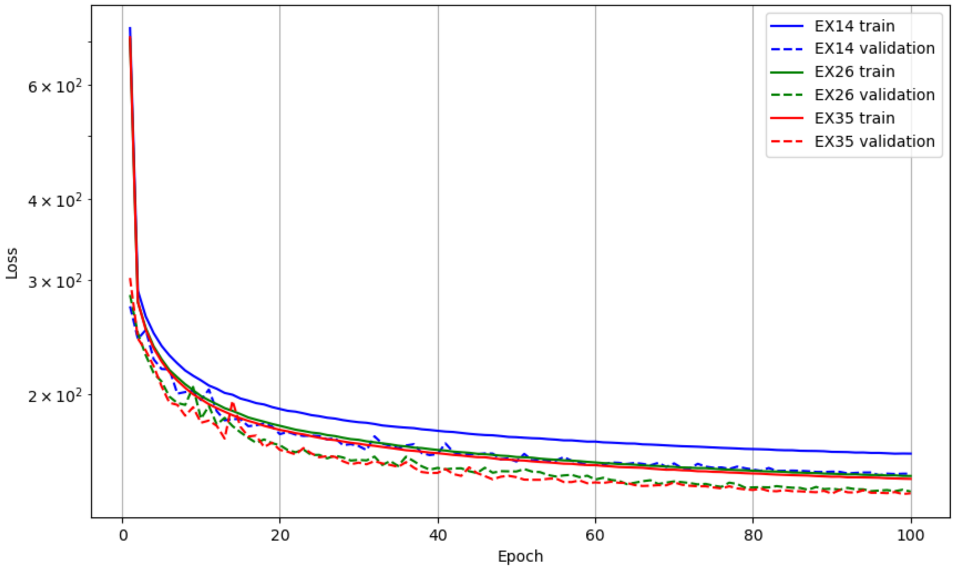
Training Stability and Overfitting: o improve training stability and reduce the risk of overfitting, we increased the EarlyStopping patience parameter from 10 to 20 epochs, providing the models with additional time to stabilize before termination. Furthermore, we refined the model architecture by incorporating pooling layers, which preserve key features while reducing the number of parameters, thereby mitigating overfitting. As a result of these adjustments, all three models—Ex14, Ex26, and Ex35—were successfully trained for 100 epochs, eliminating the previous disparity in training duration (40 epochs for Ex14 vs. 100 epochs for Ex26 and Ex35). Notably, the overall performance of Ex14 has improved substantially compared to the previous version, where training halted prematurely around 40 epochs due to early stopping. This enhancement underscores the effectiveness of the revised training strategy.

Random Seed Influence: To account for the impact of random processes, we conducted three independent training runs for each model using different random seeds. The updated Figure A1 now presents the average training and validation loss curves across these runs on a logarithmic scale in the same figure. This allows for a direct comparison of the loss trends across models at each epoch.

The revised loss curves indicate that Ex35 and Ex26 converge to similar low loss levels by the end of training, while Ex14 consistently exhibits higher losses throughout. This suggests that the differences in performance are likely attributable to the model inputs (e.g., the absence of the oxygen channel inputs in Ex14) rather than training instability or random initialization effects.

We have incorporated these findings into the revised manuscript. In the "Reconstruction Performance and Evaluation" section, we now present the average performance across three independent training runs for each experiment, ensuring robustness against variability introduced by random initialization. In the "Model Performance Evaluation During Extreme Precipitation Events" section, we showcase the best-performing outcomes from each experiment, selected from the three runs, to highlight the models' optimal capabilities in capturing extreme precipitation features.

We believe these revisions address your concerns and strengthen the reliability of our model comparisons.



The response to my question on Line 173 states that the sum of squared error loss function choice is based on findings from your previous studies. Please add the reference to that study to the manuscript. In Yang et al. (2024), Chapter 2.2, it states that mean squared error was used, shortly explain why sum of squared error is used for this work instead.

Response： While Yang et al. (2024) focused solely on reconstructing precipitation samples, our work addresses both precipitation and non-precipitation regions. Given the extreme class imbalance, a standard MSE loss would prioritize low-reflectivity regions, leading to poor performance in critical high-reflectivity zones. To resolve this, we adopt the Sum of Squared Errors (SSE): its quadratic nature magnifies large errors in precipitation regions. This implicit weighting ensures the model prioritizes physically critical regions (e.g., heavy rainfall) while maintaining fidelity across the entire profile. Unlike weighted MSE, SSE avoids subjective weight tuning and inherently aligns with meteorological priorities.

Report #2

Overview: This paper investigates the possibility of using the TB observations from the Microwave Radiance Imager-Rainfall Measurement (MWRI-RM) onboard the Fenyung-3G (FY-3G) satellite to reconstruct the 3D reflectivity profiles. The algorithm is based on a deep learning approach and uses the Ku profiles observed by the Precipitation Measurement Radar (PMR) onboard FY-3G as reference truth. Three deep learning models were trained using different sets of predictors, and two case studies are analyzed. General Comments The authors have carefully answered the reviewers' comments. The main issues raised in the first part of the review process have been addressed. In particular, the test statistics have been calculated on an independent dataset, thus avoiding the possible correlation between the training and test datasets. Good work. Prior to publication, I would like to suggest only minor changes to the text and figures. Minor Comments

1) Line 167: I guess it is the 30th of November and not the 31st of November.

Response: Thank you for catching this error. We have corrected the date to "November 30th" in the revised manuscript.

2) Lines 172-173: it is not clear to me how the authors distinguish "ocean" observations from "land" observations. Are you using a surface mask? I would like to suggest that the authors clarify this.

Response: We appreciate the reviewer's suggestion. The ocean/land classification is based on the "LandSurfaceType" variable in the Level 1 products of the Passive Microwave Radiometer (PMR). This method has been explicitly described in the revised manuscript

3) Lines 250-252: It is not clear to me whether the coastal scenario includes only samples where the radar observations are over the ocean, with part of the passive

microwave FOV over land, or all observations where the passive microwave FOV is partly over land, partly over the ocean, with no reference to the position of the radar observations. I would like to suggest that the authors clarify this.

Response: Thank you for raising this important point. The coastal scenario refers to all cases where the passive microwave field of view (FOV) contains both land and ocean components, irrespective of whether the collocated radar observations are located over land or ocean. We have added this clarification to the revised manuscript

8) Line 376-377: I would suggest that the abbreviations given in Figure 8 should also be included in the text - e.g. "It is worth noting that ground-based radar coverage is limited in remote regions, such as northern Shanxi (SX, 39.6°N, 113.0°E) and Inner Mongolia (IM, 40.0°N, 112.0°E)...". I did not understand whether when the authors speak of Shanxi (e.g., label of Figure 8) and Northern Shanxi (e.g., L. 376) they are referring to the same area. I would like to suggest that the authors clarify this.

Reponse: We thank the reviewer for highlighting this ambiguity. Both "Shanxi" (in Figure 8) and "Northern Shanxi" (in the text) refer to the same geographic area. To resolve confusion, we have adjusted the label in Figure 8 to "Northern Shanxi (SX)" and updated its coordinates to 39.8°N, 113.0°E for precision.

4) Figure 4, Figure 6 and Figure 8: it is rather strange that the lon and lat labels do not correspond to the grid shown on the map (center columns of Figure 4 and Figure 6 and panels (a), (b), and (e) of Figure 8). I would like to suggest changing the grid position.

5) Figure 5 and Figure 7: to make things clearer, I would like to suggest adding the measure units to the axes of the scatterplots - e. g., Target Reflectivity (dBZ), Reconstructed Reflectivity (dBZ). I also suggest changing the maxima of the colour bar scale. This will highlight the distribution of the scatterplots over the plane - in my opinion, the observations where the radar reflectivity is NaN are less interesting. Perhaps a maximum of 1.5 of the normalized density would be better for both figures.

6) Line 330-331: I would suggest adding the abbreviations in Figure 6 to the text. - e. g., "At 4 km altitude, PMR observations showed high reflectivity predominantly in the southern parts of Beijing (BJ, 39.5°N, 115.8°E) and central Hebei (HB, 39.1°N, 116.2°E)"

7) Line 366-368: it is not necessary to mention the position of Beijing and Central Hebei a second time in the text, while I would suggest adding the abbreviation for Tianjin.: "In addition to the precipitation echoes observed by PMR-Ku over southern Beijing, central Hebei, and Tianjin (TJ, 39.2°N, 117.0°E),..."

9) Figure 6: I would like to suggest adding the region abbreviations to all maps in the centre column.

10) Figure 8: I would suggest to add the abbreviations also to (e)

11) Figure 8 - caption: It is not clear to me what the authors are referring to when they write "Panel (b) highlights the locations of Shanxi (SX) province and Inner Mongolia, labeled in red font. ". I would like to suggest adding the red font or deleting this part of the caption.

12) Figure B1 and Figure B2: Thanks to the authors for adding these maps.Thanks to the authors for adding these maps. If possible, I would like to suggest changing the limits of the colormaps to emphasise the signal. For example, for the first panel in the left column of Figure B1 (Khanun Observations 50.3 GHz_V) a minimum value of 220 K would be more useful to highlight the precipitation signal. At the same time, it seems to me that the maxima value for the colormaps in Figure B2 are too low. For example, in the first panel of the left column of Figure B2 (Khanun Observations 50.3 GHz_V), almost the whole map seems to be characterized by the same value also at 50.3 GHz, and this seems a bit strange to me - but I don't know TB values, so maybe it's just my impression. I would also like to suggest that the titles, colorbars and lat/lon labels be enlarged - the increments between lat and lon labels can also be increased if there is too little space. I would also suggest adding the abbreviations used in Figure 6 and in Figure 8 to the plots of Figure B2.

Response: Thank you for your suggestions regarding textual and figure revisions. We have carefully incorporated all recommended changes to improve clarity and accuracy. The revised text and figures (e.g., updated labels, terminology, and visual adjustments) are highlighted in the manuscript for your convenience. We appreciate your thorough review and valuable feedback.