EGUsphere, referee comment RC1
https://doi.org/10.5194/egusphere-2022-806-RC1, 2022

Comment on egusphere-2022-806

Anonymous Referee #2

Referee comment on "Use of machine learning and principal component analysis to retrieve nitrogen dioxide (NO_2) with hyperspectral imagers and reduce noise in spectral fitting" by Joanna Joiner et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-806-RC1, 2022

The paper by Joiner et al. is a carefully, well written paper. It contains interesting results. First of all, it shows that a meaningful retrieval of NO2 is possible at a moderate spectral resolution of about 5 nm as provided by OCI. Secondly, it demonstrates a neural net PCA approach as a fast way to do the retrievals, an approach with interesting properties. As such I am in favour of publishing these results. However, I have several questions I would ask the authors to address in the final version of the paper.

page 5, line 108: "The conversion of SCD to total or tropospheric VCD can be accomplished in a straight-forward and computationally efficient manner as in current algorithms and is not addressed further here. " Nevertheless, I understand that the largest contributions to the overall error is in the conversion to vertical columns. Information on the clouds, aerosols and albedo is important for this step. It would be of interest to comment if such information is available from OCI/PACE.

line 158: "independent of radiance value." The cloud-free scenes are the most relevant ones, and these have lower radiance levels than cloud covered scenes. Does this assumption not lead to an over-optimistic representation of the results?

line 172: "The PCA concentrates the spectral features" -> The PCA concentrates on the spectral features

line 275 "nomainal"

line 280: "GLIMR, even with no noise, were not satisfactory .. are not shown". I would like to ask the authors to consider to provide them anyway, or add a few lines to table 1 or 2. Linked to figure 1, this would more explicitly show that meaningful retrievals are possible

up to 5nm resolution, but that 10nm is removing much of the useful NO2 information.

How does the NN-OMNO2 RMS compare to the OMNO2 retrieval uncertainty? It seems there is still quite a spread and some systematic effects caused by the NN approach (Fig.4). Please comment and put the results in perspective.

Data is shown for 28 January over the US, when both the solar angle and NO2 column amount is relatively high. It would be interesting to show also an example (maybe in the tables only) for the summer to check the seasonality of the differences.

line 313: "generally low bias over highly polluted areas" What is the reason for this low bias? I would expect instrument noise to lead to random effects, not a systematic bias.

line 364: "convolved with a 1 nm boxcar function" Could you explain why this is done, instead of using OMI radiances at their spectral resolution?

Figure 9: It was confusing to me what can be concluded from these plots. Several things were changed: wider spectral range, change of spectral resolution, NN versus standard retrieval. Is the effect due to the wider window, or could the NN procedure cause a lowering of the apparent noise? Is the NN distribution of SCD values more realistic than the OMNO2 one?

line 403: "could be used for emissions estimates". Most emission estimation methods make use of daily data, basically identifying the plume from a localised source, which then allows the estimation of the source (and these daily values may then be averaged in time). Would monthly-averaged maps be of use for emission estimates?