Response to Sergio DeSouza-Machado

The paper is an extension of the 2018 JOAT paper “Cloud-Assisted Retrieval of Lower-Stratospheric Water Vapor from Nadir-View Satellite Measurements," by J. Feng and Y. Huang, where the authors used a physically based retrieval that when compared to in-situ data, performed better than the AIRS L2 operational cloud clearing and subsequent retrieval above deep convective clouds. The AIRS L2 algorithm is more designed for tropospheric retrievals, and some of the inaccuracies could be attributed to that. In addition, the cloud clearing process degrades the spatial resolution of the AIRS observations (15 km) to a 45 km footprint. The current paper is an significant extension of the previous paper. Measurements of UTLS water vapor is important for climate studies, and this paper is an important step towards understanding how to improve the retrievals above convective storms, for subsequent use by the scientific community.

I find the paper well written, and would be a very welcome addition to the literature. Though I understand the paper is purely a simulation exercise, the fact that we currently enjoy almost 20 continuous years of high quality, low noise hyperspectral radiance measurements, means the paper would benet from some corrections and/or improved explanations, and a discussion of how the consequences/limitations of realistic instrument parameters would affect your simulated results.

We thank Sergio DeSouza-Machado for his constructive comments. Following his suggestion, we have clarified the noise level of the instrument used in this study and improved the instrument channel selection by removing O3 and CH4 channels.

Primarily the addition of the T/WV profiles a few hours \after" the event to the observation vector is puzzling. I'm not an expert on dynamics, so could the authors explain why 400 minutes (7 hours) were chosen? Is this some (intensification?) timescale associated with tropical cyclonic activity? *Furthermore, the authors have simulated a tropical cyclone, which lasts for a few days and hence is well tracked with NWP fields populated by data assimilation. I can't see how this would help for a mesoscale outburst over a land mass, which would have timescales of hours?*

1) **Choosing this time step is arbitrary, simply to represent the quantitative difference between products such as reanalysis and the truth atmospheric conditions.**
2) **It is a good point considering how the addition of infrared spectra may help the data assimilation process. We have not worked on this problem so far.**

Figures 3(a) 3(b) and Figure 4 are misleading. Current Infrared sounders have detectors between 650 cm-1 to 2780 cm-1, so adding far-infrared channels to your OSSE could be unrealistically \helping" the retrieval. Can you remove these channels and comment how the retrieval is affected, for example in terms of degrees of freedom?

The far infrared channel in Figures 3 and 4 is to evaluate the effects of cloud more generally. It is not included in the retrieval. We now emphasis the AIRS-like instrument specification in the paper. To avoid confusion, we mark the AIRS spectral range used in this study in Figure 4.

Furthermore, the spectrally constant 0.3 K noise estimate is extremely conservative by today's standards. For a 250 K observation, the noise in the high altitude 15 um temperature sounding channels is close to 0.6 K; for the 200 K simulated observations, the noise would be larger than 1 K (!). Similarly in the WV sounding region (1300-1700 cm-1) the noise would increase from about 0.2 K (close to what you use in your OSSE) to about 0.6 K. Again could you test and comment how this affects your retrieval performance? All these points should combine to lower the Degrees of Freedom that you retrieve. For the reasons above, the numbers you quote (3 for T(z), almost 3 for IWC) seem very optimistic for what is being retrieved from 13 km to 80 km, where the typical sounder channels rapidly run out of steam.

In the OSSE, a radiometric noise from AIRS L1B is used. Therefore, the DFS shown in the original manuscript is not biased due to the noise level.

In the original manuscript, a 0.3 K NEdT is used for plotting, just to be consistent with AIRS documentation (which is reported at a 250K reference level) and also Feng and Huang., 2018, to emphasize how the addition of cloud-induced uncertainties affects the signal detection. You are right, NEdT is indeed much higher at a cold reference temperature like you said. In the revision, we now use 0.5 K, the spectral mean value, in Figure 4.

Another point that should be mentioned is the high altitude channels have very narrow doppler linewidths. For example in the 1500 cm-1 WV region, an LBL code would show lines with widths on the order of 0.05 cm-1; the linewidths would decrease even further as you get deeper into LW IR or FarIR (eg the temperature sounding channels at 15 um). So 0.1 cm-1 from MODTran is not sufficient, even if you then convolve over an AIRS SRF which is typically 0.5-2 cm-1 wide.

A line-by-line algorithm of MODTRAN 6.0 is used. This LBL algorithm segments 100 (user-defined) sub-bins within the 0.1 cm-1 (for calculating the line shape of each molecular transition). The Doppler broadening width is therefore accurately accounted for. We now clarify it in the manuscript to avoid confusion.

Finally, you use CloudSat cloud profiles in your \observation" state vector (Page 9, line 205); but the satellite with that instrument on board has moved away from the A-Train and you no longer have co-located measurements. Besides it was purely a nadir looking instrument, and hence would more likely than not miss the DCC that AIRS regularly see. In your conclusions (or somewhere else) you should mention these facts.

Thank you for pointing it out. I clarify this important fact when introducing the instruments/ products.

Below are some additional points I have come across :

1. Abstract : Since this is an OSSE, so all your work is with simulated radiances. Hence in the second line of the abstract, please write \non-negligible impact on simulated TOA infrared radiances". By the way you need to define what TOA means.

Done. The abstract is greatly expanded as well.

2. Page 1, line 20-23 : as stated above AIRS L2 does not focus on UT/LS water vapor, though the process of cloud clearing should work best in the non-homogeneous conditions that occur when there are DCC mixed in with overshooting clouds. Perhaps you could introduce the names (and acronyms) of current sounders in line 23, around where you mention Susskind's paper etc.
Done. The introduction is greatly expanded. See L44-51.

3. Page 2, line 29 : Please also include Irion, F. W., Kahn, B. H., Schreier, M. M., Fetzer, E. J., Fishbein, E., Fu, D., Kalmus, P., Wilson, R. C., Wong, S., and Yue, Q.: Single-footprint retrievals of temperature, water vapor and cloud properties from AIRS, Atmos. Meas. Tech., 11, 971{995, https://doi.org/10.5194/amt-11-971-2018, 2018
Done. Abstract and introduction are greatly expanded, following this comment.

4. Page 3, line 85-90 : Would you consider showing a MODIS image, and an AIRS BT1231 cm-1 image for this May 16, 2015 storm? I can see AIRS granule 41 for that date would work (see also coincident MODIS 4.05 AM image). This maybe a good check of the realism of the simulations (though yours were performed for about 7 AM of that day). I did a quick check and there were 242 AIRS observations with BT1231 colder than 200 K.

We select this May 16, 2015 storm actually due to the collocation of AIRS granule 41. But collocation between AIRS and CloudSat went bad after 2015 February, therefore, we did not perform further validation tests using actual AIRS observation. We did not use simulation outputs at 4:05 am because the model has to spin up for six hours to properly generate cloud fields. I think adding actual observation can be somewhat misleading, as it is hard to explain why there is a quantitative difference between AIRS and GEM simulation in the first place. Therefore, the MODIS picture is not added. These are not mentioned in the text because there are largely not relevant to the focus we aim to present in this work.

5. Page 6, Line 145-146 and Figure 3 : you may want to point out eg 667 cm-1 and 2300-2360 cm-1 are very high altitude temperature sounding channels, whose weighting functions peak way above the tropopause (and hence the clouds you put in); similarly 1000 cm-1 region is the ozone sounding region, and O3 weighting functions peak in the stratosphere.
2300-2360 cm-1 is not included in the manuscript. Following your comments, the ozone absorption channel is removed as well because ozone is not retrieved. Similar for methane.

6. Page 6, Line 145-146 and Figure 3 : so it looks like other than the difference in magnitude, the effective radius and IWC jacobians are very similar, so it is quite difficult to unscramble them, other than constraining the effective radius in your retrieval. For completeness, you should discuss what the cloud top/cloud bottom jacobians look like.

We now add a figure showing the normalized spectral uncertainties caused by effective radius and IWC. It is found that signals from effective radii might be distinguishable through a spectral tilting pattern. There was a technical mistake in calculating DFS for effective radius. We now correct it and find that the DFS in effective radius is around 0.6.

The TOA spectra are not affected by cloud bottom (in this tropical deep convective system). In case you are interested, in the revised manuscript, we added an additional test to show whether effective radius changes across different vertical levels matter. The answer is no. Only levels around optical depth of 1 (measured from cloud top) matters to the spectral. Therefore, the effective radius we try to retrieve here is vertically constant for a profile.

7. Page 10, Figure 4 panels (a),(c) refer to my earlier comment, namely current sounders have detectors between 640-2700 cm-1, and the noise levels at 640-720 cm-1 are far larger than you use.
See previous response. In the retrieval study, only AIRS channels are used, strictly following the radiometric noise of the instrument, which is not affected by scenes temperature (except a few channels that are already excluded).

8. Page 10, Figure 4(b) is very interesting. Very often AIRS DCC observations at 1231 cm-1 are far colder than NWP tropopause temperatures, so that kink at 80 mb is required. Are there any consequences to the stability of the atmosphere and the tops of overshooting DCC, since your cold profile snaps back to the average stratospheric profile very rapidly? (by about 75 mb!!!).

You might be interested to check the accompanying paper, Feng and Huang 2021., where we show the cold BT anomaly in CO2 absorption channels above storms, from the AIRS L1b product. This cold signature is found to be typical through our retrieval study performed in Feng and Huang 2021. Your question about the stability of the atmosphere caused by temperature perturbation is very interesting. We will look into it.

9. Page 11, Line 225 : Please give a reference for the mapping from one vertical grid to another (eg the remote sensing book by Rodgers 2000, or some of the TES retrieval algorithm papers by eg Worden or Kuwalik. Irion's 2018 AMT paper mentioned above should also have that).

Added. Do you refer to the least-square interpolation using the Moore-Penrose inverse? We did not add the reference in the original manuscript because we thought it is commonly used and purely mathematical.

10. page 14, line 315 : replace \board" with \broad"
Done.

11. Figure 5, page 15 : I am impressed by the DOFs and ability of your retrieval to get the overall details correct. However as mentioned above the AIRS NeDT is much larger than you use in your OSSE.

Plus the AIRS channels typically has less information content for higher altitudes. So could you run a quick check of how this would impact retrieval performance?
See previous responses. DOFs are not biased due to NedT.

Yes, indeed AIRS contains less information content for high altitudes (DFS between 0.5-0.8 for water vapor, and between 2-4 for temperature, depending on clouds). This is why we add an additional atmospheric constraint in the framework; otherwise, when it comes into the application with real AIRS observation, the retrieval hardly converges. In the accompanying ACP paper, we found that incorporating ERA5 as the additional atmospheric constraint increases the convergence rate to 80%.

Also, DFS in IWC and effective radius are high, which can be somehow expected from the sensitivity test using DARDAR-Cloud. This is because the TOA observed radiance is strongly contributed by cloud emissions near its top. For thinner clouds, this DFS may decrease.

12. Figure 6, page 16 : please check the caption, I think you mean rows 1-5, not columns 1-5; if so there are other sentences in the caption that need fixing.
Done.

[13. Could you mention what CO2 and CH4 ppm value you used for your retrieval?
14. You have not explicitly stated which trace gases you retrieved (ozone? CH4?), if any. This goes back to assuming you are using all channels shown in Figures 3,4. Assuming you did use all channels but did not t eg CH4 and O3 and CO, perhaps you could get improved results if you removed the spectral regions that those gases cover?]

Other graces are fixed at tropical mean values and do not affect the OSSE. Following your comments, we have removed absorption channels of O3 and CH4 and clarify our channel selection. Thank you for pointing it out. We have made necessary changes to the accompanying paper as well.