**We thank the referee S. Compernolle for the useful 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.**

Overall

1/ There are indicators for bias (the MD and MRD) but not for the dispersion of differences, for example the standard deviation of the differences or the interquartile range of the differences. Please add e.g., the standard deviation of the differences to the methodology, together with the definitions for MD and MRD, and discuss the results in the manuscript, including table 1 and 2.

We added the SD of the differences in Table 1 and 2 and we briefly discuss it in the text.

2/ Although the uncertainties of S5p NO2 (p. 4) and Pandora (p. 5) are shortly mentioned, it is not discussed (e.g., in the conclusions) whether discrepancies between S5p and Pandonia are reasonable with respect to the uncertainties. Both S5p NO2 and Pandora measurements have an uncertainty provided per measurement. In the time series of co-located points of S5p NO2 and Pandora, the error bars based on the provided uncertainties can be added. It can then also be discussed whether the S5p values based on the CAMS a-priori are meaningfully different from the TM5-MP based S5p values.

We added the errorbars in the Fig. 2 (and Fig. S8 of the updated supplement), as suggested. We discuss now in more details how the observed discrepancies compares to the uncertainties as follows:
"We find that the differences between the total columns derived from the TROPOMI and Pandora instruments are on average around 10 % (or 0.12×10^15 molec./cm^–2), which is smaller than the precision of the TROPOMI summed columns used in this study (10–50%) and well below the requirements for TROPOMI observations (25–50 % for the NO2 tropospheric column and 10 % for the stratospheric column; ESA, 2017)."

We also discuss the significance of the change of a-priori as described in the following points.

3/ Minor comment: be consistent in the units for NO2 column number density, and preferably use 1015 molec cm–2 as unit in the Tables and figures, as this is very commonly used in NO2 column comparisons. Currently the authors use 1014 molec cm–2 in table 1 and 2, and 1016 molec cm–2 in e.g., Fig. 5.

All pictures and tables are corrected accordingly to this suggestion

Detailed comments

4/ Abstract, line 5. 'TROPOMI total columns underestimate ground-based observations for relatively large Pandora NO2 total columns'. It should be added here that TROPOMI overestimates for the lower columns. Also the obtained bias (absolute scale and rela-

tive), and the dispersion of the differences (e.g., the standard deviation of differences, as noted above) should be added in the abstract.

The following text was added to the abstract:
"The mean relative and absolute bias between the TROPOMI and Pandora NO2 total columns is about +10% and 0.12e15 molec./cm^2, respectively. The dispersion of these differences (estimated as their standard deviation) is 2.2e15 molec./cm^2."
[…]
"On the other hand, TROPOMI slightly overestimates (within the retrieval uncertainties) relatively small NO2 total columns."

Abstract, line 9. Here it is stated that " 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." Please add a statement on the overall agreement and/or episodes of low NO2.

We added the following text to the abstract:
"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."

Introduction. p. 2, around line 27. Here, the authors should add that there is an operational validation of S5p products by the S5P-MPC-VDAF (S5P - Mission Performance Center - Validation Analysis Facility, http://mpc-vdaf.tropomi.eu/) which includes online comparisons and validation reports using the S5p total NO2 vs Pandora from the Pandonia Global Network, including the one at the Helsinki site.

We added this text to the introduction:
"The TROPOMI/S5P NO2 products are operationally validated by the S5P-MPC-VDAF (S5P - Mission Performance Center - Validation Data Analysis Facility) using the Pandora NO2 total columns from the PGN. The operational validation results are reported every 3 months at the S5P-MPC-VDAF website (http://mpc-vdaf.tropomi.eu/)"

p. 4, line 4. I would add here that the summed total column is the one that is recommended by the data provider.

We add 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)."

p. 4, line 27 and following. More detail should be provided here:
• Is reanalysis data used ?
• make clear that CAMS global, despite the name similarity, is a very different model compared to CAMS regional

• add reference for CAMS global, the horizontal resolution, and the vertical range.
• 'better description of free troposphere': do you mean better compared to TM5-MP ?
• make more clear that you are actually constructing a hybrid profile from CAMS regional and CAMS global.
• line 29. '...using the CAMS (...) a-priori profiles'. Certainly this first time, I sug- gest to formulate instead 'using the hybrid CAMS regional/CAMS global a-priori profiles (called shorthand "CAMS a-priori profile" from now on) ' or some similar formulation.
• line 30. 'These ratios were available on the regular CAMS 0.1x0.1 grid' This sounds as if the authors obtained the AMF ratios from elsewhere. But if I under- stood well, you actually calculated the ratios yourself, using input from the hybrid CAMS regional/CAMS global profile and from the S5p product, right? Also, the procedure how to calculate the AMF ratio using CAMS a priori data and S5p NO2 input (averaging kernel, TM5-based AMF) should be explained. E.g., likely there was need for (i) a vertical regridding of the CAMS profile to match the vertical grid of the averaging kernel of S5p NO2, and (ii) an horizontal interpolation (if so, what kind of interpolation) of the CAMS global profile to the CAMS regional grid.

We try to answer all your questions by changing/adding the text at the end of Section 2.1 as follows:

"Since the retrieval of TROPOMI vertical column densities (VCDs) is sensitive to the a-priori estimate of the NO2 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 NO2 profile, resulting in a new retrieval of the tropospheric NO2 column as described by Eskes et al. (2019).
In order to analyse their impact on the comparison, below 3 km altitude we used NO2 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."

These details can be discussed here, or alternatively in an appendix or the supplement.
p. 6, line 20. 'Pandora retrievals with data quality flag value of 0, 1, 10 or 11'. Pandora measurements can occasionally become negative and even reach several Pmolec cm- 2 in the negative. This is drastically reduced when only focusing on high-quality data with 0, 10 flags. Was there any filtering on negative Pandora values, or were these averaged together with the positive values, or were these -by chance- no longer present after co-location with TROPOMI?

Negative values were filtered out (they showd negative uncertainty as well) but they actually appeared only in two cases and including those in the calculation only changes the bias by a few decimals.

p. 7, fig. 2. I share the concerns of reviewer 1 on the clarity of this figure.

We changed it according to the suggestions

p. 7, line 5. 'CAMS a priori summed column' is somewhat ambiguous. A reader could assume this is a column purely derived from CAMS information. I suggest: 'the newly derived summed column, using the CAMS a-priori profile,...,is calculated as...'

We changed this with: "The new summed column, derived using the CAMS a-priori profile, was then calculated…"

p. 7, line 2. 'ratio (R) between the tropospheric column retrievals...' This is unclear. From section 2.1, I assume R is the ratio of the original AMFtrop of the S5p NO2 product and the newly calculated AMFtrop .

Yes, thank you. This was a mistake. We rewrite as follows:
"The effect of using high-resolution CAMS a-priori $NO_2$ profiles instead of TM5-MP (as used in the standard product) in the calculation of TROPOMI VCDs was analysed by calculating an alternative summed column using the ratio (R) between 5 the tropospheric air-mass factors derived using CAMS and TM5-MP a-priori profiles, computed on the CAMS-regional grid with 0.1° resolution (see Sect.

2.1)."

p. 7, Eq (3). From the formula, it is clear that the stratospheric contribution is not updated (still based on TM5-MP), while CAMS global is nonetheless available (as the authors used it for the free troposphere). A motivation is needed why CAMS regional+global is used for the troposphere while TM5 is kept for the stratosphere.

The retrieval includes an assimilation step to minimize the bias between the TM5-MP modeled and observed stratospheric column as much as possible. This is an essential element of the retrieval and should only be replaced when the other model has a high quality stratospheric NO2 and assimilates the satellite data to get a comparable or better analysis.
At this moment CAMS-global does not include detailed stratospheric chemistry, and the NO2 profiles in the stratosphere are poor. Secondly, CAMS assimilates only tropospheric columns from OMI and GOME-2 which does not impact the stratosphere.

We add this sentence: "The stratospheric columns from TM5-MP (as in the standard product) are used in the calculation of the new summed columns, because at the moment CAMS global does not include detailed stratospheric chemistry nor accurate NO2 profile information in the stratosphere."

p. 9, Table 1.
• Regarding the slope from orthogonal regression, it should be noted in the text C4 that this technique assumes that the standard deviation from random error in y (S5p NO2 total column) and x (Pandora total column) are equal, which is not at all guaranteed. See e.g., Carroll (1996), with η of Eq (4) assumed 1, or Wu (2018), who do not recommend orthogonal distance regression.

We replace the orthogonal regression with both the least square fit slope as well as the York fit slope as recommended by Wu et al. (2018) and we add this sentence:
"The York linear regression (York et al. 2004) is used alongside the traditional least squares linear regression, since it has been shown to be an appropriate measure of fit in situations where the two sets of data have different levels of uncertainty (Wu et al., 2018)."

• What is the meaning of the number after the ± ? Is it the standard deviation of the mean? This should be explained in the table footnote. Similar for Table 2.

Yes it is. We clarify this in the captions of both tables.

p. 10, line 19. What is the impact of changing the co-location criteria (spatial and temporal) on the standard deviation of the differences and the correlation coefficient?

We add now a plot in the supplement with the correlation coefficient and the standard deviation of the differences as a function of the changing co-location criteria in the supplement and we update the text accordingly.

p. 10, line 23. What is meant by 'variability' here? The amount by which the MD changes?

This sentence is removed and replaced with: "The MD value increases with increasing temporal averaging interval by about 0.3e15molec./cm2 (2 percentage points)."

p. 12, Fig. 5 right panel. Add error bars (based on the provided uncertainties) to S5p NO2 and Pandonia points. This figure will be clearer when using points instead of lines.

Corrected

p. 12, Fig. 6. What is the meaning of the vertical error bars? The standard deviation of the values in the month? This should be explained in the caption.

Yes it is. Corrected

p. 12-13 ( about the evaluation of the effect of using CAMS a-priori profiles) + Fig. S3
• Please add in Fig. S3 error bars on the S5p NO2 TM5-MP points and on the Pandonia points. This will give an indication whether the update with the CAMS a-priori profiles is significant with respect to the uncertainties.

Corrected. Note that S3 is S8 in the revised manuscript.

• Assumed that the numbers after the ± in Table 2 are standard deviations of the mean, it seems to me that the difference between the MD calculated with TM5- MP profiles on the one hand, and the MD calculated with CAMS a-priori on the other hand, is not statistically significant. Same remark for the MRD. This should then be also reflected in the abstract and the conclusions.

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).

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."

p. 13 line 4-5. '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 pro- files slightly worsens the agreement with Pandora by increasing the positive bias. ' Looking at Fig S3 this effect seems really small to me and is probably not statistically significant.

Add in Table 2 entries for 'Pandora high' and 'Pandora low' so one can conclude what is the significance of this effect.

We updated table 2 accordingly. See also the answer to the previous point.

p. 13, Conclusions. Here, it should also be stated whether the S5p vs Pandora discrepancies are reasonable (or not) in light of the measurement uncertainties of S5p and Pandora.

Corrected as follows:
"We find that the differences between the total columns derived from the TROPOMI and Pandora instruments are on average around 10 % (or $0.12 \times 10^{15}$ molec. $cm^{-2}$ ), which is smaller than the precision of the TROPOMI summed columns used in this study (10–50 %) and well below the requirements for TROPOMI observations (25–50 % for the $NO_2$ tropospheric column and <10 % for the stratospheric column; ESA, 2017)."

p. 13, line 22. 'while low values are overestimated' A short discussion on the possible reasons should go here. Does this mean that TROPOMI has a positive systematic error at low NO2 values? Or that the Pandora instrument has a negative systematic error? Or is it somehow due to the still relatively coarse resolution of S5p NO2? And is the overestimation actually significant with respect to the uncertainties?

The overestimation of low NO2 columns suggests a possible overestimation of the stratospheric fraction of the column. Also, replacing the surface reflectivity climatology (Kleipool et al., 2008) currently used in the retrieval with higher resolution geometry-dependent information is expected to improve the comparison of the TROPOMI NO2 vertical columns with the ground-based observations.
Anyway, the reasons for this positive bias are still under investigation. We mention this in the text.

p. 15, Data availability. It should be noted that there is no general open access to the S5p Expert users Data Hub, only to the S5p Pre-Operations Data Hub. Also, the point of access for CAMS regional and CAMS global should added here, and exactly which kind of data was used (forecast, reanalysis?).

We correct that and we add this text:
"CAMS regional forecasts and analyses for the previous day, as well as CAMS global forecasts are available through Copernicus Atmosphere Monitoring Service data portal (https://atmosphere.copernicus.eu/data)."