Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC1
https://doi.org/10.5194/amt-2022-237-RC1, 2022

Comment on amt-2022-237

Anonymous Referee #1

Referee comment on "Satellite remote-sensing capability to assess tropospheric-column ratios of formaldehyde and nitrogen dioxide: case study during the Long Island Sound Tropospheric Ozone Study 2018 (LISTOS 2018) field campaign" by Matthew S. Johnson et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-237-RC1, 2022

This study evaluates OMI and TROPOMI retrievals of NO2, HCHO and FNR using aircraft measurements during the LISTOS campaign. The manuscript is well-written, and it is a good for for AMT. See my comments below.

- The authors focus on statistical results of the comparison, especially the mean biases. But I don't think that mean biases could tell much about the uncertainties of TROPOMI and OMI. The standard deviation of the mean biases is large, which made me wonder if the overestimates of underestimates are broadly consistent. If not, presenting the mean biases here may not help understand the performance of satellite retrievals. For example, how well do these retrievals capture the spatial and temporal variability of NO2, HCHO, and FNR? And how the errors in satellite retrievals affect the interpretations of the ozone sensitivity?
- It's also not clear to me how the statistical results drawn from a single field campaign can be generalized to other regions or other time periods. I'd strongly recommend the authors go beyond the statistical comparison, and have a more thorough discussions about the sources of uncertainties, and the associated errors, and whether their conclusions can be generalized.

Specific Comments:

Abstract: The abstract is lengthy. I'd suggest the authors shorten the abstract to include only the core findings of this work. For example, the first paragraph may belong to introduction.

Line 370: What are the quality flags for? Is this the same quality flag as for TROPOMI? If so, why do you choose different thresholds? Better to include references here.

Table 2: I'd suggest include an estimate of the error, such as normalized mean standard errors. NMB doesn't tell much about the precision of the retrievals.

Line 530: Maybe you could have a figure of the mean biases of HCHO to show where OMI or TROPOMI HCHO is biased high?

Line 600: I'm not sure if we could call this as a 'high pollution' day because ozone was actually low on this day. This could very well be a cold day when the lifetime of NO2 is long, and the the photolysis is low. I'm not sure how much value there is to evaluate FNR on this day. It'd be more interesting to add another day with both high ozone and high NO2.

Line 700: It is interesting to see that improved the a priori from CMAQ does not improve the retrieval performance of OMI. The authors attribute this to coarse resolution of OMI. Could this be due to the coarse resolution of cloud and surface albedo data used in the retrieval?

Line 835: While the low mean biases of FNR is low, the standard deviation is very large. The R sure is also low for FNR. Thus I don't think the errors of HCHO and NO2 could cancel out. The errors in HCHO and NO2 can offset only if the errors are **correlated**. I'd suggest the authors make a scatter plot for errors of HCHO versus NO2, and see if they are correlated.