We thank Referee #1 for the suggestions and comments on this work. The paper has been revised based on the referee's comments (in red). Detailed answers have been provided (in blue).

This study compares the difference between in situ and remote sensing observations for air pollutants in from different perspectives over Canadian Oil Sands. The manuscript is pretty informative. However, the key contribution and findings are not clear, the large number of acronyms are difficult to follow, and the text is really long and it is suggested to keep it succinct.

Done. We shortened the text by moving Section 4 to the Appendix, and reduced the number of acronyms.

Detailed comments are given below:

Abstract

Line starting with 'Compared to an in situ instrument that provides…', that applies to nadir sensors, limb and occultation sensors provide vertical profiles.

Done.

Compared to an in situ instrument that provides air quality conditions at the ground level, most remote sensing instruments (nadir viewing) are sensitive to a broad range of altitudes, often providing only integrated column observations.

Line starting with 'Elevated SO2 VCDs are clearly observed…', are observed or were observed?

Done.

Elevated SO_2 VCDs were clearly observed for times with south and south-eastern winds, particularly at 200–300 m altitude (above ground level).

Do not mix-use British or American English, like analyzed and modelled, keep consistency

Thanks! This has been addressed with our best efforts.

Line starting with 'In addition to measured wind…' Maybe not all readers across the world know what ERA-5 reanalysis is, it is a good practice to give its full form at the first appearance.

Done.

In addition to measured wind data and lidar observed boundary layer height (BLH), modelled wind profiles and BLH from ECMWF Reanalysis v5 (ERA-5) have been used to further examine the correlation between column and surface observations.

Line starting with 'The results show that…' The findings are not significantly strong or exciting to intrigue the readers.

We have rephrased it as suggested.

The results show that the height of emission sources (e.g., emissions from high stacks or near surface) will determine the ratio of measured column and surface concentration values (i.e., could show positive or negative correlation with BLH). This effect will show impacts on the comparison between column observations (e.g., from the satellite or ground-based remote sensing instruments) with surface in situ measurements.

Introduction

Maybe not 'even' in polluted urban areas, my guess is relatively high correlation is expected in un- or less polluted areas for NO2, SO2, except for O3.

We think the referee was referring to this sentence:

Alternatively, a direct comparison of ground-based VCD observations with surface concentrations does not typically produce high correlations ~~even in polluted urban areas~~ (Dieudonné et al., 2013).

For un- or less polluted areas, the satellite column observations are more challenging, due to a worse signal-to-noise ratio. So, we fully agree with the referee that this issue is not only for urban but almost "everywhere". So, we deleted "even …".

Authors may want to use acronyms less to help readers follow the storyline clearly instead of wondering the meaning of individual acronyms.

Thanks for the suggestions. We have reduced the number of acronyms, such as "WBEA", "AOSR", "DIAL", "U340", "ECCC", "PGN", etc.

Ln 15, we compare or compared?

Done.

Besides trace gas pollutants, we also compared and studied the differences between surface $PM_{2.5}$ observations with remote sensing AOD data. $PM_{2.5}$ concentration is also one of the three indicators in the Canadian Air Quality Health Index (Stieb et al., 2008).

Ln 20, the cause of what?

Done.

Utilizing information on detailed vertical wind field profiles and BLH, this work illustrates the cause of ~~and demonstrates~~ the differences between ground-based remote sensing and in situ observations, as well as studies the ratios of the surface to column values for various wind directions and BLH values.

Ln 20 again, what is the major difference of definitions on ground-level and in situ observations?

The ground-based remote sensing observations provide the total column of pollutants, whereas the in situ observations provide pollutant concentrations on the ground. We have deleted the sentence and replaced it with the following paragraph.

In short, this study is focused on the difference between total column (measured using ground-based remote sensing technique) and surface/ground-level concentrations (by in situ observations) of air pollutants. Measurements of wind profiles and BLH were used to examine their impact on that difference. The possibility of using modern reanalysis-modelled data instead of direct measurements of wind profiles and BLH were also explored. ~~Utilizing information on detailed vertical wind field profiles and BLH, this work illustrates the cause and demonstrates the differences between ground-based remote sensing and in situ observations, as well as studies the ratios of the surface to column values for various wind directions and BLH values.~~

Ln 25, this bit 'integration period differences' comes out of sudden, readers may want to know why do we care about the integration period.

Thanks! We have rephrased this part.

In Sect. ~~6~~3, the ~~integration period~~ observation conditions (mainly weather) induced differences are evaluated.

Section 2

It would be lovely to have a map of the study area that makes the paper more illustrative.

Figure 2 in the manuscript was the map (also plotted with satellite SO2 and NO2 VCD data). To make it clear, we have moved it to this section.

Fort McKay is a small town (population of 600) surrounded by seven oil sands surface and two in-situ mining facilities to the north and south. Satellite maps showing the observation site and surrounding oil sands areas are shown in Fig. 1.

Ln 5 What would the inspiration of this study be as the Fort McKay is a very small town? How would people be informed and apply the findings here to other polluted regions, as most heavy pollution that threatens people's life occur at densely populated urban areas.

We thank the referee for this insightful thought. As already provided in Section 2, one of the major outcomes of this work is that we found the pollution level at Fort McKay is largely dependent on the wind direction (and some other meteorological factors, such BLH).

Thus, the pollution level at ~~Fort McKay~~ the site is largely dependent on the wind direction. Therefore, the planning of communities close to industrial activities should consider regional climatology factors, such as prevailing wind directions.

Ln 5 I do suggest the authors use acronyms less to make the life of readers easier.

Done. For example, we removed acronyms for the Wood Buffalo Environmental Association (WBEA), Nd: YAG, U340, AOSR, etc.

Section 2.1

Just curious, would it be considered as a 'recently developed' instrument as it has at least ten years of history?

Done.

The Pandora spectrometer is a ~~recently developed~~ ground-based remote sensing instrument that measures solar and sky spectral radiation in the UV and visible part of the spectrum (Herman et al., 2009; Szykman et al., 2019).

Ln 30 What is U340 bandpass filter? What is ECCC? What is PGN?

Done. We agree with the referee there were too many unnecessary acronyms and technical details.

~~*The Pandora instrument consists of an optical head sensor, mounted on a computer-controlled sun-tracker, and connected to a commercial Avantes array spectrometer by means of an optical fibre. To allow for the detection of different absorbers, the instrument periodically measures UV spectra with the U340 bandpass filter with a cut-off limit at 380 nm on and off, with an interval of about 90 seconds. The 306–330 nm spectral interval was used for SO_2 spectral retrievals (ECCC research retrieval) and the 400 to 440 nm interval was used to retrieve NO_2 (PGN official retrieval, version nvs1p1-7) (Fioletov et al., 2016; Zhao et al., 2020).*~~

Ln 5 no. 104, does this serial number really matter?

Yes. There were two Pandora spectrometers been deployed at the site (no. 104 and 122), as described in this paragraph. The PGN and ECCC teams spend efforts in making sure the bias between the two systems has been corrected and accounted for. I.e., to have a uniform time series from 2014 to 2019. Also, when Pandora no. 122 was used, it also performed multi-axis observations (not included in this work), which reduced its direct sun (total column) sampling frequency. We could not provide too much technical details here, but still want to inform the reader of these important instrumental changes.

Ln 10 '…is given by (Herman et al., 2009)' is not a good format for reference.

Done.

A detailed description of the Pandora spectrometer and its total column NO_2 retrieval algorithm is given by Herman et al. (2009).

Ln 10 How does this Pratmo box model work and why it can remove NO2 in the stratosphere? Sometimes it's not the case the more information the merrier.

Done.

In order to isolate tropospheric NO_2 VCD from the total column VCD measured by Pandora, stratospheric NO_2 partial columns were subtracted from Pandora measurements (following the method described in Zhao et al., 2019).

Ln 15 What does the retrieved SO2 refer to? Is it from OMI? If it's from a project, we need a reference to it.

Done. The SO_2 column data is also retrieved from Pandora.

For SO_2 data, as the only sources are near the surface (as no comparable SO_2 quantities were in the stratosphere during the analyzed period), the retrieved total column SO_2 (SO_2 VCD from Pandora observations) are directly been used in this study.

Section 2.4

What is Q-switch Nd:YAG?

Done. Neodymium-doped Yttrium Aluminum Garnet (Nd:YAG) laser. To avoid more unnecessary details, we only provided the reference for the system.

The lidar simultaneously emits two wavelengths laser light (1064 and 532 nm, Neodymium-doped Yttrium Aluminum Garnet laser, see references in Strawbridge, 2013) at energies of approximately 150 mJ pulse^{-1} wavelength^{-1} and detects the backscatter signal at 1064 nm and both polarizations at 532 nm.

Section 3

What are coincident observations? Do the authors mean overlapped observations?

Yes.

Thus, only coincident (overlapped) observations from both remote sensing and in situ instruments were included in the analysis.

P9 Ln 5 What are selected wind directions? How did the authors do the selection, do you mean upwind or prevalent direction?

Done. It was described in the caption "*The white dashed lines show the centre of the wind sectors.*". Since this figure is now Fig. 1, we modified its caption accordingly.

Figure 1. Satellite maps (© Google Maps) of the Athabasca Oil Sand Region ~~(AOSR)~~ masked with satellite observations ~~selected wind directions~~. The Pandora spectrometer, sunphotometer, WindRASS, lidar, and in situ instrument were located at the observation site represented by a white circle. The two largest upgraders in the mining areas are shown by red triangles. The white dashed lines show the centre of the wind sectors. Maps are masked with pixel averaging of total column SO_2 and tropospheric column NO_2 (2018–2021) from TROPOMI satellite instrument (McLinden et al., 2020).

Conclusion

It would be better to have a concise conclusion

Done. We revised the conclusion to be more concise.

The magnitude of the SO_2 VCD and surface concentrations reach their maximum for winds from 160° ± 30° directions at 200–300 m altitudes ~~(which is about 38% and 27% higher than the value for winds from the same directions but of near-surface winds, respectively)~~.

As a result, NO_2 VCD shows a more uniform sensitivity to winds from near surface to up to 300 m (peak value at 260–290 m altitude), while in situ measured NO_2 surface concentrations show a strong sensitivity to near-surface winds (peak value at 60–70 m altitude). ~~*In cold seasons, the NO_2 surface to column ratio from 160° direction changes from 39 ppbv/DU at 70 m to 28 ppbv/DU at 200 m.*~~

~~*The horizontal transport sampling differences show that local sources have the largest impact on observations when the wind speed is low and the pollutants are not transported far from the source (as in the case of some NO_2 emissions). For elevated sources (e.g., SO_2 emitted from high stacks), the moderate wind is substantial to bring the pollutant to the measurement site. Compared to in situ instruments, remote sensing observations are more sensitive to higher wind speed conditions, i.e., transported pollutants.*~~

~~*In general, due to the complex nature of PM, the conversion from $PM_{2.5}$ to AOD (or vice versa) is not straightforward.*~~ *Here we show that linking these two measurements could be even more complicated, as they have more sampling differences than observations of trace gases.* ~~*On the positive side, both remote sensing and in situ observations show consistent uniform sensitivities to the wind speed and direction from near surface to 300 m altitude, indicating the aerosol loads in this region are more uniformly mixed than SO_2 and NO_2.*~~

Further results show replacing measured wind profiles and BLH by ERA-5 data could also preserve these features. Thus, ~~*the boundary layer height and wind profile data from*~~ *these ERA-5 data also can be utilized to reveal pollutants' vertical distribution and mixing conditions, which can be used as critical information when converting remote sensing column data to surface values.*

This analysis of surface to column ratios also shows that the column values cannot be converted to surface by just one value of the ratio. Depending on the wind direction (and season), the ratio for directions related to the pollution sources could be a factor of two larger than these from "clean" directions.