We thank reviewer 2 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.

General comments:
1. The validation is based on total columns. The reason for doing so is reasonable for me. However, we usually rely on tropospheric columns to investigate air pollution. I would recommend adding the analysis focus on tropospheric columns, even though systematic retrieval errors may exist. Such validation results will be very useful for data users to have a better sense about the current quality of the data.

We validate summed columns as they are those comparable with the ground-based Pandora observations. We do not have equivalent measurements of tropospheric NO2 from a ground-based instrument in Helsinki. On the other hand, we use the tropospheric columns for qualitative analysis as the weekly cycle, for example.

2. The comparison with OMI. The authors have performed a similar validation of OMI NO2 columns against Pandora observation. Do the validation results differ significantly from this study? I would recommend a short discussion to compare the OMI and TROPOMI validations.

We mentioned this but we write now in more details in the conclusion as follows: "As compared to previous satellite-based instruments such as OMI, the bias against ground-based observations in Helsinki is similar on average (±5% under clear sky conditions for OMI, Ialongo et al. (2016)), while the correlation coefficient is generally higher (r=0.68 for TROPOMI and r=0.5 for OMI, see Ialongo et al., 2016)."

3. The use of high-resolution profile. I expect a better performance of the NO2 products using CAMS profiles compared to those using TM5 profiles based on the experience on OMI validations. However, as shown on Page 13, the use of CAMS a-priori profiles does not improve the agreement with Pandora significantly. What is the most likely reason for this? Does it indicate that TM5 profiles are good enough for the retrieval?

Indeed the improvement is not significant on average but it is sensible for episodes with high NO2 columns as measured by Pandora. The improvement is expected to improve the retrieval under polluted conditions where the spatial variability is sharper, but we have in Helsinki also several overpass with somewhat background conditions, so that the change overall remains small (within the uncertainties). Also, Griffin et al. 2019 also stated that using high-resolution input improves the tropospheric AMF and the tropospheric NO_2 VCDs but the correction is not as significant as previously seen for OMI. That study included also a better characterization of snow-covered surfaces.

We update the text in the Sect. Results as follows:

“The comparison shows that the largest differences between the two summed columns are mostly found in cases of relatively high concentrations. In these cases, the use of CAMS profiles generally increases the TROPOMI summed columns and reduces the difference between TROPOMI and Pandora (from -28.5±3.3 % for TM5-MP to -23.7±3.5 % for CAMS). On the other hand, in cases of low concentrations, where TROPOMI tends to overestimate the VCDs compared to Pandora, the use of CAMS a-priori profiles slightly increases the positive bias (from +16.9±2.3 % for TM5-MP to +19.1±2.3 % for CAMS). Because the largest improvement is achieved for relatively high concentrations and negative biases becoming less negative, the overall MRD value increases from 11.5 % to 14 % (Table 2). According to a two-sided t-test, the differences of the two mean absolute biases (MD) in Table 2 are statistically significant at the 52% significance level. Thus, on average, the use of CAMS profiles does not improve significantly the agreement with Pandora observations.

For this smaller subset of 75 co-locations with Pandora the correlation between TM5-MP summed columns and Pandora is 0.74 and the slope of a least squares linear fit is 0.45. Using the CAMS profiles improves the agreement with Pandora in terms of correlation and slope, with their values increasing to 0.80 and 0.52, respectively. This improvement is more evident for high values of the Pandora NO2 total columns with the correlation and the linear slope increasing by 0.1 and 0.27, respectively, from TM5-MP to CAMS (Table 2).

The time series in Fig. S8 of the supplement further illustrate how using the high-resolution CAMS profiles increases the TROPOMI tropospheric columns so that the summed columns (yellow dots) become closer to Pandora's peak values (blue dots), corresponding to episodes of NO2 enhancement, but that overall the difference between the summed columns obtained using TM5-MP and CAMS remains mostly within the uncertainties of the TROPOMI NO2 retrieval.”

We clarify this also in the abstract and conclusion, respectively, as follows:

Abstract:

“Replacing the coarse a-priori NO2 profiles with high-resolution profiles from the CAMS chemical transport model improves the agreement between TROPOMI and Pandora total columns for episodes of NO2 enhancement. When only the low values of NO2 total columns or the whole dataset are taken into account, the mean bias slightly increases. The change in bias remains mostly within the uncertainties.”

Conclusion:

“In Helsinki we find that replacing the original profiles with those derived from the high-resolution CAMS regional ensemble model increases the TROPOMI NO2 tropospheric columns and partly reduces the discrepancy between TROPOMI and Pandora VCDs for episodes of relatively high NO2 concentrations, while increasing the correlation and the linear fit slope. On the other hand, the agreement does not significantly improve on average or for lower values of NO2 vertical columns. Overall, the change in bias remains mostly within the uncertainties.”

Specific comments:

1. Page 3, line 1. "The improved resolution of TROPOMI retrievals is expected to reduce the effect of dilution, due to the relatively coarse pixel size as compared to the field-of-view of the ground-based observations." I guess the authors want to say the pixel size of TROPOMI is finer than that of OMI and thus the effect of dilution is reduced. If so, what the reason for pointing out the relatively coarse pixel size as compared to the field-of-view of the ground-based observations here?

We mean here that the smaller pixels of TROPOMI (compared to OMI) will possibly reduce the dilution effect when compared to the field-of-view of the ground-based observations.
We rewrite this as: "The improved resolution of TROPOMI retrievals is expected to reduce the effect of spatial averaging compared to OMI, leading to a better agreement with the ground-based Pandora observations that has a relatively narrow field-of-view."

2. Page 3, line 29. The time format of "15.4.–30.9.2018" is a little bit confusing for readers. I recommend using the April 15- Sep 30. Same comments for Page 4, line 30.

We changed that throughout the manuscript according to the recommendations for AMT journal

3. Page 12, line 4. The authors use summed columns for TROPOMI and total columns for Pandora. Is this intended? If so, please clarify the reason in the text.

Yes it was on purpose. We explained that we used the summed over the total column product, because of the latter's sensitivity to the ratio between the stratospheric and tropospheric a-priori columns may lead to substantial systematic retrieval errors. The intermediate step of using data assimilation to first estimate the stratospheric column does remove part of this error.

We add also this sentence in the text to further clarify:
"The summed total column product is described by the data provider as the best physical estimate of the NO2 vertical column and recommended for comparison to ground-based total column observations (van Geffen et al., 2019)."

4. Page 15. Line 4. "The correlation between Pandora and TROPOMI NO2 retrievals is also in line with the results obtained by Griffin et al. (2019) over the Canadian oil sands." How those two studies are in line with each other? I recommend presenting the quantitative analysis for the consistency.

We rewrite the text 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.”