Reviewer comments for:

The differences between remote sensing and in situ air pollutants measurements over the Canadian Oil Sands

This review is for the above manuscript submitted for publication in Atmospheric Measurement Techniques. The manuscript focuses on two key intercomparisons for data collected at or near Canadian Oil Sands: 1. in situ instrumentation measuring surface-level concentrations of multiple criteria air pollutants (SO_2, NO_2, $PM_{2.5}$) and remote sensing instrumentation measuring total column concentrations of the same pollutants, and 2. Vertically resolved wind profiles measured at site and modelled wind profiles obtained from an international dataset (ERA-5). The authors conduct these comparisons primarily by linking altitude-based wind direction with concentration data. The authors find the presence of higher concentrations linked to certain wind directions irrespective of altitude for all three pollutants, and attribute these higher concentrations to key local sources in the region. Further, the authors also note the dominance of higher concentrations linked to certain altitudes (while fixing the wind direction), and present it as evidence of a strong influence of winds at that altitude on the measurement. The authors also repeat the analysis above with modelled wind data, and find similar results, suggesting that in absence of measured wind data, modelled data can be used. The authors also note the differing temporal coverage of data from in situ surface and remote sensing instrumentation, and quantify the biases in sampled and actual variations in BLH and pollutant concentrations. While the presentation of the altitude-concentration plots resolved by wind direction were very interesting, and the findings seem plausible too, I think that this manuscript lacks an underlying methological framework needed to conduct these comparisons. Additionally, the current choice of presentation is very difficult for the reader to digest, and requires substantial changes as well. I suggest the authors **revise and resubmit** this manuscript for the following major issues.

1. **Issues with motivation of the paper.**

The authors chose multiple lines to motivate this work. This includes "applying satellite measurements to surface air quality applications" such as "air pollution monitoring" and "linking satellite air quality observations and surface in situ measurements". However, it is unclear how the authors have addressed these specific knowledge gaps in the manuscript. I recommend that the authors list clear objectives in the introduction and address those objectives step by step in different sections. This would also affect the abstract and the conclusions section, which would become more streamlined.

Additionally, the authors' literature review provides little motivation to compare in situ and ground based remote sensing data, with previous data showing bias <10% (Zhao et al., ACP 2019). If anything, the authors seem to have shown that in situ measurements combined with modelled ERA-5 data works really well, and that is a substantial finding. However, some of the other minutae can be removed from the conclusions then, with only key message kept in there.

2. **Lack of uncertainty analysis**

In a paper like this with substantial use of modelled data, I think a detailed subsection on uncertainties of the different pollutants modelled or measured using different instruments is needed in the Methods section. I suggest that the authors also add a small table of the pollutants/meteorological variables measured/modelled, the instruments used to measure/model them, and their average concentrations, standard deviations, and uncertainties. Also, for atmospheric pollutant data, geometric mean and geometric standard deviation may be more representative (please check). For wind direction, median wind direction (and not mean) would be more representative. Similarly, all instances of mentioning specific concentration or DU values in the manuscript should be accompanied with associated uncertainty (e.g., on pg 10 line 4).

3. **Comparison-based research without correlation coefficients?**

It was terribly inconvenient as a reviewer to read such an extensive comparison based paper without any discussion of correlation coefficients. At numerous points in the text (e.g., pg 10 line 1, page 12 line 7, and several others), authors not only compare sub-figures with each other, but sub-figures across different figures, all of which in my mind points to the inability of the figures in the main manuscript to convey the authors' message. This includes comparisons of NO2 and SO2 altitude patterns, remote versus in situ comparisons, and also comparison of patterns based on measured versus modelled wind data. I recommend using Spearman correlations instead of Pearson correlations because that coefficient better suits atmospheric measurement data (data at 0 and around 0 has value).

I strongly recommend the use of the correlation coefficients whereever possible, and a supporting figure in the supplement showing the relationship of the variables that are the basis of the coefficients. Additionally, there are several places where the arguments the authors make (e.g., pg. 20 lines 14-18, pg. 22 lines 28-29, pg. 24 lines 13-15, pg 25 lines 5-6, Conclusion pg 30 Lines 18-20) could have easily been backed by correlations. This is also true for the relationships of surface/column ratio to BLH (Sect. 5), where the authors state positive or negative correlations without a coefficient.

4. **Choice of meteorological variables and focus on wind speed for horizontal transport**
The authors have focused Sect. 4 on a discussion of horizontal transport using wind speed and a later section on boundary layer height (Sect. 5). However, it is unclear to me that local wind speeds are representative of horizontal transport, and I believe that the only way to really model that transport is using 3D CTMs. The lack of references in Sect. 4 is also telling. Can the authors cite previous work that systematically shows local wind speeds and directions are representative of transport on the 10s of km scale? This would address the linkage of the wind speed discussion to plumes (e.g., pg 1 lines 27).

This critique above also made me question the broader theme of the paper: relationships of pollutant concentrations with specific meteorological variables. Given that they have access to detailed meteorological data, including measured temperature and several other modeled meteorological variables from ERA-5, did the authors consider conducting a PCA analysis of all meteorological data to see what are the key components that explain the data? If the variables presented here, wind direction, wind speed, and boundary layer height separate out as clear and separate components, this analysis makes sense. However, if other variables stand out such as temperature or precipitation, then the authors are making the age-old mistake of drawing conclusions from correlations when really, they want to get as close to causation as possible. And the best way to do that would be to identify the key PCA components of the underlying meteorology and then interpreting the data based on those components. At the least, the authors should consider the relationships of pollutant concentrations with altitude-resolved ventilation coefficient (see Fig. 6, Gani et al., ACP 2019) that combines the effects of wind speed and boundary layer height since, like all meteorological variables, they actually affect pollutants interlinked and not separately.

5. **Presentation of data**
There are several critical issues with the current style of presentation that need to be addressed.

- The authors have analyzed the data before presenting it. If anything, Section 6 should be Section 3, the first results section. In this current format, the authors jump straight into comparing remote sensing and insitu data using a complicated figure, that combines wind direction, concentration, and altitude, which is a difficult and unusual transition.
- The authors have conducted so many different analyses with different purpose that it is hard to keep track of all of them. I suggest adding a table at the end of the methods section listing the relevant pollutant, method used for VCD measurement, method used for in situ measurement, and if not, model used as replacement. This would allow you to cut a lot of text used for describing each analysis pipeline.
- Unnecessary length of manuscript: the manuscript is unnecessarily long, and large sections of the paper add little to the manuscript. For example, most sub-figures in Sect. 5 do not add much to the manuscript and could be placed in the Supplement. Also, Figs. 3-5 could be styled along the lines of Fig. 9, and separate detailed figures could be placed in the supplement. Additionally, as discussed above, sections 4 and 5 have relatively weak foundations, and should be shortened significantly, or large parts (e.g, Figs. 6-8) moved to the supplement. Only focus on the core messages for these sections instead of describing them in much detail.
- Several method descriptions are in other sections (e.g., pg 7 line 16-17, pg 10 lines 14-16, pg. 14 lines 6-10, pg. 20 lines 10-11, 18-20). This made reading the manuscript difficult. The authors should bring together all such descriptions back to the methods section.
- Before Fig. 10, the authors based all arguments on the surface to column ratio, but after, switch to column to surface. This is unnecessary, and should be kept consistent.

6. **Logical extrapolation and lack of explanation for multiple phenomena**

The authors have presented several phenomena (e.g., pg 11 lines 16-18, pg. 19 lines 1-2, pg. 28 lines 2-3) that they have not explained or explicitly figured out ways to address or are themselves perplexed by. I suggest going back to the literature to identify such instances happening elsewhere, or find explanations for why its happening here. One such instance seems fine to accommodate but multiple such instances in the same manuscript make it seem like the authors are simply putting data together and not interpreting it.

Minor comments

1. Pg. 3 Last sentence on lines 19-20 need a citation

2. Pg. 4 Last sentence on lines 16-18 should be moved to Sect. 2.5 and a reference to Sect. 2.5 should be made here
3. Pg. 5 line 10 Does the Pratmo model use wind data and other meteorological data. Explain.
4. Pg. 5 Last sentence on lines 14-15 what does “comparable” refer to? Comparable to what? Detail the % here.
5. Pg. 5 lines 22-23 Define fine and coarse mode by size. Also, detail the uncertainty. See comment 2 for broader issues with uncertainty.
6. Pg. 6 line 16 What does “both” refer to? Name the two polarizations.
7. Pg. 6 line 19 Why was this specific wavelength used even though other wavelengths were mentioned?
8. Pg. 9 line 21 Fig 3d suggests its 340 deg and not 280 deg in the cold season
9. Altitude figures e.g., Fig 3 the authors could draw average BLH estimates on this figure itself
10. Pg. 11 lines 14-15 cite Figs. 3d-e with this sentence
11. Figs. 3c, 3f show data for all directions; if not in the main manuscript, then in the supplement. Also, pg 12 line 1 2c is likely 3c.
12. pg 13 line 15-pg 4 line 2 phenomenon happening for NO2 but not for SO2. Why?
13. pg. 14 line 16, pg. 22 lines 10-11 add figure reference
14. Make fig. 6 in steps of 100 m and put other Figs. in the SI
15. pg. 22 lines 4-5 I disagree with the authors’ assessment re: “a general agreement between the results based on WindRASS (see circles symbols) and reanalysis (see square symbols) data…” There are visibly influential deviations in 4 out of 6 subfigures.
16. Fig. 10 need to show 280 deg as well. Could add in the SI

References

1. Zhao et al., ACP 2019 https://doi.org/10.5194/acp-19-10619-2019
2. Gani et al., ACP 2019 https://acp.copernicus.org/articles/19/6843/2019/