Review of Nawaz et al.

**General Comments**

The authors have improved the manuscript and addressed many of the comments. The additional comparisons of temperature, windspeed, and wind direction against 16 ground monitors are useful. I still have two major comments:

1. The authors say there is good agreement also against the surface observations, but how do I know that their agreement is good? Table S6 gives an $R^2$ for wind direction of 0.26 which is not very good. I would have preferred to see scatterplots rather than (or in addition to) tables to determine whether the $R^2$ values were being driven by outliers or overall model performance. The authors should better discuss the meaning of their meteorological comparison statistics and how any biases could impact their results. This could be a very positive discussion that could inform future studies about what model biases are impacting emissions assessments. Currently the manuscript is a little confusing about what is ‘good agreement’ and what is a meaningful bias that needs to be discussed.

2. I am still concerned by the lack of evaluation of model vertical structure. As the model has a strong negative bias in $NO_2$, it is possible this is due to errors in wind direction (as the authors suggest) but also to potential excessive vertical mixing. As vertical structure is essential to successful satellite interpretation, I do not agree that comparisons between vertical profiles of temperature, RH etc is out of the scope of this project. I do not see how this is different than the comparison done with the surface stations. I would again strongly suggest the authors add an evaluation against the data from the TRACER-AQ ozonesondes (https://www-air.larc.nasa.gov/cgi-bin/ArcView/traceraq.2021?SONDE=1) that include temperature, pressure, windspeed, wind direction, RH, and of course ozone. Possibly the model output was not saved for this which then certainly would be burdensome to produce? If so, then a statement about how this could be done in the future (or with other datasets like HSRL-2 or TolNet) would be helpful.

**Specific Comments**

*Line numbers refer to the version with track changes.*

Line 170 – Please add the swath size for GCAS.

Line 237 – Previously I commented “Line 190 – For comparison to TROPOMI, you need to regrid the model to the coarser TROPOMI resolution of 3.5x5.5 km2, otherwise the comparison will certainly look poor.” The response is as follows:

Fig. 2 – Is the model output at 15-min or hourly resolution? I read both in the previous text. It would be useful to know how variable Pandora is across the hour if using model hourly to know how much that matters.

**In section 3.4 we have done this already and found generally comparable performance albeit improved correlation when comparing results at a coarser resolution.**

Please add a reference to Section 3.4 about this then on line 237, otherwise as a reader I would be thrown off.

Line 446 - The authors state that there is a non-systematic difference in wind direction on the order of 20º, but then state two sentences later that there is no apparent systematic bias in meteorology. This is contradictory. Also, Tables S5-S6 give statistics for all days and non-cloudy days, but not windy vs. calm.

Line 550 – Can you give us some more information about the performance of the scheme used here (YSU?) compared to others?

Line 684 – Over how much of a larger extent?

Line 990 – The authors have discussed errors in meteorology throughout the paper including a bias of 20º in wind direction. I am not convinced errors are “minimal”. It would be better to say that: with caveats that there are some errors in meteorology, they are unlikely to fully explain the low $NO_2$ bias in the CAMx column and some of this bias may be attributable to underestimated emissions.

Data availability – the data availability statement does not include the TCEQ emissions inventory.