Summary: This manuscript compares OMI columns to Pandora columns for formaldehyde, nitrogen dioxide, and ozone. Section 2.0 references 145 Pandoras, but figures show data fewer. The authors highlight that TCNO2 has a diurnal profile that requires time-pairing when comparing Pandora to satellite. They also highlight that OMI THCHO lacks seasonality that is expected and seen in Pandora.

Response overview:

This manuscript is not ready for publication. The goal appears to be evaluation the OMI dataset using Pandora and EPIC. It needs to be better organized, with more methodological details, and improved quantification. There are a combination of inconsistent methods (Loess vs monthly avg), statements that seem to imply methods that are neither discussed nor have results reported. One of the current conclusions seem like they would be removed if the methods had been more appropriate. With additional methods and quantification, this will be a nice contribution to assessing the satellite assets currently monitoring our atmosphere.

- The introduction does not currently describe the motivation for the study and needs reorganization. The introduction discusses sources of pollutants and closes with the idea that it will compare OMI and Pandora. It does not describe why a comparison of OMI and Pandora would be useful. It appears to be a validation paper, which is good but is not clear. If this is a validation paper, the paper should include descriptive statistics (and more stations).

- The manuscript needs a methods section.

- The authors need to describe the Pandora data selection and filtering. **Filtering done by error estimate and distance from Pandora site**

- The authors describe example file names, but do not discuss their meaning. As a Pandora user, I am aware that rnvs means direct sun NO2 and rfus means sky-scan HCHO, but the average reader may not.

rfus = direct sun HCHO not sky scan rfuh = sky scan HCHO

- This raises the question why are you using sky-scan measurements for HCHO? And, how were stations selected as having sufficiently high quality sky-scan HCHO for comparison? **Not using sky scan**

- Which stations are used and where are they? Section 2.0 references 145, the acknowledgements references 63, and I did not count the number referenced in the

figures. Given that conclusion statements like "most sites" are made, it should be clear which sites were part of the analysis.

I have added **Table 1 List of Pandora locations used in this study in order of appearance**

- How OMI and Pandora were paired. Are OMI pixels within a certain distance used? Or only when the Pandora site is within the pixel geometry defined by the corners?

OMI pixels within 50 km filtered by my Fortran program

- The OMI product is insufficiently described in the document. There are no version numbers or citations. The authors have provided the URL of where to get OMI, but nothing about the data product or when they acquired the data. Websites change and the Aura website will likely be lost after Aura is decomissioned. **The AVDC website is not dependent on the AURA mission.** More details about the data product need to be included in the publication for posterity. According to https://ozoneaq.gsfc.nasa.gov/products/ozone/, "Overpass (OVP) products are a weighted average of data within a defined range to a set of ground station locations." What the distance of the defined range? There is a 50km version that is very clear about its distance, but the README for the standard OVP products is not clear.

- If you're going to use a Loess fit, it would be good to introduce it somewhere. The Loess (not Lowess right?). **No Lowess, the Loess routine is different**. **I gave a reference.** fit seems unnecessary given that all your other plots use simple running means. **My mistake in writing. Lowess is similar in purpose to a running average but reduces the effect of outliers.** Perhaps you could explain why it is appropriate for Pan 180 NO2 in Figure 1, but not Figure 7. **All smooth curves are Lowess**

Figure 7 caption now reads. "Both OMI (black) and Pandora (red) then have a Lowess(3-month) running average applied."

Figure 11 now has "The smooth curve is Lowess(6 Months)."

- Figures are often scatter plots where the markers are so dense that often only a cluster is visible. The distribution of values is not decipherable. I recommend creating some sort of synthesis plots. **I am not sure what a "synthesis" plot might be. Other than adding a Lowess fit, I do not see the value of a different form.**

- The longer time-series would benefit from some sort of statistical analysis that quantified "agreement" and "disagreement" in the abstract. Right now, there is little more than visual analysis of datasets that were processed by others and downloaded.

I assume that you are referring to the ozone timeseries in Figs 12 and 13

- Similarly, it would be nice to quantify seasonality. XX% higher in JJA than DJF or similar. The conclusions starts with a paragraph about seasonality, but right now the manuscript simply says it is seen in one dataset and not the other.

That is correct. The amount of seasonal variation for TCHCHO varies depending on the site.

For most midlatitude sites, the seasonal variation is significant and occurs during summer. This has been added to the Summary

- The idea that Pandora would agree best if paired in time seems like an obvious conclusion. Figure 7 and analysis could be simplified by highlighting (or citing) the diurnal variation in the methods sections as the reason for time-pairing.

- The conclusion that OMI "underestimates" the degree of atmospheric pollution does not seem novel or quite accurate. OMI only "underestimates" pollution if we assume that overpass (13:30LST) is representative of the whole day. We know that vehicular emissions clearly peak at rush hour, so we would expect columns not to peak at Aura overpass (13:30LST). There is much evidence of this understanding in the literature. For example Anenberg et al. (doi:10.1016/S2542-5196(21)00255-2) use a series of ratios to translate overpass-time data to daily averages (see Figure S1 and discussion). The submitted manuscript should cite existing works highlighting the coincidence of the local minimum as a need for temporal co-sampling rather than highlighting this as a finding. **Added Lamsal, L. & Duncan, Bryan & Yoshida, Yasuko & Krotkov, Nickolay & Pickering, Kenneth & Streets, David & Lu, Zifeng. (2015). U.S. NO2 trends (2005–2013): EPA Air Quality System (AQS) data versus improved observations from the Ozone Monitoring Instrument (OMI). Atmospheric Environment. 110. 10.1016/j.atmosenv.2015.03.055**.

- The PGN website requests that "The PGN is a bilateral project supported with funding from NASA and ESA." be added to the acknowledgements. https://www.pandonia-global-network.org/home/documents/pgn-data-use-guidelines/ **Done**

Line-by-line notes:

- 35: the abstract discusses seasonal dependences, but isn't clear one what would be "big" or "little".

- 36-39: the abstract and conclusions assert that OMI is not observing near the surface, but the authors only show that it fails to capture seasonality. Could the failure have to do with reference sector correction? Or some other failure? Could you explain why you specifically think it is a failure to sense the lowest levels.

Since both Pandora and surface measurements (Wang et al., 2022) see the seasonal dependence of HCHO it is likely that OMI is not seeing into the boundary layer or is averaging over non-vegetated areas. The conclusion has been modified.

. For most sites, OMI does not observe the strong seasonal variation of TCHCHO that is clearly seen in the Pandora data and in surface measurements (Wang et al., 2022). The lack of OMI seasonal variation in TCHCHO at most sites suggests that OMI may not be seeing the lowest layers of the HCHO variation or may be averaging over non-vegetated areas. The amount of seasonal variation for TCHCHO varies depending on the site. For most midlatitude sites, the seasonal variation is significant and occurs during summer.

- 39-41: It seems obvious that excluding rush-hour from the comparison with 13:30 would be good. Why is this a noteworthy finding? (see discussion above)

The abstract now has a sentence: "Even when Pandora data is averaged between 13:00 and 14:00 hours local time OMI underestimates TCNO2." **One can see this in figure 7**

- 43-44: Agreement and disagreement should be put into some sort of context. Is there a pattern (under-estimating high values, over-estimating at low latitude) or is it random? What does agreement mean (bias within X? correlation above Y?)?

- 45: Does EPIC provide any particular meaningful result? **EPIC ozone retrievals are consistent with Pandora and suggests that the calibrations are consistent**.

- 54-55: Is the point that most methane that later forms HCHO comes from these sources? Or is this arguing that the majority of HCHO comes from this specific pathway (more so than isoprene + other methane sources)? **The results from this data paper do not address this problem**

- 70: Are these citations only for the first half of the sentence? If so, what are the citations for the rest and their ranking? **The citations apply to the entire sentence and have now been moved to the end**

- 79-104: This discussion does not mention surface monitors that sample in situ air or airborne in-situ measurements. The apparent focus is column integrals, but the sentences that introduces it simply says "typically measured by." Given that surface monitors and in-situ air sampling are more common, I think this needs clarification. **The sentence now reads "TCHCHO, TCNO2 and TCO in the atmosphere are typically measured by satellite and ground-based instruments."**

- 83: Given the timing of this submission, it is worth noting the TEMPO and GEMS satellites if this is a list. If this is really the methods section, then I don't think you use several of the data sources in this list. **This paper was written and submitted before TEMPO data were available. GEMS has few underlying Pandora sites. A paper discussing geostationary time dependent data is being written.**

- 117: seems weird to note that the website is Austrian. **Now reads "Austrian project website"**

- 119-122: Rather than providing file names, perhaps it would be better to discuss the meaning of the codes. For example, my read of the names is that you're using direct sun for the NO2 and skyscan for the HCHO. However, you do not discuss that. You also make no mention of data filtering of any kind. **I only used direct-sun data as mentioned earlier. The file names are all direct sun. The OMI overpass filtering is for data within 50 km of the Pandora location. All of the comparisons for OMI data are mid-day so that solar zenith angle filtering is not needed. Pandora data for single-day comparisons are selected for those days where it is likely that Pandora is observing clear sky.**

- 126-127: Did you use their other measurements? **I only used the direct-sun Pandora data as mentioned, rfus5p1-8, rnvs3p1-8, and rout2p1-8**

- 129-136: Using a single week a representative of day-of-week distributions is not a good idea. Friday July 8 might have been an outlier with winds blowing from a specific

source that was active but downwind on Thursday. Why is this specific week a good case study? **I was not doing a case study, just showing two examples. Other weeks are somewhat different. I picked a week in July and a week in September.**

- 133: "summer seasonal dependence" should probably be "summer peak"? **No**

- 197: To me, it looks like NO2 peaks in the DJF period in the lower left and lower right plots for OMI (green line). Perhaps the seasonality in OMI is larger than the seasonality in Pandora. **The TCNO2 peaks are in DJF and track each other for PAN and OMI with Pandora values > OMI values. The new Figure 7 shows this more clearly.**

- 206-207: "not statistically different" -- this stands out to me because I do not see any statistics anywhere. Nor do I see a discussion of how differences will be tested for significance. t-test? Welch's? Mann-Whitney? To see a statement like this, I would expect to see some data characterizations (mean+-std) for both datasets and/or the differences. **You are right. That sentence has been removed as being meaningless**

- 211-212: Can you at least state which Pandora sites you looked at? **See Table 1**

- 217: How did you measure the "cases of agreement"? For example, did you consider the uncertainty in either measurement? **I left out the criteria, now added within 10%.**

- 226-233: Is this the only site where the OVP file is not aligned with the Pandora site? If so, why is it a good site to show? **OMI Toronto East is Latitude: 43.740 deg. Longitude: -79.270 deg. The Pandora site is at 43.740^{O}N, -79.270^{O}W They are exactly aligned**

- 234: Why are 10a and 10b not 10 and 11? **They now are. Figures have been modified and renumbered**

- 234: Figure 10a -- what is the AVG for Pandora? Is this 13:50 to 14:50? Or some other window? **13:20 +/- 0:20 Local sun time (GMT + Longitude/15) Now in Figure caption**

- 235-241: The discussion of highway vs heating seems speculative. **The highway has nothing to do with heating. NO2 is produced continuously from traffic. The heating is seasonal but is mostly electrical heat and not from burning gas. Therefore, no seasonal local NO2 emission from heating.**

- 250: There is currently no discussion of Figure 11 or 12. **The discussion is above the figures (now 12 and 13) and is just to show that the calibration of the Pandoras is consistent with OMI.**

- 253: "good agreement ... at most sites" - is this comment based on the 4 sites shown? or was more analysis done? It is good that here there is a statement about what the difference is, but are we really talking about just one day? **More sites have been looked at and a separate paper has been written on ozone comparison for a special issue on EPIC. This is just a sample that was not selected for good or bad agreement.**

- 267: "for most sites" should be quantified. Of the N sites, M do not... I say this because the paper uses a small set of Pandora as representative. The conclusion could be interpreted as "most" Pandora sites. This may be true, but the paper does not show that. Instead it relies on a few (3?) case study sites.

The number of TCO comparisons has been expanded to 10 sites (Figs. 14 -15), all showing fairly good agreement. While this does not prove that all Pandora sites are good (6 are known to have stray light problems) it is representative of most of the sites in the Pandonia website that produce TCO data.

Citation: https://doi.org/10.5194/egusphere-2024-1216-RC2