GPROF-NN: A neural network based implementation of the Goddard Profiling Algorithm

The manuscript presents an investigation on using Neural Networks to develop a passive microwave precipitation retrieval model. The authors evaluate their proposed global-scale models using GMI and MHS brightness temperatures compared to the widely used Bayesian-based Goddard Profiling algorithm. In this study, one year of data is utilized to train and validate the NNs models and GPM combined product and MRMS datasets are considered as the reference/ground truth.

The topic is certainly attractive. Satellite precipitation estimation using NNs has been a popular line of research and some papers have recently appeared in the literature. The manuscript is relevant to AMT and follows on to previous NN-related articles by authors including this journal (e.g., Simon Pfreundschus et al. 2018). The manuscript is well written and well organized, the NN methodology sounds, and I do not see any issue with their discussions and general analytical approach. However, the introduction lacks an adequate review of the literature of recent studies, and also data preparation for inputs into the CNN model is poorly explained.

There are some points that should be brought to the attention of the authors that should be easy to address.

Many of your readers may not be familiar with NNs terminology. Would you please highlight the advantages of the NNs, e.g., related CNN and QRNN methods, that today are popular in the satellite precipitation community compared to other ML techniques.
As mentioned before, the study lacks an adequate review of the recent literature about using NNs for satellite precipitation estimation. I suggest some relevant papers (but are not limited to) that are worth reviewing. Please briefly explain already published works in the literature, their challenges/their methodologies, etc., and mention how your work is different from them. What are the open questions you try to address that the previous studies have not considered?
As an example, in Lines 65-70: I understand that you specifically explore the potentials for NNs algorithm in GPROF, so please acknowledge other studies that have already discussed using spatial features in retrieving precipitation.

- *Li, Z., Wen, Y., Schreier, M., Behrangi, A., Hong, Y. and Lambrigtsen, B., 2021. Advancing satellite precipitation retrievals with data-driven approaches: Is black box model explainable?. Earth and Space Science, 8(2), p.e2020EA001423.*
- *Afzali Gorooh, V., Akbari Asanjan, A., Nguyen, P., Hsu, K. and Sorooshian, S., 2022. Deep neural network high SpatioTEmporal resolution Precipitation estimation (Deep-STEP) using Passive Microwave and Infrared Data. Journal of Hydrometeorology.*
- *Sanò, P., Panegrossi, G., Casella, D., Marra, A.C., D'Adderio, L.P., Rysman, J.F. and Dietrich, S., 2018. The passive microwave neural network precipitation retrieval (PNPR) algorithm for the CONICAL scanning Global Microwave Imager (GMI) radiometer. Remote Sensing, 10(7), p.1122.*
- *Ehsani, M.R., Zarei, A., Gupta, H.V., Barnard, K., Lyons, E. and Behrangi, A., 2022. NowCasting-nets: Representation Learning to Mitigate Latency Gap of Satellite Precipitation Products using Convolutional and Recurrent Neural Networks. IEEE Transactions on Geoscience and Remote Sensing.*
-

Lines 13 and 435: How do you define accuracy? Please elaborate on the reported improvements.
Line 15: how do you see the spatial consistency in precipitation retrievals? Does this sentence refer to visualization of derived precipitation rates over Hurricane Harvey for one or two orbital tracks? Please report some statistics for the general spatial detection skills of your proposed models.

Lines 25, Section 3.3: I think the authors need to be cautious in reporting processing time and computational cost comparisons. It is obvious that pixel-wise predictions are faster compared to convolutional-based systems when models are trained and are ready to use. I mean the comparison between GPROF and NN 1D makes sense but including NN 3D is 'Comparing Apples to Oranges'. Processing time means when we have everything set up and ready, let's say we receive one or more orbital tracks (different channels have different footprints, etc.), how long does it take from getting a set of brightness temperatures (Level 1 product) to get the corresponding precipitation maps.

The data preprocessing steps are not clear in the methodology. I suggest summarizing all the training process and prediction (here means after train and validation stage) steps in a numbered list, especially for the CNN algorithm in the methodology section.

Data and method Section: Please clearly explain how many channels are used as inputs to the NN models? What type of resampling/rescaling/interpolation methods do you use? Different radiometers/imagers/sounders have different bands and resolutions, how do you address this problem?
How do you define 18 surface types? Are they generated by TELSEM classification algorithm? Please provide a clear picture of the source of data, pre-processing steps, etc. in this section.
Background material for GPROF Algorithm is well described and cited in previous papers. So please summarize Section 2.2 and please explain more about the innovative parts of your investigation and the proposed models.
Please define all acronyms just the first time you use them. Then use the acronyms in the rest of the manuscript.

Line 200: How many trainable parameters do the NNs algorithms have? Is one year of information enough for training and validating the models?
Line 220: Please add some information about the training stage of models. For example, what are the size of 3D inputs to the CNN model in the training stage? how do you pre-process data to come up with input training samples?

Figure 5. Define regions A and B in the figure.
Please explain the augmentation process in the training stage of CNN model development.

Results Section: from my understanding, the authors only used one year of data for training and validation of NN models. Also, the first three days of 12 months (36 days) are used for test purposes. It is not clear, how many samples (bins) are exactly utilized for testing and reporting the scatter plots and biases? What is the threshold for rain/no-rain samples?

Figures 6 and 7. I do not see a good reason for the color used in these figures, and I find it confusing, commonly blue-red colors would reveal more features. Please find similar figures as the example in Utsumi et al 2020 paper. Also, please report some common statistical indices (related to scatterplots) or detection skill metrics to reveal the discrepancies/improvements. It is better to judge the performance based on statistical indices along with visual assessments.

- *Utsumi, N., Turk, F.J., Haddad, Z.S., Kirstetter, P.-E., Kim, H., 2020. Evaluation of Precipitation Vertical Profiles Estimated by GPM-Era Satellite-Based Passive Microwave Retrievals. Journal of Hydrometeorology 22, 95–112. https://doi.org/10.1175/JHM-D-20-0160.1*

Figure 7: What are the vertical white lines in the last panel of the figure? (Lowest right scatter plot)?

Figure 8 and the associated discussions in this section: The authors mentioned that they have used 18 surfaces classed. Did they regroup the precipitation over different surface types in order to report the statistics? Or here they just report 4 types out of 18? How do the proposed models perform on arid land surface types?

Line 303: Again, how many samples are used to calculate Bias, MSE, etc in each pixels/5-degree box?

Section 3.2: This section presents a visualization of precipitation rates over one or two orbital tracks during Hurricane Harvey. Would you please report some basic statistical indices such as the probability of detection, missed ratio, etc.

Figures 13 and 14. Please show CMB and MRMS products in both figures. Please use the commonly used blue red colorbar and colormap for presenting precipitation rates. Revise the figure in a way that the rain rates less than 1 mm/h are not eliminated. I see that the figures are patchy, and the spatial patterns of precipitation rates are not obvious.
Please remove the colorful background from figure 14. and again, it is miss leading when the precipitation rates less than 1 mm/h in panels c, d, g, h is not shown in the figures.

Section 3.3 as mentioned before, I suggest removing this part or please add more information for different stages of developing NN 1D and NN 3D models, to avoid confusion for the readers. I understand that GPUs, TPUs, etc. can be used to train deep neural networks, and the processing time when everything is ready for the model can be fast for pixel-wise NN 1D. Using NN 3D may be relatively fast in precipitation estimation (prediction phase), but the data preprocessing takes time and is not mentioned here.

Line 413, 461: Please avoid using “simply” replacing or developing. It is not simple!
Line 440: Again, please review the study by Li et al. 2021 and more recent ones that use CNN and PMW data are a part of their input datasets. It is worth mentioning previous works at least in the introduction. Also, it is already established that using neighboring information (spatial features) improves the satellite retrievals both in capturing the amount and the location of events.

- *Li, Z., Wen, Y., Schreier, M., Behrangi, A., Hong, Y. and Lambrigtsen, B., 2021. Advancing satellite precipitation retrievals with data-driven approaches: Is black box model explainable?. Earth and Space Science, 8(2), p.e2020EA001423.*
- *Afzali Gorooh, V., Akbari Asanjan, A., Nguyen, P., Hsu, K. and Sorooshian, S., 2022. Deep neural network high SpatioTEmporal resolution Precipitation estimation (Deep-STEP) using Passive Microwave and Infrared Data. Journal of Hydrometeorology.*
  and many more,…

Line 440-445, 452: No evidence has been reported or shown that the model is trained properly. At least please mention the number of samples in the training and testing process, how do the authors

select the hyperparameters? How many parameters do the NN models have compared to GPROF? The Hurricane Harvey event was just a visual representation of retrievals. By adding statistic indices such as pixel- or window-wise correlation, false alarm, missed ratio, etc., the reader can find the improvements and the differences (not only by reporting average bias and visual assessments).
Line 455: Quoting from the manuscript "an additional neural network model was required to transform the data from the retrieval database into a form that is amenable for training a CNN…", I invite the authors to clearly explain the process in the manuscript. It is not clear!
Line 477: I suggest replacing "warming climate" with something like changing climate.