DETAILED COMMENTS for AMT Manuscript No. AMT-2020-518 "An observing system simulation experiment (OSSE)-based assessment of the retrieval of above-cloud temperature and water vapor using hyperspectral infrared sounder" by J. Feng et. al

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 benefit from some corrections and/or improved explanations, and a discussion of how the consequences/limitations of realistic instrument parameters would affect your simulated results.

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?

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?

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.

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, a 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.

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.

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.

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.

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

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.

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.

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.

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.

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!!!).
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).
10. page 14, line 315 : replace "board" with "broad"
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?
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.
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 fit eg CH4 and O3 and CO, perhaps you could get improved results if you removed the spectral regions that those gases cover?