Response to Referee Comment 2 (RC2) from Anonymous Referee # 1 (*Referee comment in Italic*, response in blue).

Review of manuscript "A New Machine Learning based Analysis for Improving Satellite Retrieved Atmospheric Composition Data: OMI SO2 as an Example" by Can Li et al. 2022

In the manuscript a new and interesting method to minimize noise and artifacts using machine learning has been presented and applied to OMI SO2 data. The paper is well written and only minor revisions are needed.

*My **main comment** is that an optimization (or sensitivity analysis) of the applied NN architecture is missing. A simple NN architecture from a completely different scientific field ("for reconstruction of RGB images from hyperspectral radiances") was used. In general a simple architecure is fine as a starting point, but what I am clearly missing is a sensitivity analysis of much the NN architecture is affecting the results.*

Furthermore I suggest some restructuring of the paper - I think it would be good to start with the simple approach of using the linear interpolation model and then begin the NN model.

Althoug the comparison and different maps are nice to see, I suggest to also add line plots as a function e.g. latitude such that is easier to see biases and differences.

We thank the referee for providing several helpful suggestions. The NN architecture used in this study has also been applied to a few other studies on different topics and we have added references to those. Additionally, we have also conducted more tests on the architecture, please see below our response to the specific suggestion.

We have also carefully considered the suggestion to move the part of the paper on the linear interpolation model to the front. We agree that this change could help to justify the use of NN in this study. A potential issue is that the linear model results are compared with NN analyzed results, so it would be necessary to introduce the NN approach first. Because of this, we have elected to keep the overall structure of the paper unchanged. We have added, in the methodology section, that a simple linear interpolation model does not work quite as well as the NN based approach.

We have also added line plots to several figures following the suggestions.

Detailed comments:

- *Sect.1, L49: Please provide numbers or references for the background noise of OMI SO2 SCDs.*

Reference added.

- *Sect 1. L80ff: The 20 DU limit of the FP-ILM retrieval only applies to the SO2 LH retrievals. It was not yet applied to SO2 VCD retrievals.*

We have clarified this point in the revised paper.

- *Sect 2.1, L109: How are pixels with enhanced SO2 after volcanic eruptions detected/filtered? Do you apply a VCD threshold? Please describe.*

In the PCA retrieval algorithm, this is done by examining the O_3 residuals at two wavelength pairs (313/314 nm and 314/315 nm). We have added the reference for this in the revised manuscript.

- *Sect 2.1, L128: By how much do the monthly medians/stddev change every month? Do you see jumps in the results from one month to the other?*

The monthly medians and standard deviation are strongly affected by SZA. When calculated as a function of latitude, both show seasonal changes (please see figure below).

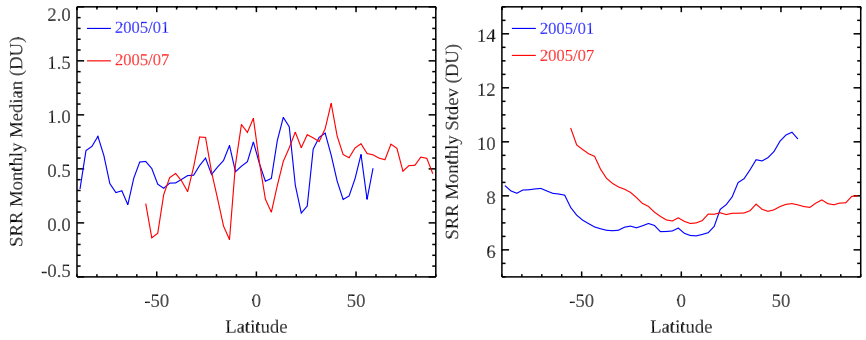

Figure. (Left) Monthly median and (right) standard deviation of SCD/RMS (SRR) ratios for different latitude bands for January and July 2005.

- *Sect. 2.2, L146: I am a bit concerned that the a1/a2 factors are based on trial and error and there is no robust criterion to determine them. This makes appying the whole method to other sensors (or even to OMI for a different time span) difficult - I assume the factors will vary over time, especially related to instrument degradation. Can you perhaps show how sensitive your results are to (small) changes in a1/a2?*

Following your suggestion and also that from the Referee #2, we have conducted sensitivity tests by changing a1 and a2 by +/-10%. While the NN analyzed SCDs show some sensitivity to a1 and a2, overall the sensitivity is not very large (< 0.1 DU for polluted areas and 0.03 DU or less for clean areas). With that, we think it would probably be better to use a constant set of a1 and a2 parameters for long-term data analysis. For applications to other instruments or datasets, a new set of parameters are most likely

needed given the differences between instruments/retrievals. We have added these results and the discussion to the revised paper.

- *Sect 2.3, L180ff: As already mentioned in my main comment, I am a bit concerned about the choice of the NN architecture. Although it is a good starting point to use a simple archictecture, it definitely needs to be optimized for the specific problem, especially independently for your NN1 and NN2.*

  Is there a reason for the choice of activation functions? I.e. using soft-sign and then sigmoid is not really common - I suggest to use ReLu or something related for both hidden layers.

Thank you for the suggestion. We have conducted some additional tests. In one, we altered the number of nodes in the hidden layers (to 20 for both layers and then to 30). We found only marginal changes in the performance of the NNs. For example, for training for March 2005, the RMSE of NN SCDs are 0.0331 DU, 0.0330 DU, and 0.0329 DU for our current setup, 20 nodes per hidden layer, and 30 nodes per hidden layer, respectively.

In another test, we used ReLU as the activation function for both hidden layers. We notice that the new setup speeds up the training and reduces the RMSE of NN SCDs from 0.0331 DU to 0.0319 DU for March 2005. But once we applied the trained NNs to the entire month, we notice that the new setup overall increases the SCDs over background areas by ~0.01-0.02 DU (see figure below). Of course, the pixel selection scheme has been optimized for our existing setup, so this may not be a completely fair comparison. We have added this discussion to the revised manuscript (with the figure added to the supplemental information) and we intend to conduct more extensive tests in follow-up studies.

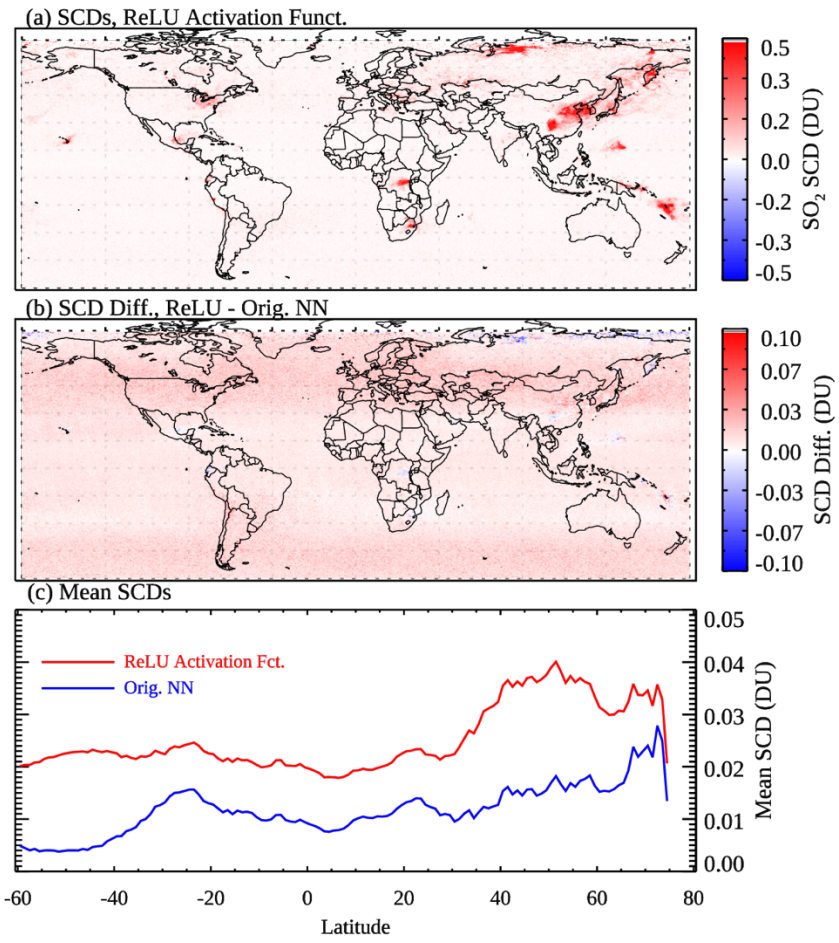

Figure. (a) Monthly mean analyzed SCDs for March 2005 using ReLU as the activation function in both hidden layers of the NNs. (b) The differences in the monthly mean analyzed SCDs for the same month between the NNs using ReLU as the activation function and the original NNs (soft sign and sigmoid as the activation function for the first and second hidden layer, respectively). (c) The mean SO_2 SCDs for 1-degree latitude bands from the two NN architectures over generally clean areas (monthly mean SRR < 3).

- *Sect. 4.3: I suggest to put this in front before you apply the NNs, to show that a simple linear interpolation is not sufficient and therefore you apply NNs.*

As mentioned above, the linear model results are compared with NN results in this section, and it is therefore necessary to introduce NN results first. We feel that it is probably better to keep the section in its current place.

- *Figure 4&5: Suggest to add line plot of SCD as a function of latitude. With this plot you probably better see biases and the differences*

Line plots added.

- *Figure 5: Suggest to use different color scale from -0.1 - 0.1*

We experimented with a few different color scales and selected -0.2 to 0.2 DU in the updated figure.

- *Figure6b: Relative differences are always problematic for SO2 plot since in clean areas the SCD is close to zero and hence the relative difference becomes extremly high, as can be seen in the figure. Suggest to use absolute differences here.*

We did try to plot absolute differences instead of relative differences for Figure 6 (please see below). We feel that it is more difficult to distinguish between polluted and clean areas from the absolute differences. For this reason, we have elected to keep the relative difference plot in Figure 6.

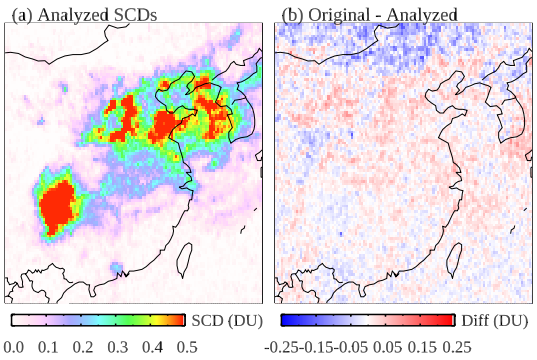

Figure. Same as Figure 6 but showing absolute differences in (b).

- *Figure 7: The maps are rather confusing and do not provide addtional information. I sgueest to rather repleace them by line plots as a function of e.g. latitude (see my comment above)*

Following your suggestion and also that from Referee #2, we have changed Figure 7 by adding line plots and also moving some subplots (maps) to the supplemental information. These changes make Figure 7 more readable and easier to interpret. We have also updated the text in the revised paper accordingly.

- *Figure 9&10: Suggest to add line plots of SCD as a function of e.g. latitude. With this plot you probably better see biases and the differences.*

Line plots added.