We thank the referee 1 for the comments and we answer to the specific questions below. The referee's comments are in black while the answers by the authors are in blue.

1) Ialongo et al. claim that a-priori profiles have been replaced with high-resolution CAMS profiles (e.g. in the abstract p.1 l. 7; p.4 l. 25-27; p.14 l. 3-5). However, this is not true when reading the method section (p.7 l. 1-7); in fact, the tropospheric columns are simply scaled with the tropospheric CAMS columns (not profiles). Replacing the a priori profile shape with the profile shape of a high resolution model is a common technique to improve satellite tropospheric NO2 columns. However, to do this new AMF have to be estimated, e.g. Goldberg et al. (2019), McLinden et al. (2014); Russell et al. (2011), Palmer et al. (2001); Martin et al. (2002) and lots more. The a priori vertical column densities do not have a linear relation to the TROPOMI tropospheric columns. To replace the standard low resolution profile shape with that from a high resolution regional model, an new AMF has to be estimated; the relationship is not simple due to the radiative transfer in the atmosphere.
In the comparison, it can be seen that this is not a good method as the columns are simply scaled, leading to a worse product than the standard tropospheric columns. As the CAMS model is a high resolution model near a city or hot spot, these columns will be larger than for the lower resolution TM5-MP model, leading to R>1 (in eq. 3), and thus all TROPOMI columns are scaled up. Thus, it is intuitive that the scaled columns are better for high concentrations, but overall worse.
I would suggest to either use CAMS to estimate new AMFs (similar to the references provided above), or to cut this part out of the manuscript.
If CAMS is used to estimate the AMF, more description of the model is needed, from the description on p.4 l.25-30 it is not clear what time stamp was used. Is an hourly output used? Are these interpolated to the time of the overpass?
I am also confused, why CAMS above 3km was used (3-5km). The largest impact on the tropospheric AMF comes from the high concentrations near the surface (in the boundary layer) around cities or other NOx sources. High-resolution models are used to improve the satellite tropospheric columns, because of the improved profile shape primarily in the boundary layer close to the emission sources, not to correct for the profile shape of the free-troposphere.

The approach to replace the a-priori is very briefly described on page 4, line 22-30, by referring to the Product User Manual (PUM) where the procedure is described. This approach provides a new estimate of the tropospheric column by using the full profile to recompute the air-mass factor. On page 7 we describe how the total column comparison is made, by updating only the troposphere (new a-priori) and keeping the stratosphere unchanged (eq. 3).

The response of the referee made us realise that the explanation how this is done was too short. Indeed the reader may get the impression that we simply use ratios of tropospheric columns. This is not the case and we clarify this in sect. 2.3 as well. As mentioned, the recipe to replace the a-priori is described by Eskes et al. (2019). This approach makes use of the averaging kernels and involves integrals over the profiles, so the full profile shape is used. This new profile shape leads to a new AMF. As mentioned by the referee, there is no direct relation between the a-priori column and the retrieved column, since only the profile shape determines the AMF. This approach, based on the averaging kernels, works if only the a-priori profile of NO_2 is replaced and no other inputs for the retrieval are changed. The approach makes use of the fact that NO_2 is optically thin (which is valid except for incidental extremely high tropospheric columns).

The referee mentions several papers where the air-mass factors were recomputed. For instance, McLinden et al. (2014), but also the POMINO product over China (Lin, J. T. et al., 2014) introduce high-resolution regional model outputs to improve the retrievals on a regional scale, similar to what is presented in our paper. However, in these papers not only the a-priori is replaced, but also other aspects of the retrieval are modified, such as the use of alternative (high-resolution) albedo maps or the explicit treatment of aerosols. In these cases, indeed, the radiative transfer calculation has to be done again to compute the impact on the tropospheric air-mass factor, because these changes also lead to a change of the averaging kernels. The approach described in the PUM no longer works.

The averaging kernels in case of clear and weakly clouded scenes decreases when moving from the tropopause down to the surface. Indeed, as mentioned by the referee the column amount above 3 km is small compared to the column amount in the boundary layer. But, because the sensitivity in the free troposphere is much higher (e.g. factor of 3 is normal) we find that this small free troposphere column still has a substantial impact on the AMF especially in the more rural areas. The regional models are not designed to describe the free troposphere accurately and produce unrealistically low NO_2 above 2-3 km. This is why we combined profile information from the CAMS-global system (3 km to tropopause) with the CAMS-regional profiles below 3 km.

To explain this also in the paper, we extended the last paragraph of section 2.1 (page 4):

"Since the retrieval of TROPOMI vertical column densities (VCDs) is sensitive to the a-priori estimate of the NO_2 profile shape, the accuracy of the VCDs may be improved by using a-priori profiles from a chemical transport model (CTM) with a higher resolution than the 1°×1° of TM5-MP (Williams et al., 2017). The air-mass factor (AMF) can be recomputed using an alternative a-priori NO_2 profile, resulting in a new retrieval of the tropospheric NO_2 column as described by Eskes et al. (2019).

In order to analyse their impact on the comparison, below 3 km altitude we used NO_2 profiles from the CAMS regional ENSEMBLE model (Météo-France, 2016; Marécal et al., 2015) as an alternative to the TM5-MP profiles. The CAMS regional ENSEMBLE is a median of seven European CTMs, and the data are provided on a regular 0.1°×0.1° grid over Europe on 8 vertical levels up to 5 km altitude. In addition, the CAMS global model was used to generate the profiles above 3 km altitude with the assumption that this model gives a more reliable description of NOx in the free troposphere. Data for CAMS global are provided on a regular 0.4°×0.4° grid on 60 model levels reaching up to 0.1 hPa (Flemming et al., 2015). In particular, we used the ratios between TROPOMI tropospheric air-mass factors derived using the hybrid CAMS regional/global a-priori profile (henceforth "CAMS a-priori") and the TM5-MP a-priori profile (see Sect 2.3). These ratios were provided on the regular CAMS 0.1°×0.1° grid for the period 30 April to 30 September 2018.

In order to minimize representativeness errors during the comparison, certain considerations were taken into account so that the fields could be correctly sampled in space and time. Horizontally, all available gridded data were interpolated to the CAMS regional, 0.1°×0.1° grid. Source grids in this process were either the TROPOMI native grid, which is different for each orbit, the CAMS global grid or the TM5-MP grid. Horizontal interpolation of retrieval columns was realized by means of a weighted average of all individual columns within a target grid cell. Intensive variables (e.g. temperatures, pressures, averaging kernels, the tropopause layer index etc.) were interpolated horizontally using bilinear regridding. Modelled fields were also interpolated in time, based on the satellite overpass time over Central Europe. All vertical levels of source data were linearly interpolated to the TM5-MP vertical levels and all subsequent integrations to columns were performed based on those levels. Pressures at each of those levels were calculated based on the surface pressure and the hybrid coefficients included in the TROPOMI product, which originate in TM5-MP. For the column integrations, all concentrations were converted to densities based on temperature and pressure profiles provided by TM5-MP."

2) The Kumpula AQ in situ measurements are converted from surface concentrations to total columns, based on the correlation between the PANDORA and in situ measurements. One concern is that these two instruments are not co-located and are quite likely measuring two different airmasses. Especially, since the in situ measurements are taken near an airport, and thus have likely high concentrations near the surface that may or may not be captured by PANDORA, depending on the winds etc. Further, the good correlation is primarily driven by three measurements that measured high amounts of NO2 for the PANDORA and in situ measurements. I would suggest cutting this figure (Fig. 5), since it is not used for any qualitative comparison, a similar figure is provided in Fig. 2.

This was a misunderstanding. The AQ station and Pandora are indeed co-located. They are about 100 meters from each other in the Kumpula area of Helsinki. The confusion came perhaps from the two points in Fig. 1. We clarify this in the text. We find figure 5 important to visualize the temporal correspondence between in situ measurements and satellite observations; we remove now the lines to make it clearer as suggested by the referee n.3.

We add this sentence in section 2.2: “This station, also known as SMEAR III station (Järvi et al., 2009), is located close to the Pandora instrument (about 100 m distance).”

3) A little more can be done in this paper in terms of validation. Here are some suggestions:

3a) On p.4, l.1-3 Ialongo et al. claim that the differences should be small between the OFFL and NRTI version. I think this paper would provide a good opportunity to quantitatively identify the differences between the NO2 NRTI and OFFL version (e.g. similar as Garane et al., 2019 who quantified the differences between the OFFL and NRTI TROPOMI O3 columns to ground-based observations).

The NRTI data are not stored and are replaced with the OFFL in the sentinel data hub, so the NRTI data are not available for a comparison in the past. Nevertheless, there is an operational validation of S5p products by the S5P-MPC-VDAF (S5P - Mission Performance Center - Validation Data Analysis Facility, http://mpc-vdaf.tropomi.eu/), which includes online comparisons between both NRTI and OFFL NO2 products and the Pandora NO2 total columns from the Pandonia Global Network, including the Helsinki site. The results are summarized in 3-montly validation reports and they show almost identical results (see for example the last report here: http://mpc-vdaf.tropomi.eu/ProjectDir/reports/pdf/S5P-MPC-IASB-ROCVR-04.0.0-20190923_FINAL.pdf)

We add this document as reference to the text and we mention the operational validation activities as also suggested by referee n.3.

3b) There may be limited measurements available but perhaps looking at the differences between TROPOMI and PANDORA NO2 columns in terms of TROPOMI’s SZA, cloud fraction etc. similar as in Beak et al. (2017) Fig. 5 or Fig. 7

We add plots in the supplement including the bias vs SZA and CRF but we note that we apply already a screening to the data that removes cloudy pixels and high SZA values. There is an apparent increase in bias (first positive, then negative) with increasing CRF but less clear with SZA. We also analyse the bias vs the time of the day and pixel number and we update the text as follows:

“Figure S6 in the Supplement includes the absolute differences between TROPOMI and Pandora NO2 total columns as a function of TROPOMI SZA (solar zenith angle) and CRF (cloud radiative fraction) (upper and lower panel, respectively) within the range of values allowed after the TROPOMI data screening (QA value >0.75). While the dependence between the differences and SZA values is not clear, the differences for SZA above 45˚ are generally larger (between -3 and +1e15 molec./cm2) than for smaller SZA values (0 to 1e15 molec./cm2). Similarly, larger CRF values correspond to larger (positive or negative) absolute differences.

Since S5P has often two valid overpasses per day at the latitude of Helsinki 60°N), it is possible to study the NO2 daily variability between about 12 and 15LT. The S5P overpass time typically corresponds to the NO2 daily local minimum (between the morning and afternoon peaks due to commuter traffic), observed for example in the NO2 surface concentration measurements from Kumpula AQ site (Fig. S7). Figure 5 (upper panel) shows TROPOMI and Pandora NO2 total columns as a function of the time of the day between 12 and 15 LT. Both datasets show an enhancement around 13:30LT and lower NO2 levels before and after. The relative differences between TROPOMI and Pandora NO2 total columns do not show a clear dependence on the time of the day (Fig. 5, lower panel), but the dispersion (standard deviation of the relative differences) is larger (about 30%) before 13:30 LT than afterwards (21%). Increasing time of the day also corresponds to increasing pixel number (filled colour dots in Fig. 5, lower panel), since the first overpass of the day corresponds to the left side of the orbit (smaller pixel numbers) while the second overpass to the right side (higher pixels number). No clear dependence between the relative differences and the pixel size (larger at the edges and smaller in the center of the swath) was observed."

- Further, adding a boxplot showing the differences between the TROPOMI and PANDORA columns binned in low, medium, high columns (e.g. 0-0.6 , 0.6-1 , >1 10^16 molec/cm2) would also improve the paper and provide more contents to the discussion. This is already discussed on p.10 l.1-5, but a figure would help.

We added a box plot in the supplement as suggested.

- The paper would improve if the time period of the comparison could be increased maybe use 1 year of data (April 2018 to April 2019). Maybe one concern would be data in the winter time with snow cover, but the difference between summer and winter observations could also be investigated.

This is unfortunately not possible because we have no measurements from Pandora for winter or for year 2019 due to maintenance. TROPOMI data also are not available at Helsinki latitude for more than a couple of months in winter, after the quality flag screening. Further analysis will be perhaps the focus of a future work, when a larger amount of data are collected.

Minor comments
Figure 2: The lines are confusing and misleading, the columns are completely unknown when no measurements are taken. I would suggest replacing the line plot with a scatter plot, at the very least for the TROPOMI, and PANDORA 10min avg. measurements.

We changed figure 2 according to the suggestions.

Figure 3: It's hard to tell the difference between weekdays and weekends. I would suggest replacing the "weekend marker" with a triangle marker (or something similar). It is also sufficient to reduce the size to a 1-column plot.

We changed figure 3 according to the suggestions.

P. 2 l. 5: "Netherlands" -> "Netherlands Space Office"

Changed

p. 3 l. 10: According to the AMT author guidelines dates should be written as dd month year: "on the 13th October" -> "on 13 October"

Changed

p.3 l. 14: "UV-Visible (UVVIS)" -> "UV-VIS" (as defined on p.2 l. 24)

Changed

p.3 l. 20 DOAS already defined on p.2 l. 25

Removed

p.3 l. 29: "15.04-30.09.2018" -> "15 April to 30 September 2018"

Changed

p. 3l. 32,p.4. l. 1: NRT->NRTI

Changed

p.4 l. 12 : 15.04.2018-30.09.2018 -> 15 April to 30 September 2019

Changed

p.4. l. 18 -21: maybe move Fig. S1 from the supplement into the main paper. It is discussed here in a few sentences and seems important.

We think that the supplement is more appropriate for such technical maps.

p.6 l. 3: FMI not defined, please define. Also, are these ground-based measurements publically available? If, so please provide the link where it can be downloaded.

Changed

p. 10 l. 11: Figure S2 -> Fig. S2 (from AMT author guidelines)

Changed

p.10 l. 25-30: as suggested in the previous section, this can be cut together with Fig.5

We leave it together with the picture.

p. 13 l. 22: "We find this partially. . ." -> this has not been concluded or found from the analysis in this paper; maybe change it to : "This is partly due to the profile shapes of the low resolution TM5-MP model used to compute the standard TROPOMI tropospheric NO2 columns and thus. . ."

We change the sentence as: "This is partly due to the low resolution of the TM5-MP profile shapes used to compute the tropospheric air-mass factors and thus the vertical columns."

p. 15 mention that this study is using summer observations only (unless the time period has been changed, see previous suggestions), with no snow cover (?)

Added

p.15 l. 4: the comparison to the results from Griffin et al. could be a bit more quantitatively: were the results similar, how similar? Include some numbers.

We refer now to the correlation coefficient and bias values as follows:
"The correlation between Pandora and TROPOMI NO2 retrievals is also in line with the results obtained over the Canadian oil sands (r=0.70 according to Griffin et al., 2019). On the other hand, Griffin et al. (2019) report a mean negative bias up to -30%, as expected for very polluted sites, while we find a smaller positive bias (on average about 10%) over a relatively less polluted site like Helsinki."