Anonymous Referee #2:

Thank you very much for the useful and constructive remarks. As described below, we have modified the manuscript according to suggestions and provided clarifications where necessary. We hope that the revised manuscript has improved in respect to the original paper. Please find a rebuttal against each point below.

Black, bold, italic: Referee's comments

Black: Author's reply

Changes in the original discussion paper are highlighted in yellow and attached below

1) p.3 l. 31: Is this really the first? Around the same time: Evaluating Sentinel- 5P TROPOMI tropospheric NO2 column densities with airborne and Pandora spectrometers near New York City and Long Island Sound, Laura M. Judd et al.; https://doi.org/10.5194/amt-2020-151; I would suggest deleting this comment and possibly include a reference to this paper within the manuscript.

The study https://doi.org/10.5194/amt-2020-**151** was indeed submitted to AMT in the same week as the study under review (https://doi.org/10.5194/amt-2020-**148**). We have adapted the paragraph in the manuscript and we have added a proper reference now it is available:

This is one of the first publications assessing TROPOMI NO_2 retrievals over strongly polluted regions based on the comparison with airborne remote sensing observations and it is one of the first airborne spectrometer data sets coinciding in space and time with a large amount of fully sampled satellite pixels. At the same time the study of Judd et al. (2020) on the Long Island Sound Tropospheric Ozone Study (LISTOS) campaign in the New York City/Long Island Sound region has been submitted. Earlier studies reporting on the validation of spaceborne observations based on airborne spectrometer data, such as Heue et al. (2005), Constantin et al. (2016), Lamsal et al. (2017), Broccardo et al. (2018), and Merlaud et al. (2020) have shown high potential but are scarce, mainly due to the relatively large pixel footprint of TROPOMI's predecessors with respect to the area that can be covered with an airborne mapping spectrometer.

However, we would find it more than fair that in the study https://doi.org/10.5194/amt-2020-151, the statement "This is the first airborne spectrometer dataset to be used to evaluate the TROPOMI tropospheric NO2 product." would be changed as well and a proper reference to this study would be added.

2) p. 4 l. 16: The TROPOMI tropospheric columns are up to _12km. There is NO2 above 6.5km, the NO2 profile is not 0. Over cities and enhanced areas this will not be a big factor, but this should be discussed and mentioned. A typical amount of NO2 from 6.5 to 12km over Belgium would be useful to mention – maybe using CAMS or TMP.

Thanks for the interesting comment. As suggested we checked the partial NO_2 column between 6.5 km and tropopause, based on the interpolated TM5-MP a priori NO_2 profiles over Brussels and Antwerp for the four campaign days. The TM5-MP a priori NO_2 profiles are provided in Fig. 3a (the figure was updated in the manuscript as there was a mistake in the plot in the conversion from VMR to partial columns). Note that the tropopause is around 16 km (defined from the temperature profile) instead of the suggested 12 km on these days. The partial column between 6.5 and 16 km ranges between 2.8 and 4.7 x 10^{14} molec cm^{-2}. We can also refer to a similar question/answer (comment 2) in https://amt.copernicus.org/preprints/amt-2020-151/amt-2020-151-AC1-supplement.pdf. Judd et al. (2020) reports a partial column of 2 x 10^{14} molec cm^{-2}, but the aircraft altitude is 9 km instead of 6.5 km.

The impact on the comparison and conclusions is expected to be small and would generally increase the bias between the TROPOMI NO_2 VCD product and APEX retrievals, as the latter could be underestimated. Note that the effective impact is difficult to assess as airborne measurements are in fact sensitive to NO_2 above the flight altitude of 6.5 km, however, indeed with reduced sensitivity as can be observed in Fig. 3b. Retrieved SCDs are the sum of the measured differential slant column and the residual amount of NO_2 in a reference spectrum acquired over a clean area during the same flight (SCD = DSCD + SCDref). The residual amount in the reference spectrum is a tropospheric VCD (corrected for the stratospheric content) estimated in this work from mobile DOAS measurements (but can also be derived from for example a model or MAX-DOAS observations like done in a number of other studies). In principle SCDref contains implicitly a contribution from the upper troposphere (> 6.5 km). However, also these measurements, similar to MAX-DOAS measurements, have a reduced sensitivity to the upper troposphere. In case there are temporal/spatial changes in the NO_2 field in the upper troposphere between reference area and measured area this should be implicitly measured in the DSCD.

We have added a discussion in the manuscript in Sect. 5.2.2: "Note that APEX observations have reduced sensitivity to the NO_2 above the aircraft altitude of 6.5 km (see Fig. 3b), while the TROPOMI NO_2 VCD is defined up to the tropopause (approximately 16 km on the campaign days). The TM5-MP NO_2 partial columns between 6.5 and 16 km range between 2.8 and 4.7 x 10^{14} molec cm^{-2}. Retrieved APEX SCDs are the sum of the measured differential slant column and the residual amount of NO_2 in a reference spectrum acquired over a clean area during the same flight. SCDref is derived from a tropospheric VCD, estimated in this work from mobile DOAS measurements. In principle SCDref contains implicitly a contribution from the upper troposphere. However, also these measurements have a reduced sensitivity to the upper troposphere. In case there are temporal or spatial changes in the NO_2 field in the upper troposphere between the reference area and observed area, this should be implicitly measured in the DSCD. As the amount of NO_2 in the upper troposphere appears to be small compared to the total column over polluted sites and as the APEX retrievals still have some sensitivity to it, we expect any impact on the comparisons to be minimal."

3) p. 4/5 I think the cloud fraction should be mentioned. It's mentioned that for the flights it was mainly clear sky, but what is the range of the cloud fractions for the TROPOMI observations? Some of this could be of course due to aerosols, but I think it would be good to know the cloud fraction (nitrogen dioxide window) assumed in the TROPOMI retrieval. I just noticed this is mentioned later in Sect. 4.3.3 , but it would be good to include it in this section.

Ok, we have specified the TROPOMI cloud fraction in Section 2: "Flights took place in mostly cloud-free conditions and on days with good visibility. For flights on 27 to 29 June, there was a cloud fraction of less than 1% for the TROPOMI NO_2 retrieval window at 440 nm. Only on 26 June (Flight #1), conditions were not fully optimal with few scattered clouds and some light haze and aerosols (cloud fraction of 12%)."

4) p.10 l. 1 Could the difference of the AMF come from the different height? APEX is from the surface to 6.5km; for TROPOMI it's higher.

It has certainly an effect on the tropospheric AMF: due to scattering and absorption, the sensitivity to NO_2 decreases towards the ground surface and the decrease in sensitivity is stronger with increasing platform altitude due to the larger scattering probability above the absorbing layer. See for example Fig. 8 in https://doi.org/10.5194/amt-12-211-2019.

We have added the following sentence to clarify this: "This can be partly explained by a stronger decrease in sensitivity with increasing platform altitude due to the larger scattering probability above the absorbing layer."

5) Sect 4.3.2 Albedo: the TROPOMI AMF could be re-calculated using the APEX albedo and the impact can be directly estimated. I think the study would benefit from looking at the impact of correcting for the albedo. I'm not sure if this would be possible to do within a reasonable amount of time, if this is too time consuming, just mention it at least.

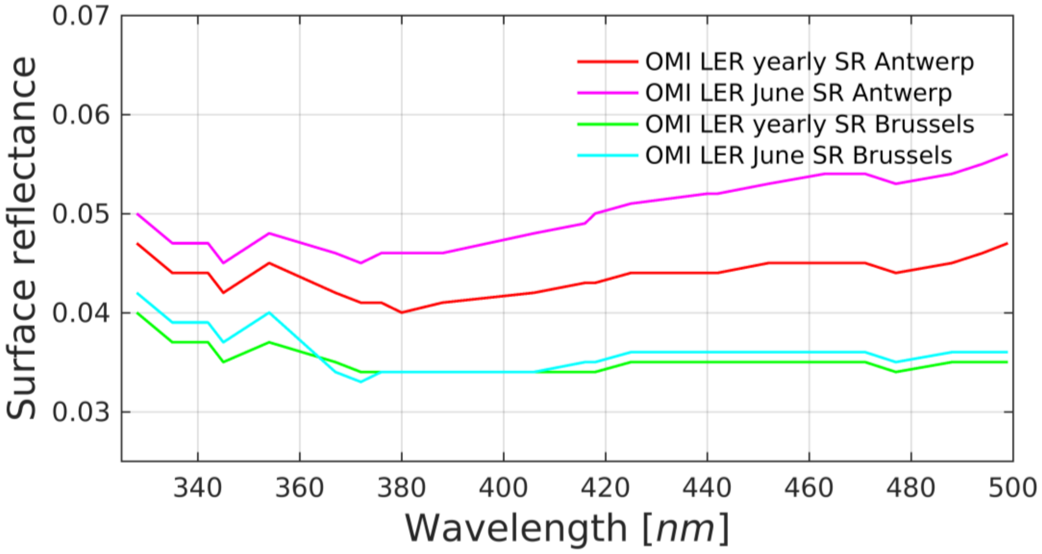
Recalculating the TROPOMI tropospheric AMF (and NO_2 VCDs) based on the APEX albedo is not as straightforward as for example replacing the a priori NO_2 profiles based on the provided AKs, especially as the main authors of this study are not involved in the operational TROPOMI NO_2 retrievals. We agree it would be an added-value to directly study the impact on the NO_2 retrievals, instead of the albedo comparison tests done in this study. However, we propose to keep this as part of a future study: new APEX flights are foreseen over the two target areas in summer 2021 (and also other validation activities will take place later this year and next year → see reply to comment 8). As the TROPOMI LER (under development – see last paragraph on page 11) should be available by then, we suggest to compare the new APEX retrievals with the TROPOMI retrievals based on the initial OMI LER and new TROPOMI LER product to asses the impact.

To make clear that a study on the direct impact of the albedo is in the pipeline we have added the following at the end of the last paragraph on page 11: "New APEX validation flights over the Antwerp and Brussels region are foreseen for summer 2021 and will be valuable to assess 1) the retrieval impact of replacing the OMI LER by the TROPOMI LER, and 2) the v2 reprocessing of the TROPOMI NO_2 product".

6) Sect. 4.3.2: albedo is wavelength dependent; albedos at 3 different wavelengths are compared The. How big is the impact of? the wavelengths difference This should be discussed, e.g. look at the OMI albedo and include the relative difference for these different wavelengths (over Brussels and Antwerp).

The albedo is indeed wavelength dependent and therefore a statement was present in the next to last paragraph to warn for the difficulties when comparing different albedo products: "Even if a direct comparison of different albedo products is not trivial due to BRDF-effects and albedo wavelength dependencies, among other…"

We followed your suggestion and tried to quantify the wavelength dependency based on the OMI LER. Below is a plot of the surface reflectance over Antwerp and Brussels for the 23 wavelength bands (and for both the yearly and monthly OMI LER product (June)). For TROPOMI NO_2 retrievals, the OMI LER product at 440 nm is used. Note that we used the MODIS MCD43A3 product at 470 nm and APEX SR is at 490 nm (middle of the APEX NO_2 fitting interval). The relative differences are 0 for the pixel over Brussels and 2.3% (yearly) – 3.8% (monthly) over Antwerp. Note that the relative difference between 440-470 and 440-490 is the same in all cases. Based on these tests, the impact of the wavelength dependency seems to be small at these wavelengths.



We have added the following in the paper: "The albedo is wavelength dependent and albedo products at different wavelengths have been compared: OMI LER at 440 nm, MODIS MCD43A3 at 470 nm and APEX albedo at 490 nm. The wavelength dependency has been assessed by analysing the relative difference of the OMI LER albedo over Brussels and Antwerp between 440 nm, and 470-490 nm for both the yearly and monthly OMI LER product (June). Overall, the OMI LER albedo increases slightly with wavelength but the increase is smaller than 4% between 440 and 490 nm."

7) p. 13, l. 16: it could be due to meteorology; e.g. lower wind speeds can increase the VCD enhancement even though emissions do not increase VCDs can be higher for stagnant winds, there could also be factors that potentially increased the lifetime of NOx (e.g. OH, O3, and NOx concentration) for that particular day. I think meteorology should be mentioned as a potential influence; look at the wind speeds and direction for these days (the wind speed is definitely lower). If both TROPOMI and APEX observed higher VCDs on June 29, this would not be due to the APEX instrument troubles.

Good comment! Indeed, meteorology is certainly playing a role here, as well as other factors increasing the NOx lifetime. I was surprised at first glance not to see a "weekend effect", which we usually see for example in MAX-DOAS data in Brussels, mainly monitoring traffic emissions. However, thinking about it we are mainly looking at emissions from petrochemical industry, which is probably constant each day of the week…this in contrast to traffic emissions. And like you comment lower wind speeds and other factors can increase the VCD, even when emissions are stable. We have altered the paragraph as follows:" Although June 29 is a Saturday, the NO_2 VCDs observed over the Antwerp harbour are slightly higher than on June 27, both in the APEX and TROPOMI data. The prevailing emissions in Antwerp from petrochemical industry are expected to be rather constant in contrast to traffic emissions, but meteorology, for example a more stagnant wind speed (3.7 m s^{-1} on 27 June and 2.3 m s^{-1} on 29 June, on average), and other factors that can potentially increase the lifetime of NOx, might explain the slight NO_2 VCD increase observed on June 29. However, when …

Note that there is some misunderstanding here: the instabilities with the instrument are no explanation for the fact that we observe higher VCDs on June 29. We mentioned this to explain why we don't have data over the city center on June 29, as we couldn't analyse the first three flight lines. The data we show and compared with TROPOMI are not affected by the encountered instrumental issues. In fact both were 2 different remarks related to the 29 June VCD map and we suggest to split it in two (small) paragraphs for clarity.

8) p. 22, l. 21: Can you really conclude this if your comparison is done over a small area and over a short time period, seasons are not considered (e.g. snow)? Re-phrase this, or add "over Belgium in the summer time."

Thanks for pointing this out. We agree that the statement is "too strong" based on the data set we currently have. We have nuanced this in the conclusion: "The case study over polluted regions in Belgium in summer time demonstrates that the TROPOMI tropospheric NO_2 product meets the mission requirements in terms of precision and accuracy."

Note however that some nuance was already present in the last paragraph of the conclusion, i.e. mentioning that more flights/data are needed under different geophysical conditions: "…The main focus was to quantify the TROPOMI retrieval uncertainties in polluted regions and results from the comparison with APEX data, acquired over Belgium in summer time, have shown that the TROPOMI tropospheric NO_2 product meets the mission requirements in terms of accuracy and precision. However, additional validation studies are required and are currently planned, focusing on more sites with different geophysical properties and varying pollution levels, including background areas, extreme albedo sites, other seasons, and cloudy scenes, among others, in order to assess as well the performance in suchlike conditions."

Note as well that new flights will take place over the two sites in summer 2021. Also other validation activities, involving airborne instruments, will take place later this year and next year over several sites in Europe. For example, recurrent flights will take place over the cities of Bucharest and Berlin covering different seasons. Still under construction but here's a link to upcoming TROPOMI validation activities: https://s5pcampaigns.aeronomie.be/

9) p.1 l. 24-26: "When the absolute value . . ., when comparing APEX NO2 VCDs with TM5-MP based and CAMS-based NO2 VCDs, respectively." I suggest re-wording this sentence, e.g.: The absolute difference is on average xx molec cm-2 (16%) and xx molec cm-2 (9%) compared to . . .

We have applied the suggestion in the abstract (and also in the conclusion).

10) p.1 l. 26: Which accuracy requirement; maybe change it to "mission accuracy requirement"

Good suggestion and corrected throughout the paper.

11) p.1 l. 29-30; suggest re-wording: Something like: The current TROPOMI data underestimate localized enhancements and overestimate background values by approximately 1-2x 1015 molec cm-2 (10- 20%).

Thank you for the suggestion! Note that for the same reason as for the earlier comment #8 (and also comment #8 from Reviewer #1) we have also added some nuance.
In the abstract: "For a case study in the Antwerp region, the current TROPOMI data underestimates localised enhancements and overestimates background values by approximately 1-2 x 10^{15} molec cm^{-2} (10- 20%)."

In the conclusion:" The TROPOMI spatial resolution is limited to resolve fine-scale urban NO_2 plumes and can cause a considerable smoothing effect in case of the observation of strongly polluted scenes with steep gradients. This depends both on the instrument pixel size and the amount of heterogeneity in the NO_2 field. The high-resolution APEX retrievals allow to monitor the effective horizontal variability in the NO_2 field at much finer scale. In Sect. 6, the impact of smearing of the effective signal due to the finite satellite pixel size was studied for the Antwerp region based on a downsampling approach of the APEX retrievals. Assuming a pixel size of 25 to 20 km^2, equivalent to the initial 3.5 km x 7 km and new TROPOMI 3.5 km x 5.5 km spatial resolution (at nadir), the TROPOMI data underestimates localised enhancements and overestimates urban background values by approximately 1-2 x 10^{15} molec cm^{-2}, on average, or 10% - 20%, for the Antwerp case study. The average under- and overestimation is further reduced to 0.6-0.9 x 10^{15} molec cm^{-2}, or smaller than 10%, when increasing the pixel size to 1 km^2. Therefore, detailed air quality studies at the city scale still require observations at higher spatial resolution, in the order of 1 km^2 or better, in order to resolve all fine-scale structures within the typical heterogeneous NO_2 field."

12) p.3 l. 13: "studied in Sect. 6" change to "see Sect. 6"

Corrected.

13) p.4 l. 4 Air pollution levels over Belgium. . . Do you have a reference that can be included here?

This statement was mainly based on own experience while looking at satellite data, where in Europe always a hotspot over North of Belgium/South of The Netherlands is popping up, together wit the Po valley, Paris, Ruhr area, etc.

I couldn't find a proper peer reviewed journal paper but we suggest to add following reference from a Greenpeace study where Flanders, or more specifically Antwerp, is mentioned as one of the 50 biggest NO_2 hotspots in te world. Only the German Ruhr and Paris are the other two European hotspots appearing in the list:

https://www.greenpeace.org.au/research/new-satellite-data-reveals-worlds-largest-air-pollution-emission-hotspots-greenpeace-media-briefing/

also the in-situ data from the environmental network on following website is showing how Belgium is flirting with or exceeding the EU annual limit value of 40 µg/m3 for NO_2:

https://www.eea.europa.eu/themes/air/country-fact-sheets/2019-country-fact-sheets/belgium-air-pollution-country

14) p.6, l. 16 VNIR; this should be defined, maybe in the previous sentence were the two channels are mentioned

As suggested, we have changed the previous sentence to: "APEX records backscattered solar radiation in the visible, (near-)infrared regions of the electromagnetic spectrum, covering the 370 to 2540 nm wavelength range in two channels, a visible/near-infrared channel (VNIR) and a short-wave infrared channel (SWIR)."

15) p.7 l. 28, mention the height of the layers (between surface and xx km)

Profile height has been added: "…and (2) daily NO_2 vertical profiles from the TM5-MP model on a 1° × 1° grid and covering 34 vertical layers (between surface and TOA)."