Review of: Highly resolved mapping of NO_2 vertical column densities from GeoTASO measurements over a megacity and industrial area during the KORUS-AQ campaign (Choo et al., 2022)

The manuscript discusses results from the KORUS-AQ campaign and demonstrates for the first time the NO_2 VCD horizontal distribution over highly urbanized / industrialised regions in South Korea, based on airborne imaging data. The scientific content of the paper fits within the scope of AMT, although the study lacks novelty. The paper is short on the actual data retrievals and focuses more on the interpretation of the NO_2 spatiotemporal variability over the different regions. Regarding the VCD retrievals, a number of choices/assumptions are made without proper argumentation which are not consistent with other similar studies that included extensive explanations/analysis. This can affect the results and the made conclusions, especially regarding the absolute VCD values. I'm referring for example to the fitting interval, reference SCD, surface reflectance product. Also the error analysis has some serious flaws and the 'validation' lacks a lot of details in order to assess the validity of the comparisons. Overall, I would recommend publication in AMT. However, major revisions (detailed below) need to be conducted in the paper before publication. Please note that all questions below should not only be addressed in the author's reply, but also in the manuscript.

General comments

-Please extend section 2.1. with campaign information. Some information about the campaigns is scattered in the manuscript, but a clear campaign section shortly discussing the number of flights, time and duration of flights, ROI, SZA change during flights, environmental conditions, e.g. cloud fraction, etc. is missing while it would improve interpretation. Also a table would help indicating the different flights, their time of flight and the region of interest. This would also help to understand certain choices made for the comparisons (see later comments on that).

-p.5 l.150: One reference is used to analyse the whole data set if I understand it correctly. Three questions that should be properly addressed: 1) do you use one spectrum or an average of spectra over your reference area in order to improve the SNR? ; 2) What is the residual amount in the reference spectrum used and how was this determined. Do you use the value 6.751 x 1015? This is not really clear from the manuscript. This value seems high for a considered background area over the ocean and can result in an overestimation of all retrievals shown in this work. Have you compared with the OMI value retrieved for that day or an average + st. dev. OMI value to get an idea on the NO_2 variability for this area? ; 3) for airborne hyperspectral imagers often a daily reference is preferred as the spectral properties of the instrument can change resulting in an along-track and across-track drift. Please prove that this is not the case with this instrument and the right choice was made, i.e. by showing RMS of the fit on the day when the reference was taken and RMS for flights further away from the reference date.

-p.5 l.155: This NO_2 spectral window is not very optimal and moreover very narrow, while the instrument properties allow a more optimal and larger window for NO_2 fits, e.g. like the 425-490 nm recommended in the literature. Please show some fit results with this larger window (or similar larger window) and show the VCD differences with the narrow window to clarify the impact of the choice made and also give a clear motivation for choosing this non-optimal window 425-450 nm for your retrievals.

p.6 l.175: It is shown in several studies, some of them cited in this work, that surface reflectance can have a very strong impact on the AMF. Is GeoTASO not absolutely calibrated in order to derive a surface

reflectance product from the instrument itself? If not, why isn't the MODIS product MCD43A1 or MCD43A3 used like during the LISTOS campaign? It provides a much higher spatial resolution matching better the spatial resolution of your retrievals and it is proven to be a reliable product. I'm worried that the coarse resolution product you have chosen strongly impacts your AMF, thus also your VCD retrievals.

Sect. 2.5: You clarified later on that SSA and AOD are also derived from CMAQ. This should be described here as well, besides the vertical NO_2 profile.

Sect 3.1: The spatial binning (0.01°) has not been explained in the manuscript, but only mentioned in the relevant figure captions. Do you bin spectra prior to the retrievals to increase the SNR or do you bin the VCDs afterwards? In the latter case, why do you bin the data (native resolution of 250 m) to this coarser resolution? You are losing some spatial detail while NO_2 is a species with strong spatiotemporal variability.

p.7 l.204 and Figure. 4: I would like to see a zoom on figure 4 on the expressways (eventually at the native spatial resolution). In Figure 4, it is difficult to judge if there is a stronger NO_2 pattern downwind than upwind from the expressway line sources, but this is something that should be present in the data if the made statements are correct. You could also do it for the afternoon with a different dynamic scale for the VCD map.

P.7 l.206: So it means the background NO_2 VCD is around 1×10^{16}? This seems pretty high (see also earlier comment on the SCDref). Can you demonstrate that this is somehow consistent with OMI/TROPOMI retrievals? Please do the same for statement on p. 9 l. 255

p.9 l.364: The error on the SCDref is an important source of uncertainty as well, not taken into account in this study. Please include.

p.9 l.273: Please clarify how these percentages were obtained. This is unclear. An important uncertainty regarding the AMF is also the uncertainty related to the a priori NO_2 profile shape. This is mentioned as well in the conclusion but it is completely ignored in this uncertainty study. Please include.

p.10 l.283: 7.3% uncertainty on the AMF is really on the optimistic side and not in line with other studies that report around 15%-20%, or larger. Please revise your calculations and/or explain. Just to clarify: it is perfectly fine if you obtain other numbers than in other studies but this should be clarified based on an in-depth analysis. Like mentioned in the previous comment it is unclear where the percentages are coming from, while in other studies these are based on discussed sensitivity tests. Are the numbers provided in Table 6 an average for the flight on 9 June 2016 or what does it represent exactly?

Sect. 3.3. <u>all</u> the performed comparisons should report how many measurements/data points are compared (eventually in the plots), in order to assess how statistically relevant the comparisons are.

p.10 l.308: Elaborate on this part, eventually with a plot, as many things are unclear! Why are OMI NO2 VCDs only available on 10 June? How did you perform the comparison exactly? Setting a radius is a valid strategy when comparing ground based measurements with airborne or spaceborne data, but not when comparing airborne with satellite. In the latter case you should average the airborne pixels within the larger satellite pixel footprints in order to perform a fair comparison. The time constraint can be kept but could be extended to 1 hour to increase amount of compared measurements. Slope is 0.43: not clear without plot what is under- / overestimating what.

Sect. 3.3.1: first sentence should be rewritten as it is unclear. Again not clear: do you compare all pixels within 0.5 km and 30 min with the station measurement or do you average the airborne pixels and do you compare the averaged value? You have hourly station measurements and many flights over the whole campaign period (I think? Much details are missing on flights), but you don't seem to have many data points. Please clarify this. Why not focusing on all possible comparisons between airborne and ground-based measurements instead of restricting to AM and PM in order to improve your statistics. Like it is presented now: not really statistically relevant and without any proper discussion on the non-linear relation between columns and surface concentration, I suggest to leave this section out of the paper.

p.11 l.325: Why do you restrict to 5, 9, 10 June? You should clarify the reasons. If you only had flights over this region on these dates that is a proper explanation, but these details have not been provided.

Figure 5: I have a hard time to understand the AMF you obtained. Normally the AMF should be highly correlated with the surface reflectance which isn't really the case here. One (partly) explanation is maybe the rough albedo product used. But this doesn't explain the strong striping in the AMF, especially in the western part. Can you elaborate on this in the manuscript to explain the reasons?

Figure 10: I would expect that PANDORA retrievals are higher than the GeoTASO retrievals as the ground-based measurements are more sensitive to the bulk of NO2 close to the surface, but it is the other way around. This could be related to the choices made in the VCD retrievals like mentioned in earlier comments which can strongly affect the absolute values. Are these findings consistent with other GeoTASO-PANDORA comparisons like from the DISCOVER-AQ or LISTOS campaign?

Minor comments

p.3 l.79: SWING has not been operated over Antwerp. The example of Antwerp shown is a simulation. SWING has been operated on a UAV over Romania. It is also better to refer to https://amt.copernicus.org/articles/11/551/2018/amt-11-551-2018.pdf

p.3 l.81: "regional radiative transfer models" → I think you are referring to regional air quality models here?

p.3 l.89: How is transboundary pollution defined here?

p.4 l.120: Be more clear on how exactly you do the comparison: if an airborne overpass matches the 30 min constraint, do you average spatially all the pixels within 1 km radius? Or do you compare the different individual pixels with the Pandora retrieval? Same comment for p. 5 l. 131.

p.5 l.143: What is the exact reason of flying so high to measure tropospheric species? At 3 km you would already be well above the PBL. And while indeed the higher flight altitude allows to measure the NO2 in the free troposphere as well, it has the drawback that you are losing some sensitivity towards the NO2 in the PBL and especially at the surface due to larger scattering and absorption probability.

p.7 l.210: Also the PBL plays a role and it's rise in the afternoon + transport/accumulation of emitted NO_2 in the PBL during the day. But the latter is already stated later on in this paragraph.

p.8 l.240: I agree with the statement that they can be highly complementary, but it should be clarified to the reader that the relation VCD-surface concentration is highly non-linear. Depending on meteorology it is possible that strong columns are detected while surface concentrations are low. Especially in case of industrial emissions, where the emissions happen at a certain altitude (stacks/chimneys).

p.13 l.376: Link doesn't work

Table 2 and 5: Average mileage is average mileage per car, per day?

Technical corrections

p.1 l.24: please replace 'data' by 'VCDs'

p.2 l.31: you might add 'domestic heating' as well

p.2 l.38: cites --> cities

p.2 l.45: what is the point of mentioning UTC here if you don't specify a time?

p.2 l.53: is there a need to repeat here the spaceborne sensors? This was done in the previous paragraph

p.3 l.90: replace “from May to June 2016.” By “…, organized from May to June 2016.”

p.3 l.94: In this study, NO2 VCD retrieval was conducted using solar backscattered radiance observed from GeoTASO over South Korea during the KORUS-AQ campaign --> This sentence is redundant, please remove.

p.4 l.106: Amnyeon --> Amnyeon region?

p.4 l.118: O3 → O_3

p.4 l.122: “. “ after Notably

p.5 l.132: “.” at end of sentence.

p. 6 l.184: stratospheric NO2 → stratospheric and free tropospheric NO_2

p. 8 l.219: decrease --> decreases

p. 9 l.268: Boersma et al. 2004 --> Boersma et al., 2004

Figure 1 caption: please repeat the campaign period here and the number of flights.

Figure 3 caption: “.” at end of sentence.

Figure 4 caption: remove 'the' in 'the each panel'

Figue 6 caption "into to" --> please correct + what is the meaning of the color-coding of the arrows?