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.”

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."

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 NO_2 (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 NO_2 and Pandora measurements have an uncertainty provided per measurement. In the time series of co-located points of S5p NO_2 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 NO_2 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 NO_2 column number density, and preferably use 1015 molec cm−2 as unit in the Tables and figures, as this is very commonly used in NO_2 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 NO_2 total columns'. It should be added here that TROPOMI overestimates for the lower columns. Also the obtained bias (absolute scale and relative), 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 NO_2 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 NO_2 total columns."

Abstract, line 9. Here it is stated that " Replacing the coarse a-priori NO_2 profiles with high-resolution profiles from the CAMS chemical transport model improves the agreement between TROPOMI and Pandora total columns for episodes of NO_2 enhancement." Please add a statement on the overall agreement and/or episodes of low NO_2.

We added the following text to the abstract:
"When only the low values of NO_2 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 NO_2 vs Pandora from the Pandonia Global Network, including the one at the Helsinki site.

We added this text to the introduction:
"The TROPOMI/S5P NO_2 products are operationally validated by the S5P-MPC-VDAF (S5P - Mission Performance Center - Validation Data Analysis Facility) using the Pandora NO_2 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 NO_2 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 0.12×10^{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).”

Comparison of TROPOMI/Sentinel 5 Precursor NO_2 observations with ground-based measurements in Helsinki

Iolanda Ialongo[1], Henrik Virta[1], Henk Eskes[2], Jari Hovila[1], and John Douros[2]

[1]Space and Earth Observation Centre, Finnish Meteorological Institute, Helsinki, Finland.
[2]Royal Netherlands Meteorological Institute, De Bilt, Netherlands.

Correspondence: Iolanda Ialongo (iolanda.ialongo@fmi.fi)

Abstract. We present a comparison between satellite-based TROPOMI (TROPOspheric Monitoring Instrument) NO_2 products and ground-based observations in Helsinki (Finland). TROPOMI NO_2 total (summed) columns are compared with the measurements performed by the Pandora spectrometer ~~during April–September~~ between April and September 2018. ~~We find a~~ The mean relative and absolute bias between the TROPOMI and Pandora NO_2 total columns is about 10 % and 0.12×10^{15} molec. cm^{-2}, respectively. The dispersion of these differences (estimated as their standard deviation) is 2.2×10^{15} molec. cm^{-2}. We find high correlation (r = 0.68) between satellite- and ground-based data, but also that TROPOMI total columns underestimate ground-based observations for relatively large Pandora NO_2 total columns, corresponding to episodes of relatively elevated pollution. This is expected because of the relatively large size of the TROPOMI ground pixel (3.5×7 km) and the a-priori used in the retrieval compared to the relatively small field-of-view of the Pandora instrument. On the other hand, TROPOMI slightly overestimates (within the retrieval uncertainties) relatively small NO_2 total columns. Replacing the coarse a-priori NO_2 profiles with high-resolution profiles from the CAMS chemical transport model improves the agreement between TROPOMI and Pandora total columns for episodes of NO_2 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.

We also analyse the consistency between satellite-based data and in situ NO_2 surface concentrations measured at the Helsinki-Kumpula air quality station (located a few metres from the Pandora spectrometer). We find similar day-to-day variability between TROPOMI, Pandora and in situ measurements, with NO_2 enhancements observed during the same days. Both satellite- and ground-based data show a similar weekly cycle, with lower NO_2 levels during the weekend compared to the weekdays as a result of reduced emissions from traffic and industrial activities (as expected in urban sites). The TROPOMI NO_2 maps reveal also spatial features, such as the main traffic ways and the airport area, as well as the effect of the prevailing south-west wind patterns.

This is one of the first works in which TROPOMI NO_2 retrievals are validated against ground-based observations and the results provide an early evaluation of their applicability for monitoring pollution levels in urban sites. Overall, TROPOMI retrievals are valuable to complement the ground-based air quality data (available with high temporal resolution) for describing the spatio-temporal variability of NO_2, even in a relatively small city like Helsinki.

1 Introduction

Nitrogen oxides ($NO_x = NO + NO_2$) play an important role in tropospheric chemistry, participating in ozone and aerosol production. NO_x are mainly generated by combustion processes from anthropogenic pollution sources (including transportation, energy production and other industrial activities), and they are toxic in high concentrations at the surface (US-EPA, 2019). The NO_2 amount in the atmosphere can be measured using satellite-based instruments. Launched in October 2017, TROPOMI (TROPOspheric Monitoring Instrument), the only payload on-board the European Space Agency's (ESA) Sentinel-5 Precursor (S5P) satellite, is expected to revolutionise the way we monitor air pollution from space because of its unprecedented spatial resolution (3.5 × 7 km at the beginning of the mission and 3.5 × 5.5 km since 6 August 2019) and high signal-to-noise ratio. TROPOMI (jointly developed by the Netherlands Space Office and ESA) is designed to retrieve the concentrations of several atmospheric constituents including ozone, NO_2, SO_2, CO, CH_4, CH_2O, aerosol properties as well as surface UV radiation. TROPOMI derives information on atmospheric NO_2 concentrations by measuring the solar light back-scattered by the atmosphere and the Earth's surface. Due to its high spatial resolution, TROPOMI observations are particularly suitable to monitoring polluting emission sources at city level. The S5P mission is part of the Space Component of the European Copernicus Earth Observation Programme.

TROPOMI builds on the experience from previous polar orbiting instruments such as the Dutch-Finnish Ozone Monitoring Instrument (OMI), which has been operating on-board NASA's EOS (Earth Observing System) Aura satellite (Levelt et al., 2006) since late 2004. OMI NO_2 observations have been used in several air quality applications and the main achievements have been recently summarised by Levelt et al. (2018). The results achieved using OMI NO_2 retrievals include estimating top-down polluting emissions, analysing changes in the pollution levels over the period of 13 years, and verifying the success of environmental policy measures (e.g., Beirle et al., 2011; Castellanos and Boersma, 2012; Streets et al., 2013; Lu et al., 2015; Lamsal et al., 2015; Duncan et al., 2016; Krotkov et al., 2016; Liu et al., 2017). Also, OMI observations have been used for monitoring the NO_2 weekly cycle over urban sites (Beirle et al., 2003; Boersma et al., 2009; de Foy et al., 2016). Recently, a reprocessing of the OMI NO_2 dataset has become available (Boersma et al., 2018) as deliverable of the European QA4ECV project. Many of the QA4ECV OMI retrieval developments have been incorporated in the TROPOMI NO_2 retrieval processor.

Since TROPOMI/S5P is a very recent mission, accurate validation against independent ground-based measurements is needed in order to evaluate the quality of the retrieval. Recently, the Pandonia Global Network (PGN), including a network of ground-based Pandora spectrometers, has been established to provide reference measurements of NO_2 total columns for validating satellite-based retrievals. Pandora measures direct sunlight in the ~~UV-VIS (~~ultraviolet-visible ~~)~~ spectral range (280–525 nm) and provides NO_2 total columns using the direct-sun DOAS (Differential Optical Absorption Spectroscopy) technique (Herman et al., 2009). Recently, Zhao et al. (2019) presented a method to derive NO_2 total columns from Pandora zenith-sky measurements as well. The TROPOMI/S5P NO_2 products are operationally validated by the S5P-MPC-VDAF (S5P - Mission Performance Center - Validation Data Analysis Facility) using the Pandora NO_2 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/).

Very recently, Griffin et al. (2019) presented first results of the validation of TROPOMI NO_2 retrievals over the Canadian oil sands using air-mass factors calculated with the high-resolution GEM-MACH model. They show how the TROPOMI NO_2 vertical column densities are highly correlated with ground-based observations and ~~agree within~~ have a negative bias of 15–30 %. In this work, we evaluate the quality of TROPOMI NO_2 vertical columns against ground-based observations in the urban site of Helsinki (60.2° N; 24.95° E). Helsinki is a city with about half a million inhabitants, surrounded by a larger urban area (including the city of Espoo in the west and Vantaa in the north-east). Satellite-based NO_2 observations from OMI instrument in Helsinki were previously validated by Ialongo et al. (2016), finding that ~~OMI generally underestimates~~ the bias between OMI and Pandora total columns ~~by 5–30 %~~ranges between −30 % and 5 %, depending on the retrieval algorithm and parameters. 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~~ spatial averaging compared to OMI, leading to a better agreement with the ground-based ~~observations~~Pandora observations that have a relatively narrow field-of-view.

The satellite- and ground-based data used in the analysis are described in Sect. 2. The results of the comparison between TROPOMI NO_2 retrievals and ground-based Pandora total columns are shown in Sect. 3. The temporal correlation with in situ NO_2 surface concentration measurements and the NO_2 weekly cycle are also analysed. Finally, the conclusions are presented in Sect. 4.

2 Data and methodology

2.1 TROPOMI NO_2 observations

TROPOMI is a passive sensing hyperspectral nadir-viewing imager aboard the Sentinel-5 Precursor (S5P) satellite, launched on ~~the 13th~~ 13 October 2017. S5P is a near-polar sun-synchronous orbit satellite flying at an altitude of 817 km, with an overpass time at ascending node (LTAN) of 13:30 ~~(local time~~ local time (LT) and a repeat cycle of 17 days (KNMI, 2017). TROPOMI is operated in a non-scanning push broom configuration, with an instantaneous field-of-view of 108° and a measurement period of about 1 second. This results in a swath width of approx. 2600 km, an along-track resolution of 7 km and daily global coverage (KNMI, 2017). TROPOMI's four separate spectrometers measure the ultraviolet (UV), ~~UV-visible (UVIS~~UV-Visible (UV-VIS), near-infrared (NIR) and short-wavelength infrared (SWIR) spectral bands, of which the NIR and SWIR bands are new as compared to its predecessor OMI (Veefkind et al., 2012).

The NO_2 columns are derived using TROPOMI's UVIS spectrometer backscattered solar radiation measurements in the 405–465 nm wavelength range (van Geffen et al., 2015, 2019). The swath is divided into 450 individual measurement pixels, which results in a near-nadir resolution of 7×3.5 km. The total NO_2 slant column density is retrieved from the Level 1b UVIS radiance and solar irradiance spectra using ~~Differential Optical Absorption Spectroscopy (DOAS)~~ the DOAS method (Platt and Stutz, 2008). The species fitted by TROPOMI and their corresponding literature cross sections can be found in van Geffen et al. (2019). Tropospheric and stratospheric slant column densities are separated from the total slant column using a data assimilation system based on the TM5-MP chemical transport model, after which they are converted into vertical column densities using a look-up table of altitude dependent air-mass factors (AMF) and information on the vertical distribution of

NO_2 from TM5-MP available with a horizontal resolution of 1° × 1° and a time step of 30 minutes (van Geffen et al., 2019; Boersma et al., 2018; Williams et al., 2017).

The instrument, the NO_2 retrieval and assimilation scheme, and the data product have been described in detail by Veefkind et al. (2012), KNMI (2017), KNMI (2018), Eskes et al. (2019) and van Geffen et al. (2019).

We used reprocessed (RPRO) TROPOMI NO_2 data files, processor version 1.2.2, for the entire study period of ~~15.4.–30.9.2018.~~ 15 April to 30 September 2018. Reprocessed data files are occasionally generated using older sensing data as new processor algorithm versions become available. Version 1.2.x includes retrieval enhancements for high solar zenith angle and snow covered scenes (Eskes et al., 2019), both of which are important for high latitude locations such as Helsinki. The time period of this study did not, however, include any days with snow cover. Additionally, offline (OFFL) and near-real time (~~NRT~~NRTI) NO_2 products are also available. Offline data files are the main TROPOMI data product and are made available within about two weeks from the sensing time, whereas ~~NRT~~ NRTI files are available within 3 hours of measurement time. ~~NRT~~ NRTI files are generated using forecast TM5-MP data rather than analysis data as with offline and reprocessed files (van Geffen et al., 2019), but the differences between the offline/reprocessed and near-real time products are generally small (Lambert et al., 2019).

The TROPOMI NO_2 product used in the comparison was the summed total column, which is the sum of the tropospheric and stratospheric vertical column densities. It was chosen over the total column product, since 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. The summed total column product is described by the data provider as the best physical estimate of the NO_2 vertical column and recommended for comparison to ground-based total column observations (van Geffen et al., 2019). The precision values of the summed total columns used in the analysis stay within the range $(0.5–4.5) \times 10^{15}$ molec. cm^{-2} (or about 10–50 %). The data before 30 April 2018 were downloaded from the Sentinel-5P Expert Users Data Hub (https://s5pexp.copernicus.eu/dhus) as part of the S5P validation team activities, and starting from this date from the S5P Pre-Operations Data Hub (https://s5phub.copernicus.eu/dhus).

Figure 1 shows the TROPOMI NO_2 tropospheric columns over Helsinki averaged over the period ~~15.4.–30.9.2018.~~ 15 April to 30 September 2018. The largest enhancements are visible over the main traffic lanes as well as the Helsinki-Vantaa airport and surrounding area. Overall, the NO_2 levels during weekends (Fig. 1, right panel) are smaller than those observed during weekdays (Fig. 1, left panel) by about 30 %. This is typical for urban sites due to the weekly variability of traffic-related emissions, which are relatively higher during working days (from Monday to Friday). We also note that the NO_2 spatial distribution shown in Fig. 1 is partially affected by systematic wind patterns, which causes the NO_2 levels in the eastern part of the area to become relatively higher than the western part. Fig. S1 in the supplementary material shows the difference between the NO_2 tropospheric columns (normalised to the maximum value in the area) for all wind and low wind speed (less than 3 m s^{-1}) conditions. The pixels in red and blue in Fig. S1 indicate the area where the NO_2 levels are relatively higher or lower, respectively, due to the wind patterns. This is related to the prevailing wind directions from south-west over the Helsinki capital region.

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 $1° \times 1°$ 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. ~~It~~ The CAMS regional ENSEMBLE is a median of seven European CTMs, and the data are provided on a regular $0.1° \times 0.1°$ 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 ~~NO_x~~ NOx in the free troposphere. Data for CAMS global are provided on a regular $0.4° \times 0.4°$ 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 ~~CAMS~~ hybrid CAMS regional/global a-priori profile (henceforth "CAMS a-priori") and the TM5-MP a-priori ~~profiles~~profile (see Sect 2.3). These ratios were ~~available~~ derived on the regular CAMS $0.1° \times 0.1°$ grid ~~, between 30.4. –30.9.2018.~~ 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, $0.1° \times 0.1°$ 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.

2.2 Ground-based NO_2 observations

The NO_2 total columns measured by the ground-based Pandora instrument #105 located in the district of Kumpula, Helsinki, Finland (60.20° N, 24.96° E), are compared to the TROPOMI NO_2 retrievals. The Pandora system is composed of a spectrometer connected by a fibre optic cable to a sensor head with 1.6° FOV (field-of-view). A sun-tracking device allows the optical head to point at the centre of the Sun with a precision of 0.013° (Herman et al., 2009). Pandora performs direct-sun measurements in the UV-VIS spectral range (280–525 nm) and provides NO_2 total vertical column densities, among other products.

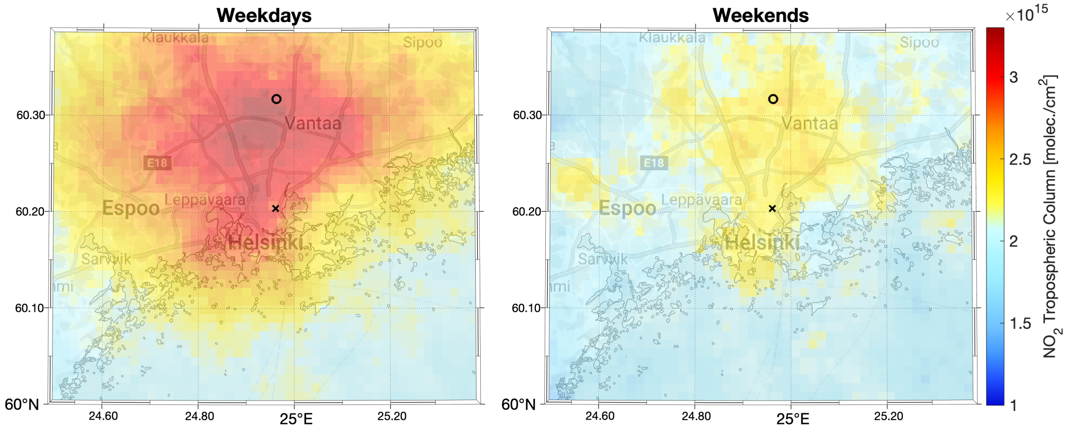

Figure 1. Average TROPOMI NO_2 tropospheric columns over Helsinki during the period ~~15.4.–30.9.2018.~~ 15 April to 30 September 2018. The left and right panels correspond to weekdays and weekends, respectively. The data have been binned and averaged to a 1 km resolution grid. The locations of the Kumpula ground-based station and the Helsinki-Vantaa airport are shown with a black cross and circle, respectively.

The NO_2 total column retrieval is based on the DOAS spectral fitting technique (e.g., Cede et al., 2006), with NO_2 and O_3 being the trace gases fitted. The algorithm derives the relative NO_2 slant column densities (SCDs) from the 400–440 nm spectral band and converts them to absolute SCDs using a statistically estimated reference spectrum obtained using the Minimum-Amount Langley-Extrapolation method (MLE) (Herman et al., 2009).

The Pandora SCD retrieval employs a temperature correction to the cross-sections used in the spectral fitting procedure based on modelled monthly average NO_2 and temperature profiles and high-resolution temperature-dependent cross sections by Vandaele et al. (1998) for NO_2 and Serdyuchenko et al. (2014) for O_3 (as in the TROPOMI retrieval). We note that, while TROPOMI uses the ECMWF operational model as its source for atmospheric temperature profiles (van Geffen et al., 2019), Pandora uses a precalculated atmospheric temperature for a typical NO_2 profile (Cede, 2019). Due to the nature of direct-sun measurements no Ring effect correction is needed for Pandora (Herman et al., 2009).

The NO_2 columns are available about every 1.5 minutes. The full description of the Pandora instrument and the algorithm for the inversion methodology has been presented by Herman et al. (2009). The nominal clear-sky precision of the Pandora NO_2 total column retrievals is in the order of ~~3×10^{14} molec. cm^{-2}~~ 0.3×10^{15} molec. cm^{-2} with an accuracy of about $\pm 1.3 \times 10^{15}$ molec. cm^{-2}. The accuracy depends on the uncertainties in the MLE-calculated reference spectrum, difference between the actual and assumed atmospheric temperature profiles, and uncertainties in the laboratory-determined absorption cross sections (Herman et al., 2009). At typical Helsinki concentrations (6×10^{15} molec. cm^{-2}) and AMF values (2.0) most of the systematic errors are due to uncertainties in the reference spectrum (Sect. 3.3 in Herman et al. (2009)). Pandora #105 is part of the Pandonia global network and the observations used in this paper were processed following the Pandonia procedure and distributed at http://pandonia.net/data/.

The NO_2 surface concentrations available from the Kumpula, Helsinki, air quality (AQ) station were ~~also~~ used in order to analyse the temporal correspondence between surface NO_2 concentrations and TROPOMI vertical columns. ~~The surface~~

~~concentration data were obtained from FMI's databases as hourly averaged measurements. Kumpula station is~~ This station, also known as the SMEAR III station (Järvi et al., 2009), is located close to the Pandora instrument (about 100 m distance), and is classified as a semi-urban site~~, and its surface NO_2 concentrations~~. Nitrogen oxides are measured using ~~an online trace level gas analyser based on the ultraviolet fluorescence method.~~ a chemiluminescence-based analyser (HORIBA APNA-360, Kato and Yoneda, 1997 NO_x and NO measurements from the station are available from the SmartSMEAR online service in intervals of one minute and in units of ppb (https://avaa.tdata.fi/web/smart), while NO_2 measurements are available from the FMI (Finnish Meteorological Institute) measurement database as hourly averaged concentrations in units of $\mu g\,m^{-3}$ (no open access). The air quality data were linearly interpolated to TROPOMI overpass times when compared with collocated Pandora and TROPOMI data. The middle of the one hour averaging period was used as the time stamp for the AQ measurements, as it was found that this resulted in the best correlation with collocated Pandora measurements.

2.3 Methodology

We evaluate the agreement between TROPOMI and Pandora NO_2 vertical column densities by calculating the mean absolute difference (MD), the mean relative difference (MRD), the dispersion (i.e., the standard deviation) of the differences (σ), the correlation coefficient (r), and the ~~slope of an orthogonal linear fit~~ slopes of ordinary least squares and York linear regression fits for the measurements. The MD is defined as the average difference between the TROPOMI and Pandora VCDs in ~~equation~~ Eq. (1), whereas the MRD is the average of these differences when normalised with Pandora's VCD (~~equation~~ Eq. (2)).

$$\mathrm{MD} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{VCD}_{\mathrm{TROPOMI},i} - \mathrm{VCD}_{\mathrm{Pandora},i}) \tag{1}$$

$$\mathrm{MRD} = 100\,\% \times \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{VCD}_{\mathrm{TROPOMI},i} - \mathrm{VCD}_{\mathrm{Pandora},i}}{\mathrm{VCD}_{\mathrm{Pandora},i}} \tag{2}$$

A positive MD or MRD is thus an indication of TROPOMI overestimation, and negative an indication of TROPOMI underestimation. The York linear regression (York et al., 2004) is used alongside the traditional least squares linear regression, since it has been shown to be appropriate in situations where the two sets of data have different levels of uncertainty (Wu and Yu, 2018). We also analyse weekdays and weekends separately, and the results are presented in Sect. 3.

Both TROPOMI and Pandora data were separately filtered according to a set of quality assurance criteria, after which the remaining temporally co-located measurements were compared with each other. For TROPOMI, only measurements with a data quality value ~~(QA)~~ QA >0.75 are used, which disqualifies scenes with a cloud radiance fraction >0.5, some scenes covered by snow or ice, and scenes that have been determined to include errors or problematic retrievals. Further details on the QA value are provided in the appendices of van Geffen et al. (2019). Only TROPOMI pixels including the Helsinki Pandora station were considered for the comparison. Also, only Pandora retrievals with data quality flag value of 0, 1, 10 or 11, corresponding to so-called assured and not-assured high or medium quality data (Cede, 2019), were taken into account. Pandora measurements within 10 minutes of TROPOMI overpass were averaged to get the Pandora-component of the validation data pairs. Wind speed data (average from the four lowest pressure levels: 925, 950, 975 and 1000 hPa) available from the European Centre for Medium-Range Weather Forecasts (ECMWF) as part of the ERA5 reanalysis product (https://cds.climate.copernicus.eu) were associated with each data pair in order to quantify the effect of advection on the NO_2 concentrations. The wind data were linearly interpolated to the Helsinki Pandora station's coordinates and the overpass time of each TROPOMI pixel used in the comparison.

Furthermore, we analyse the effect of the co-location choices on the MD~~and MRDvalue~~ , MRD, standard deviation of the differences σ and correlation coefficient by varying both the maximum distance from the ground-based station and the averaging time interval for Pandora measurements around the S5P overpass time. The results are presented in Sect. 3. When calculating these values for increasing maximum distances, we also required that in all cases the TROPOMI pixel above the station had to have a valid measurement fulfilling our quality criteria.

The effect of using ~~CAMS regional ENSEMBLE instead of TM5-MP~~ 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 the tropospheric ~~column retrievals derived from the~~ air-mass factors derived using CAMS and TM5-MP ~~NO_2 profile shapes~~a-priori profiles, computed on the CAMS-regional grid with 0.1° resolution ~~.~~ (see Sect. 2.1). For each available orbit we used the value of R in the CAMS grid pixel that included the Pandora station. The ~~CAMS a-priori summed column~~new summed column, derived using the CAMS a-priori profile, was then calculated from the tropospheric and stratospheric NO_2 VCDs of the standard L2 product as

$$\mathrm{VCD}_{\mathrm{summed,\,CAMS}} = \mathrm{R} \times \mathrm{VCD}_{\mathrm{tropos,\,TM5\text{-}MP}} + \mathrm{VCD}_{\mathrm{stratos,\,TM5\text{-}MP}} \,. \tag{3}$$

The ~~new~~ 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 NO_2 profile information in the stratosphere. The new TROPOMI-CAMS summed columns calculated using ~~equation~~ Eq. (3) were then also compared to the Pandora total columns, and the results are presented in ~~table 2 and figure~~ Table 2 and Fig. 8. Apart from these two instances, all tables and figures in this paper use standard TROPOMI data products (i.e. based on TM5-MP a-priori profiles).

3 Results

Figure 2 shows the time series of the NO_2 measurements used in the analysis, covering the period ~~April–September~~ April to September 2018. The Pandora NO_2 total columns are shown in their original time resolution (blue ~~line~~dots) as well as averaged 10 minutes around the S5P overpass (red ~~line~~dots). The latter are used in the quantitative comparison to the TROPOMI NO_2 summed columns (yellow ~~line). We note that S5P often has two valid overpasses per day (ranging from 12 to 15 local time) at the latitude of Helsinki (60°N).~~ diamonds). The hourly NO_2 surface concentrations measured at Kumpula AQ station are also shown on the right hand y-axis (black line). The Pandora total columns and the surface concentrations show similar peaks and day-to-day variability (blue dots and black line, respectively), which shows how the Pandora observations are sensitive to the changes in the NO_2 levels occurring at the surface. We note that the collocated TROPOMI and Pandora vertical columns (yellow ~~and red line~~diamonds and red dots, respectively, in Fig. 2) also mostly follow the same day-to-day variability. The

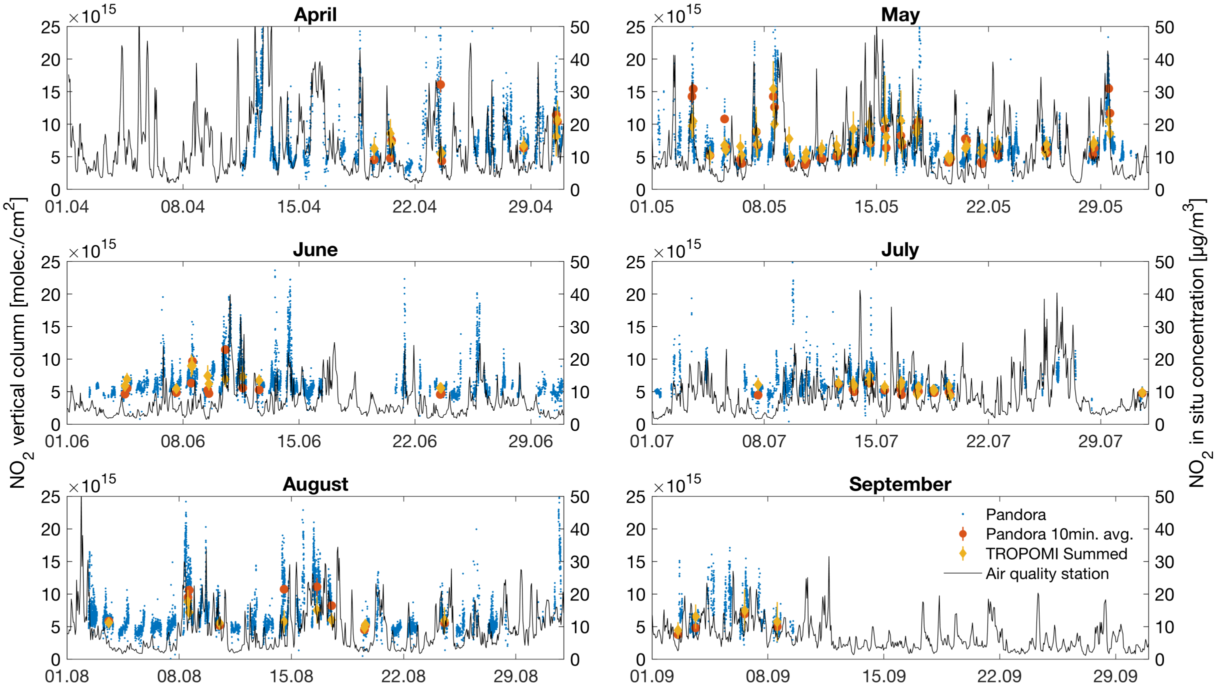

Figure 2. Time series of co-located ~~NO_2~~ Pandora total ~~/~~and TROPOMI summed NO_2 columns during the period ~~15.4.–30.9.2018. The blue line indicates~~ 15 April to 30 September 2018. Blue dots are all the available Pandora observations; ~~the~~ red ~~line indicates~~ dots are the Pandora observations averaged 10 minutes before and after S5P's overpass ~~;~~ (with standard errors of the mean as error bars); yellow ~~line indicates~~ diamonds are TROPOMI summed columns of the pixels ~~located above~~ including the ground-based Pandora station (with retrieval precisions as error bars). The black line (right y-axis) indicates the NO_2 surface concentrations from the in situ measurements at the Kumpula AQ station.

largest differences between TROPOMI and Pandora vertical columns, with TROPOMI smaller than Pandora, correspond to relatively high NO_2 enhancements measured at the surface (black line in Fig. 2). This is expected, as the comparatively large size of the TROPOMI pixels leads to greater spatial averaging compared to the Pandora field-of-view.

In order to further compare satellite- and ground-based collocated observations, ~~Figure~~ Fig. 3 shows the scatter plot between Pandora and TROPOMI total columns from the overpasses presented in Fig. 2. The ~~colour of the filled dots indicates the wind speed, and the red circles correspond to weekend observations~~filled dots correspond to weekdays while the empty circles to the weekends. The colour indicates the corresponding wind speed. The weekend overpasses fall mostly into the bottom-left area of the scatter plot, corresponding to relatively small NO_2 total columns from both Pandora and TROPOMI retrievals. This is expected due to the NO_2 weekly cycle over urban sites, i.e. reduced polluting emissions from traffic during the weekend compared to the weekdays. Furthermore, the overpasses corresponding to high wind speed values (green-yellow ~~dots~~ colours in Fig. 3) also fall into the bottom-left area of the scatter plot. In these cases, the dilution by the wind acts to reduce the NO_2 levels. Overall, the data points are quite close to the one-to-one line, except for some cases with elevated NO_2 total columns

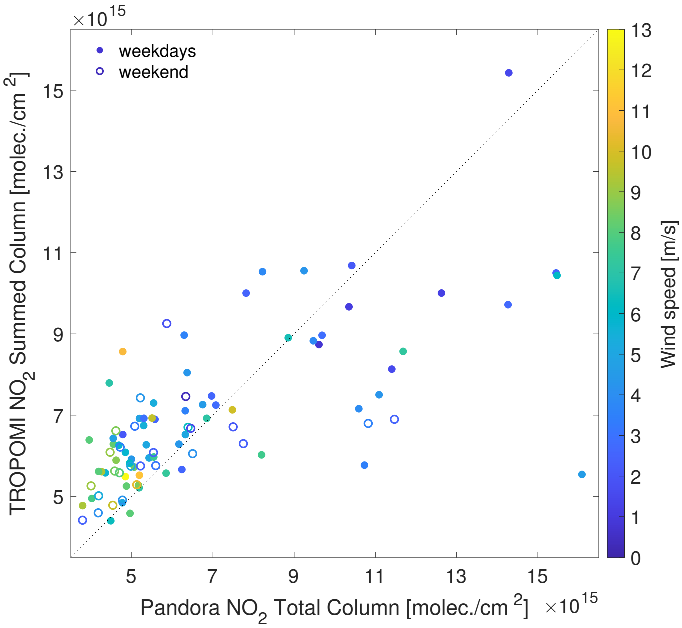

Figure 3. Scatter plot between Pandora and TROPOMI vertical columns. ~~Filled colours indicate the corresponding wind speed,~~ The filled dots correspond to weekdays while ~~red~~ the empty circles ~~correspond~~ to ~~weekend overpasses~~the weekends. The colour indicates the wind speed interpolated at the overpass time. The 1:1 line is plotted as dotted line.

measured by Pandora. These cases correspond to NO_2 enhancements with small wind speed (below $3\,\mathrm{m\,s^{-1}}$), when the spatial dilution effect of TROPOMI's ground footprint as compared to Pandora's narrow field-of-view is especially pronounced.

~~Table~~Table 1 summarises the results of the comparison between TROPOMI and Pandora in terms of mean relative difference (MRD), mean difference (MD), standard deviation of the difference (σ), correlation coefficient (r), ~~orthogonal linear least squares fit slope and~~ slopes of linear and York regression fits, and number of overpasses (n). The overall MRD and MD values are (9.9 ± 2.6) % and ~~$(1.2 \pm 2.2) \times 10^{14}$ molec. cm^{-2}~~$(0.12 \pm 0.22) \times 10^{15}$ molec. cm^{-2}, respectively, meaning that on average TROPOMI slightly overestimates the NO_2 total columns. The dispersion of these absolute differences, calculated as their standard deviation, is 2.2×10^{15} molec. cm^{-2}. The correlation coefficient is high ($r = 0.68$). When considering only weekdays, the MD and MRD values become slightly smaller (MRD=(9.0 ± 3.3) %) but the change remains within the uncertainties. This is expected, as weekday observations contain a number of collocations where the difference between TROPOMI and Pandora vertical columns is exceedingly negative (Fig. 3), corresponding to NO_2 enhancements measured by Pandora. Correspondingly, the MRD and MD values for the weekend (typically associated with lower NO_2 levels) are larger. When taking into account only overpasses with Pandora NO_2 columns larger than ~~10^{16} molec. cm^{-2}~~10×10^{15} molec. cm^{-2}, the bias becomes exceedingly negative (about -28 % or ~~$(-36.0 \pm 7.0) \times 10^{14}$ molec. cm^{-2}~~$(-3.60 \pm 0.70) \times 10^{15}$ molec. cm^{-2}), meaning that TROPOMI underestimates the NO_2 total columns when NO_2 enhancements occur. When considering overpasses below that threshold, the bias is positive (about 17 %). These two effects partially cancel each other when the data set is considered as a whole. Figure S2

Table 1. Statistics of the comparison between TROPOMI and Pandora NO_2 total columns. The ~~values outside (inside)~~ uncertainties are the ~~parentheses~~ corresponding standard errors of the mean. The uncertainty estimates used in the York fit are ~~obtained~~ pixel-specific precisions for ~~Pandora retrievals with at least medium~~ TROPOMI (~~high~~included in the data product)~~quality~~, and standard errors of the mean for Pandora as calculated for the set of measurements within 10 minutes of the S5P overpass.

	MRD^a ~~(%)~~	MD^b ~~($\times 10^{14}$ molec. cm^{-2})~~	~~r^c~~ σ^c	r^d	slope ~~^{d}~~ $_{LS}^e$	~~n^e~~ slope$_Y^f$	n^g
all data	9.9 ± 2.6 ~~(10.1 ± 3.6)~~	~~1.2 ± 2.2 (0.8 ± 3.2)~~ 0.12 ± 0.22	2.2	0.68 ~~(0.66)~~	~~0.51 (0.50)~~ 0.42	0.36	94 ~~(~~
Pandora HQ^h	10.1 ± 3.6	0.08 ± 0.32	2.4	0.66	0.41	0.33	56 ~~)~~
weekdays	9.0 ± 3.3	~~0.2 ± 2.9~~ 0.02 ± 0.29	2.3	0.68	~~0.51~~ 0.42	0.37	67
weekends	12.1 ± 4.4	~~3.8 ± 3.2~~ 0.38 ± 0.32	1.7	0.46	~~0.34~~ 0.26	0.32	27
Pandora high ~~^{f}~~ ^{i}	~~-28.2 ± 4.8~~ -28.1 ± 4.8	~~-36.0 ± 7.0~~ -3.60 ± 0.70	2.7	0.31	~~1.80~~ 0.38	0.19	15
Pandora low ~~^{g}~~ ^{j}	17.1 ± 2.2	~~8.3 ± 1.2~~ 0.83 ± 0.12	1.1	0.72	~~0.95~~ 0.69	0.61	79

^{a} Mean Relative Difference [%]; ^{b} Mean Difference [$\times 10^{15}$ molec. cm^{-2}]; ^{c} Standard deviation of absolute bias [$\times 10^{15}$ molec. cm^{-2}]; ^{d} Correlation coefficient; ^{e} Least squares linear fit slope; ^{f} York linear fit slope; ^{g} Number of collocations; ^{h} High quality Pandora observations (QA value 0 or 10); ^{i} Pandora NO_2 total columns $\geq 10 \times 10^{15}$ molec. cm^{-2}; ^{j} Pandora NO_2 total columns $< 10 \times 10^{15}$ molec. cm^{-2}.

in the Supplement illustrates in more details how the bias changes from positive (about 10^{15} molec. cm^{-2}) to negative (almost -4×10^{15} molec. cm^{-2}) for increasing values of Pandora NO_2 total columns. The standard deviation of differences and the correlation coefficient are smaller for weekend overpasses and low Pandora NO_2 total columns compared to weekdays and high Pandora NO_2 total columns. We also note that taking into account only Pandora retrievals with the highest quality flagging (0 or 10) does not have a substantial effect on the results of the comparison (~~shown in parentheses in Tab.~~second row of Table 1), but it reduces the amount of data available for the comparison by about ~~30 %~~ 40 % (as compared to the case where also medium quality data are included).

Figure 4 shows how the choice of the overpass criteria affects the calculated MD value (a similar plot for the MRD is shown in ~~Figure~~Fig. ~~S2~~ S3 of the Supplement). In the analysis presented so far we have included ~~measurements from only~~ only measurements from those TROPOMI pixels which include the Pandora ground-based station. It is also possible to average the contribution from all those pixels which fall within a certain distance from the station. Figure 4 (upper panel) shows how the MD gradually shifts towards negative values (from about 0.1×10^{15} to -0.5×10^{15} molec. cm^{-2}) when the radius increases from 5 to 30 km. This suggests that averaging over a larger area causes the resulting TROPOMI vertical columns (used in the comparison) to become smaller than those obtained from the single overlaying pixel because of the inhomogeneous spatial distribution of NO_2, so that the mean concentrations decrease with increasing distance. The MD (and MRD) value for the overlaying pixel criterion is very similar to the value obtained for the distance of 5 km, even if the number of collocations is not exactly the same. ~~The overall effect of the spatial collocation choice stays within about 6 percentage points (or 6×10^{14} molec. cm^{-2}).~~ Also, the correlation coefficient value decreases and the standard deviation of the differences increases while the radius increases (upper panels in Fig. S4 and S5, respectively, in the supplement).

Similarly, Fig. 4 (lower panel) shows how the MD value changes when the Pandora observations are averaged over an increasing time range, from 5 to 55 min around the overpass time of the satellite. The MD value increases with increasing temporal averaging interval by about 0.3×10^{15} molec. cm^{-2} (2 percentage points). Averaging over an increasing time range generally slightly reduces the Pandora total column values used in the comparison with TROPOMI, making the MD more positive. The ~~variability remains around 2 percentage points (or 3×10^{14} molec. cm^{-2}).~~ correlation coefficient value decreases until 20 km radius while slightly increasing for larger radius values while the standard deviation of the differences behaves in the opposite way (lower panels in Fig. S4 and S5, respectively, in the supplement).

Figure S6 in the supplement includes the absolute differences between TROPOMI and Pandora NO_2 total columns as a function of TROPOMI SZA (solar zenith angle) and CRF (cloud radiance fraction) (upper and lower panel, respectively) within the range of values allowed after the TROPOMI data screening (QA value >0.75). The differences between satellite- and ground-based retrievals for SZA above 45° are generally larger (between −3 and 1×10^{15} molec. cm^{-2}) than for smaller values (0 to 1×10^{15} molec. cm^{-2}). 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 NO_2 daily variability in the time range between about 12 and 15 LT. The S5P overpass time typically corresponds to the NO_2 daily local minimum (between the morning and afternoon peaks due to commuter traffic), observed for example in the NO_2 surface concentration measurements from Kumpula AQ site (Fig. S7). Figure 5 (upper panel) shows TROPOMI and Pandora NO_2 total columns as a function of the time of the day between 12 and 15 LT. Both datasets show an enhancement around 13:30 LT and lower NO_2 levels before and after. The relative differences between TROPOMI and Pandora NO_2 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 before 13:30 LT (about 30 %) than afterwards (21 %). Increasing time of the day also corresponds to increasing pixel number (colour of the filled dots in Fig. 5, lower panel), with the first overpass of the day corresponding to the left side of the swath (smaller pixel numbers) and 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 centre of the swath) was observed.

In order to better compare the temporal variability of the NO_2 vertical columns and surface concentrations, we employ a simple empirical conversion based on the linear regression between Pandora vertical columns and surface concentrations measured at the Kumpula AQ site, at the satellite overpass time (Fig. 6, left panel). From the results of the linear fit (showing high correlation, r = ~~0.71~~0.74), we convert the surface concentrations into total columns and compare the results to the TROPOMI and Pandora time series (Fig. 6, right panel). We note how the three datasets show a very similar temporal variability, with NO_2 peaks occurring during the same days. We particularly note NO_2 enhancements in May and during the first half of August.

We also analyse the NO_2 weekly cycle as seen from the different datasets. Figure 7 shows the Pandora NO_2 total columns, TROPOMI summed and tropospheric columns and surface concentrations at the Kumpula air quality station as a function of the day of the week. The data are normalised by the corresponding weekly mean value. We note that all datasets show smaller values on Saturdays and Sundays, as expected from the weekly cycle of NO_x emissions typical of urban sites. The NO_2 surface concentrations show about ~~50 %~~ 30–50 % smaller values in the weekend compared to the weekly average, while TROPOMI tropospheric columns are about ~~30 %~~ 20–30 % lower. Pandora and TROPOMI summed NO_2 vertical columns are also lower in

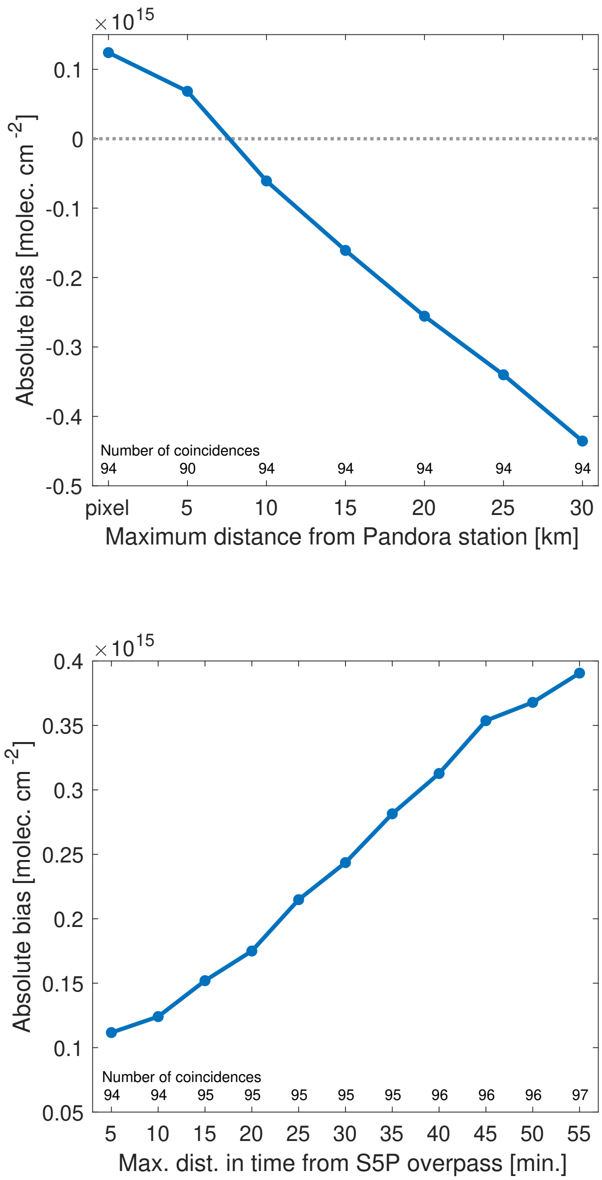

Figure 4. Mean absolute difference between TROPOMI summed and Pandora total columns as a function of the maximum distance between the centre of the pixel and the ground-based station (upper panel), and as a function of the maximum time difference from TROPOMI overpass (lower panel). The number of coincidences for different collocation criteria are shown above the x-axis. Note that in the upper panel we also require that the TROPOMI pixel above Pandora station contains a valid measurement (QA value >0.75). Thus the number of coincidences does not increase with distance.

the weekends (compared to the corresponding weekly means), but only by about ~~15–20 %~~10–20 %. This is because no weekend effect is expected in the stratospheric fraction of the NO_2 column. Surface NO_2 concentration measurements can be expected

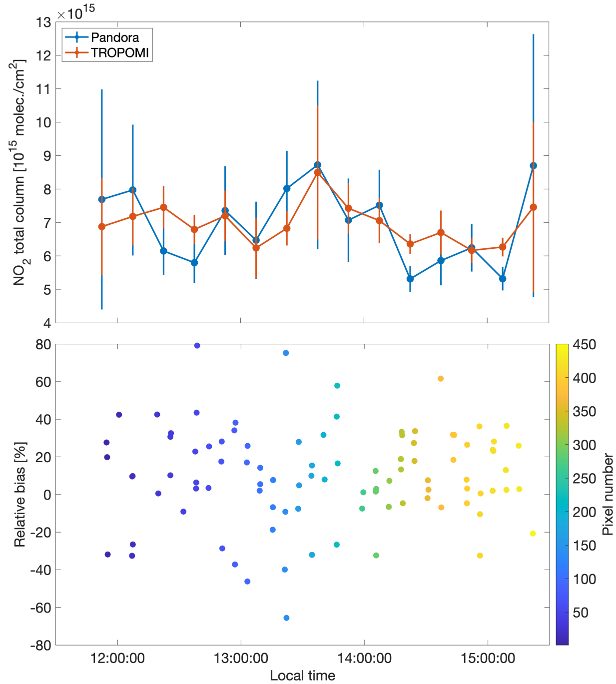

Figure 5. Upper panel: TROPOMI NO_2 summed columns and Pandora total columns as a function of the time of the day between about 12 and 15 LT. The error bars are the standard deviation of the mean. Lower panel: Relative difference between TROPOMI summed columns and Pandora total columns as a function of the time of the day. Filled colours correspond to the TROPOMI pixel number.

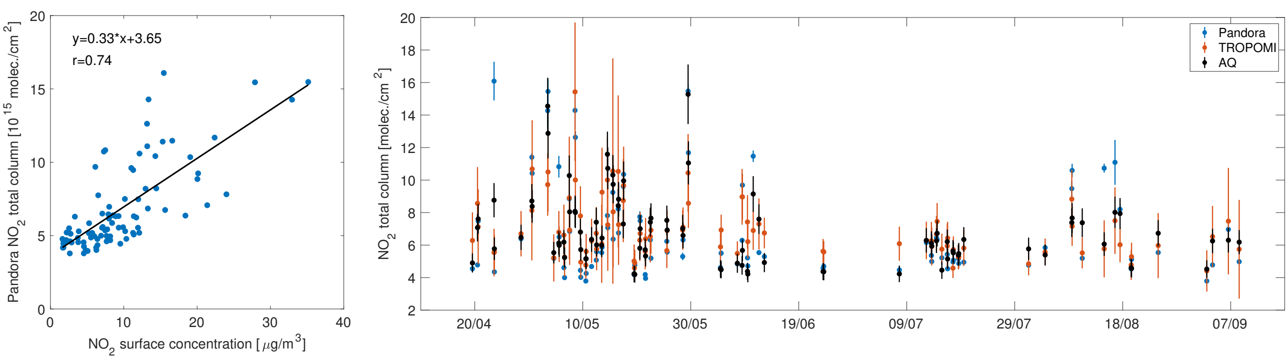

Figure 6. ~~NO_2 time~~ Right panel: Time series of NO_2 total columns from Pandora (blue), TROPOMI (red) and Kumpula AQ station (black) at the satellite overpass time~~(right panel)~~. The surface concentrations are empirically converted to total columns using the results of the linear regression between Pandora total columns and surface concentration data (left panel).

to show a larger difference between weekend and weekdays due to their greater sensitivity to changes in polluting emissions at the surface (especially from traffic in the urban environment). The results are consistent with those found using nine years of OMI NO_2 observations in Helsinki (Ialongo et al., 2016).

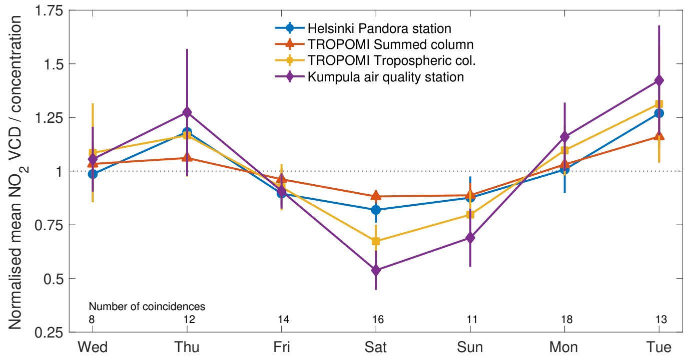

Figure 7. NO_2 weekly cycle in Helsinki. The ~~time-averaged~~ average of temporally co-located values for each day of the week for Pandora total columns (blue line), TROPOMI summed ~~(red line)~~ and tropospheric columns (red and yellow line, respectively)~~columns~~, and surface concentrations as measured at the Kumpula AQ ~~site~~ station (purple line) ~~for each day of the week~~ are shown. ~~The~~ Error bars represent corresponding standard errors of the mean. All values have been normalised by the ~~corresponding~~ weekly mean of each data set.

Finally, we evaluate the effect of using the NO_2 a-priori profiles derived from the high-resolution CAMS regional ENSEMBLE model, instead of profiles from the TM5-MP CTM as used in TROPOMI's standard product, in the calculation of NO_2 vertical column densities. ~~In Fig.8 we compare both the~~ Figure 8 shows the comparison of the standard product summed columns and the summed columns derived using the CAMS a-priori profiles, calculated as described in ~~Section~~Sect. 2.3, to the Pandora total columns (analogously to Fig. 3). Only those overpasses (n=75) for which both a-priori summed columns were available were included in the comparison. The statistics are presented in Table 2 and the corresponding time series in Fig. ~~S3~~ S8 of the supplement. 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 ~~worsens the agreement with Pandora by increasing the positive bias~~ 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 only at the 52 % significance level. Thus, on average, the use of CAMS profiles does not significantly improve 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 ~~an orthogonal~~ a least squares linear fit is ~~0.52~~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 ~~0.58~~0.52, respectively. This improvement is more evident for high values of the Pandora NO_2 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. ~~S3 in~~ S8 of the supplement further ~~shows~~ show

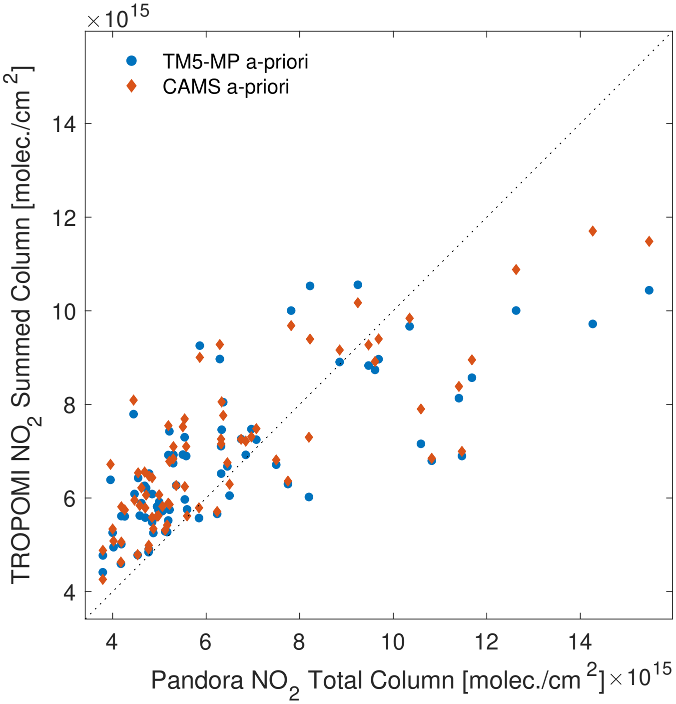

Figure 8. Scatter plot between Pandora and TROPOMI summed columns derived using CAMS regional ~~ENSEMBLE~~ and TM5-MP a-priori NO_2 profiles (blue dots and red diamonds, respectively). The comparison includes only those overpasses for which both summed columns were available at the same time during the time interval ~~30.4.–30.9.2018.~~ 30 April to 30 September 2018. The 1:1 line is plotted as dotted line.

how using the high-resolution CAMS profiles increases the TROPOMI tropospheric columns so that the summed columns (yellow ~~line~~dots) become closer to Pandora's peak values (blue ~~line~~dots), corresponding to episodes of NO_2 enhancement, but that overall the difference between the summed columns obtained using TM5-MP and CAMS remains mostly within the uncertainties of the TROPOMI NO_2 retrieval.

4 Conclusions

We showed the results of the comparison between satellite-based TROPOMI/S5P NO_2 products and ground-based observations at a medium-sized urban site, Helsinki (Finland). We find that the differences between the total columns derived from the TROPOMI and Pandora instruments are on average ~~less than~~ around 10 % (or ~~1.2×10^{14} molec. cm^{-2}~~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 NO_2 tropospheric column; ESA, 2017). We also note that the~~ (25–50 % for the NO_2 tro
The day-to-day and weekly NO_2 variability (typical of urban sites) is reproduced well by the TROPOMI retrievals, similarly to

Table 2. Statistics of the comparisons between TROPOMI summed columns calculated using two different a-priori NO_2 profiles (TM5-MP and CAMS regional~~ENSEMBLE~~) and Pandora total columns during ~~30.4~~30 April to 30 September 2018. The uncertainties are given as standard errors of the mean. ~~–30.9.2018.~~

	MRD[a] ~~(%)~~	MD[b] ~~($\times 10^{14}$ molec. cm^{-2})~~	~~r[c]~~ σ[c]	r[d]	slope~~[d]~~ $^{e}_{LS}$	n~~[e]~~[f]
TM5-MP	11.5 ± 2.7	~~3.1 ± 2.0~~ 0.31 ± 0.20	1.8	0.74	~~0.52~~ 0.45	75
CAMS	14.0 ± 2.6	~~4.9 ± 1.8~~ 0.49 ± 0.18	1.6	0.80	~~0.58~~ 0.52	75
TM5-MP high[g]	-28.5 ± 3.3	-3.48 ± 0.44	1.3	0.67	0.55	9
CAMS high[g]	-23.7 ± 3.5	-2.86 ± 0.41	1.2	0.77	0.82	9
TM5-MP low[h]	16.9 ± 2.3	0.83 ± 0.13	1.0	0.75	0.71	66
CAMS low[h]	19.1 ± 2.3	0.95 ± 0.12	0.97	0.78	0.72	66

[a] Mean Relative Difference [%]; [b] Mean Difference [$\times 10^{15}$ molec. cm^{-2}]; [c] Standard deviation of absolute bias [$\times 10^{15}$ molec. cm^{-2}]; [d] Correlation coefficient; [e] Least squares linear fit slope; [f] Number of collocations; [g] Pandora NO_2 total columns $\geq 10 \times 10^{15}$ molec. cm^{-2}; [h] Pandora NO_2 total columns $< 10 \times 10^{15}$ molec. cm^{-2}.

Pandora and in situ surface observations from the local air quality station. This confirms that the satellite-based TROPOMI/S5P NO_2 retrievals are sensitive to changes in air pollution levels occurring at the surface.

In general, we find that TROPOMI NO_2 summed columns are smaller than Pandora total columns for relatively high concentrations, while low values are overestimated. ~~We find this partially related to the~~ This is partly due to the low resolution of the TM5-MP ~~model~~ profile shapes used ~~in the TROPOMI retrieval~~ to compute the tropospheric air-mass factors and thus the ~~tropospheric~~ vertical columns. Because of the relatively coarse resolution of the TM5-MP a-priori profiles in the standard product, TROPOMI tropospheric columns are expected to have a negative bias over polluted areas where the peak in the NO_2 profile is close to the surface, and where the boundary layer column is underestimated in the a-priori. Also, the time variability of the column amounts at the measurement site may be underestimated due to the a-priori. In the same way, the concentrations away from major sources may be somewhat overestimated. In Helsinki we find that replacing the original profiles with those derived from the high-resolution ~~regional CAMS~~ CAMS regional ensemble model increases the TROPOMI NO_2 tropospheric columns and partly reduces the discrepancy between TROPOMI and Pandora VCDs for ~~situations with~~ episodes of relatively high NO_2 concentrations~~.~~, while increasing the correlation and linear fit slope. On the other hand, the agreement does not significantly improve on average or for lower values of NO_2 vertical columns. Overall, the change in bias remains mostly within the uncertainties.

The overestimation of low NO_2 columns suggests a possible overestimation of the stratospheric fraction of the column. Also, replacing the surface reflectance 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 NO_2 vertical columns with the ground-based observations.

As compared to previous satellite-based instruments such as OMI, the ~~mean~~ bias against ground-based observations in Helsinki is ~~of the same order of magnitude~~ similar on average (±5 % under clear sky conditions for OMI, Ialongo et al., 2016), while the correlation coefficient is generally higher for TROPOMI (r = 0.68 for TROPOMI and r = 0.5 for OMI, see Ialongo et al., 2016). The correlation between Pandora and TROPOMI NO_2 retrievals is also in line with the results obtained ~~by Griffin et al. (2019)~~ 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.

Overall, the analysis of TROPOMI NO_2 observations in the Helsinki area shows high correlation with ground-based observations, as well as demonstrates TROPOMI's capability to properly reproduce the temporal (day-to-day and weekly) variability of the surface NO_2 concentrations. This is a confirmation that satellite-based observations can bring additional information on the temporal and spatial variability of ~~NO_2~~ NO_2 in the neighbourhood of major cities, in addition to traditional air quality measurements.

Data availability. The re-processed TROPOMI data before 30 April 2018 were downloaded from the Sentinel-5P Expert Users Data Hub (https://s5pexp.copernicus.eu/dhus, no open access) as part of the S5P validation team activities, and after that date from the S5P Pre-Operations Data Hub (https://s5phub.copernicus.eu/dhus, open access). Pandora #105 total column data belong to the Pandonia network and are available at . Surface concentration data at the Kumpula air quality station were obtained from the FMI measurement database (no open access); an alternative source is the SmartSMEAR service (https://avaa.tdata.fi/web/smart, open access). 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). The wind data are part of the ECMWF ERA5 reanalysis product and were downloaded from the Climate Data Store (https://cds.climate.copernicus.eu).

Author contributions. I.I. and H.V. designed the content of the paper ans carried on the data analysis. E.H. and J.D. provided their expertise on TROPOMI NO2 retrievals and provided the CAMS model calculations. J.H. was responsible for the Pandora data. All the authors participated in writing the manuscripts.

Competing interests. No competing interests are present.

Acknowledgements. The research activity has been supported by the EU Horizon 2020 project E-Shape (grant agreement n.820852), by the Academy of Finland (Project ILMApilot, n.303876) and by ESA EO Science for Society funding scheme (Project DACES).

References

Beirle, S., Platt, U., Wenig, M., and Wagner, T.: Weekly cycle of NO_2 by GOME measurements: a signature of anthropogenic sources, Atmospheric Chemistry and Physics, 3, 2225–2232, https://doi.org/10.5194/acp-3-2225-2003, 2003.

Beirle, S., Boersma, K. F., Platt, U., Lawrence, M. G., and Wagner, T.: Megacity Emissions and Lifetimes of Nitrogen Oxides Probed from Space, Science, 333, 1737–1739, https://doi.org/10.1126/science.1207824, 2011.

Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., and Eskes, H. J.: Validation of urban NO_2 concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities, Atmospheric Chemistry and Physics, 9, 3867–3879, https://doi.org/10.5194/acp-9-3867-2009, 2009.

Boersma, K. F., Eskes, H. J., Richter, A., De Smedt, I., Lorente, A., Beirle, S., van Geffen, J. H. G. M., Zara, M., Peters, E., Van Roozendael, M., Wagner, T., Maasakkers, J. D., van der A, R. J., Nightingale, J., De Rudder, A., Irie, H., Pinardi, G., Lambert, J.-C., and Compernolle, S. C.: Improving algorithms and uncertainty estimates for satellite NO_2 retrievals: results from the quality assurance for the essential climate variables (QA4ECV) project, Atmospheric Measurement Techniques, 11, 6651–6678, https://doi.org/10.5194/amt-11-6651-2018, https://www.atmos-meas-tech.net/11/6651/2018/, 2018.

Castellanos, P. and Boersma, K. F.: Reductions in nitrogen oxides over Europe driven by environmental policy and economic recession, Scientific Reports, 2, 265, https://doi.org/10.1038/srep00265, 2012.

Cede, A.: Manual for Blick Software Suite 1.6, LuftBlick OG, Mutters, Austria, http://pandonia.net/media/documents/BlickSoftwareSuite_Manual_v11.pdf, version 11, 2019.

Cede, A., Herman, J., Richter, A., Krotkov, N., and Burrows, J.: Measurements of nitrogen dioxide total column amounts using a Brewer double spectrophotometer in direct Sun mode, Journal of Geophysical Research: Atmospheres, 111, D5, https://doi.org/10.1029/2005JD006585, 2006.

de Foy, B., Lu, Z., and Streets, D. G.: Impacts of control strategies, the Great Recession and weekday variations on NO_2 columns above North American cities, Atmospheric Environment, 138, 74–86, https://doi.org/https://doi.org/10.1016/j.atmosenv.2016.04.038, 2016.

Duncan, B. N., Lamsal, L. N., Thompson, A. M., Yoshida, Y., Lu, Z., Streets, D. G., Hurwitz, M. M., and Pickering, K. E.: A space-based, high-resolution view of notable changes in urban NOx pollution around the world (2005–2014), Journal of Geophysical Research: Atmospheres, 121, 976–996, https://doi.org/10.1002/2015JD024121, 2016.

ESA: Sentinel-5 Precursor Calibration and Validation Plan for the Operational Phase, Tech. Rep. ESA-EOPG-CSCOP-PL-0073, European Space Agency (ESA), https://sentinels.copernicus.eu/documents/247904/2474724/Sentinel-5P-Calibration-and-Validation-Plan.pdf, issue 1, 2017.

Eskes, H., van Geffen, J., Boersma, F., Eichmann, K.-U., Apituley, A., Pedergnana, M., Sneep, M., Veefkind, J. P., and Loyola, D.: Sentinel-5 precursor/TROPOMI Level 2 Product User Manual Nitrogendioxide, Tech. Rep. S5P-KNMI-L2-0021-MA, Koninklijk Nederlands Meteorologisch Instituut (KNMI), https://sentinels.copernicus.eu/documents/247904/2474726/Sentinel-5P-Level-2-Product-User-Manual-Nitrogen-Dioxide, CI-7570-PUM, issue 3.0.0, 2019.

Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A.-M., Diamantakis, M., Engelen, R. J., Gaudel, A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V.-H., Richter, A., Schultz, M. G., Stein, O., and Tsikerdekis, A.: Tropospheric chemistry in the Integrated Forecasting System of ECMWF, Geoscientific Model Development, 8, 975–1003, https://doi.org/10.5194/gmd-8-975-2015, https://www.geosci-model-dev.net/8/975/2015/, 2015.

Griffin, D., Zhao, X., McLinden, C. A., Boersma, F., Bourassa, A., Dammers, E., Degenstein, D., Eskes, H., Fehr, L., Fioletov, V., Hayden, K., Kharol, S. K., Li, S.-M., Makar, P., Martin, R. V., Mihele, C., Mittermeier, R. L., Krotkov, N., Sneep, M., Lamsal, L. N., ter Linden, M., van Geffen, J., Veefkind, P., and Wolde, M.: High-Resolution Mapping of Nitrogen Dioxide With TROPOMI: First Results and Validation Over the Canadian Oil Sands, Geophysical Research Letters, 46, 1049–1060, https://doi.org/10.1029/2018GL081095, 2019.

Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M., and Abuhassan, N.: NO_2 column amounts from ground-based Pandora and MF-DOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, Journal of Geophysical Research: Atmospheres, 114, D13, https://doi.org/10.1029/2009JD011848, 2009.

Ialongo, I., Herman, J., Krotkov, N., Lamsal, L., Boersma, K. F., Hovila, J., and Tamminen, J.: Comparison of OMI NO_2 observations and their seasonal and weekly cycles with ground-based measurements in Helsinki, Atmospheric Measurement Techniques, 9, 5203–5212, https://doi.org/10.5194/amt-9-5203-2016, 2016.

Järvi, L., Hannuniemi, H., Hussein, T., Junninen, H., Aalto, P. P., Hillamo, R., Mäkelä, T., Keronen, P., Siivola, E., Vesala, T., and Kulmala, M.: The urban measurement station SMEAR III: Continuous monitoring of air pollution and surface-atmosphere interactions in Helsinki, Finland, Boreal Environment Research, 14, 86–109, http://www.borenv.net/BER/pdfs/ber14/ber14A-086.pdf, 2009.

Kato, J. and Yoneda, A.: Air Pollution Monitoring Systems, AP-360 Series, Tech. rep., HORIBA, Inc., Kyoto, Japan, https://www.horiba.com/uploads/media/RE01-06-029.pdf, 1997.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2008JD010290, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD010290, 2008.

KNMI: Algorithm theoretical basis document for the TROPOMI L01b data processor, Tech. Rep. S5P-KNMI-L01B-0009-SD, Koninklijk Nederlands Meteorologisch Instituut (KNMI), https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-TROPOMI-Level-1B-ATBD, CI-6480-ATBD, issue 8.0.0, 2017.

KNMI: Sentinel 5 precursor/TROPOMI KNMI and SRON level 2 Input Output Data Definition, Tech. Rep. S5P-KNMI-L2-0009-SD, Koninklijk Nederlands Meteorologisch Instituut (KNMI), https://sentinel.esa.int/documents/247904/3119978/Sentinel-5P-Level-2-Input-Output-Data-Definition, issue 10.0.0, 2018.

Krotkov, N. A., McLinden, C. A., Li, C., Lamsal, L. N., Celarier, E. A., Marchenko, S. V., Swartz, W. H., Bucsela, E. J., Joiner, J., Duncan, B. N., Boersma, K. F., Veefkind, J. P., Levelt, P. F., Fioletov, V. E., Dickerson, R. R., He, H., Lu, Z., and Streets, D. G.: Aura OMI observations of regional SO_2 and NO_2 pollution changes from 2005 to 2015, Atmospheric Chemistry and Physics, 16, 4605–4629, https://doi.org/10.5194/acp-16-4605-2016, 2016.

Lambert, J.-C., Keppens, A., Hubert, D., Langerock, B., Eichmann, K.-U., Kleipool, Q., Sneep, M., Verhoelst, T., Wagner, T., Weber, M., Ahn, C., Argyrouli, A., Balis, D., Chan, K., Compernolle, S., De Smedt, I., Eskes, H., Fjæraa, A., Garane, K., Gleason, J., Goutail, F., Granville, J., Hedelt, P., Heue, K.-P., Jaross, G., Koukouli, M., Landgraf, J., Lutz, R., Niemejer, S., Pazmiño, A., Pinardi, G., Pommereau, J.-P., Richter, A., Rozemeijer, N., Sha, M., Stein Zweers, D., Theys, N., Tilstra, G., Torres, O., Valks, P., Vigouroux, C., and Wang, P.: Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products – #04: April 2018 – August 2019., S5P MPC Routine Operations Consolidated Validation Report series, Issue #04, Version 04.0.0, 129pp, http://s5p-mpc-vdaf.aeronomie.be/ProjectDir/reports/pdf/S5P-MPC-IASB-ROCVR-04.0.0-20190923_FINAL.pdf, 2019.

Lamsal, L. N., Duncan, B. N., Yoshida, Y., Krotkov, N. A., Pickering, K. E., Streets, D. G., and Lu, Z.: U.S. NO_2 trends (2005–2013): EPA Air Quality System (AQS) data versus improved observations from the Ozone Monitoring Instrument (OMI), Atmospheric Environment, 110, 130–143, https://doi.org/10.1016/j.atmosenv.2015.03.055, 2015.

Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H.: The ozone monitoring instrument, IEEE Transactions on Geoscience and Remote Sensing, 44, 1093–1101, https://doi.org/10.1109/TGRS.2006.872333, 2006.

Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., Stein Zweers, D. C., Duncan, B. N., Streets, D. G., Eskes, H., van der A, R., McLinden, C., Fioletov, V., Carn, S., de Laat, J., DeLand, M., Marchenko, S., McPeters, R., Ziemke, J., Fu, D., Liu, X., Pickering, K., Apituley, A., González Abad, G., Arola, A., Boersma, F., Chan Miller, C., Chance, K., de Graaf, M., Hakkarainen, J., Hassinen, S., Ialongo, I., Kleipool, Q., Krotkov, N., Li, C., Lamsal, L., Newman, P., Nowlan, C., Suleiman, R., Tilstra, L. G., Torres, O., Wang, H., and Wargan, K.: The Ozone Monitoring Instrument: overview of 14 years in space, Atmospheric Chemistry and Physics, 18, 5699–5745, https://doi.org/10.5194/acp-18-5699-2018, 2018.

Liu, F., Beirle, S., Zhang, Q., van der A, R. J., Zheng, B., Tong, D., and He, K.: NOx emission trends over Chinese cities estimated from OMI observations during 2005 to 2015, Atmospheric Chemistry and Physics, 17, 9261–9275, https://doi.org/10.5194/acp-17-9261-2017, 2017.

Lu, Z., Streets, D. G., de Foy, B., Lamsal, L. N., Duncan, B. N., and Xing, J.: Emissions of nitrogen oxides from US urban areas: estimation from Ozone Monitoring Instrument retrievals for 2005–2014, Atmospheric Chemistry and Physics, 15, 10 367–10 383, https://doi.org/10.5194/acp-15-10367-2015, 2015.

Marécal, V., Peuch, V. H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., van der Gon, H. A. C. D., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouille, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, A., van Velthoven, P., van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geoscientific Model Development, 8, 2777–2813, https://doi.org/10.5194/gmd-8-2777-2015, 2015.

Météo-France: Regional Production, Description of the operational models and of the ENSEMBLE system, Tech. rep., Copernicus Atmosphere Monitoring Service (CAMS), https://atmosphere.copernicus.eu/sites/default/files/2018-02/CAMS50_factsheet_201610_v2.pdf, Ref.: CAMS_50_2015SC1_Models_Factsheets_201610_v2, 2016.

Platt, U. and Stutz, J.: Differential Optical Absorption Spectroscopy: Principles and Applications, Springer-Verlag, Berlin, Germany, https://doi.org/10.1007/978-3-540-75776-4, 2008.

Serdyuchenko, A., Gorshelev, V., Weber, M., Chehade, W., and Burrows, J. P.: High spectral resolution ozone absorption cross-sections — Part 2: Temperature dependence, Atmospheric Measurement Techniques, 7, 625–636, https://doi.org/10.5194/amt-7-625-2014, 2014.

Streets, D., Canty, T., Carmichael, G., de Foy, B., Dickerson, R., Duncan, B., Edwards, D., Haynes, J., Henze, D., Houyoux, M., Jacob, D., Krotkov, N., Lamsal, L., Liu, Y., Lu, Z., Martin, R., Pfister, G., Pinder, R., Salawitch, R., and Wecht, K.: Emissions estimation from satellite retrievals: A review of current capability, Atmospheric Environment, 77, 1011–1042, https://doi.org/10.1016/j.atmosenv.2013.05.051, 2013.

US-EPA: Nitrogen Dioxide (NO2) Pollution, available at: https:// www.epa.gov/no2-pollution, latest access 6 August 2019, 2019.

van Geffen, J. H. G. M., Boersma, K. F., Van Roozendael, M., Hendrick, F., Mahieu, E., De Smedt, I., Sneep, M., and Veefkind, J. P.: Improved spectral fitting of nitrogen dioxide from OMI in the 405–465 nm window, Atmospheric Measurement Techniques, 8, 1685–1699, https://doi.org/10.5194/amt-8-1685-2015, 2015.

van Geffen, J. H. G. M., Eskes, H. J., Boersma, K. F., Maasakkers, J. D., and Veefkind, J. P.: TROPOMI ATBD of the total and tropospheric NO_2 data products, Tech. Rep. S5P-KNMI-L2-0005-RP, Koninklijk Nederlands Meteorologisch Instituut (KNMI), https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-NO2-data-products, CI-7430-ATBD, issue 1.4.0, 2019.

Vandaele, A., Hermans, C., Simon, P., Carleer, M., Colin, R., Fally, S., Mérienne, M., Jenouvrier, A., and Coquart, B.: Measurements of the NO_2 absorption cross-section from 42000 cm^{-1} to 10000 cm^{-1} (238–1000 nm) at 220 K and 294 K, Journal of Quantitative Spectroscopy & Radiative Transfer, 59, 171–184, https://doi.org/10.1016/S0022-4073(97)00168-4, 1998.

Veefkind, J. P., Aben, I., McMullan, K., Forster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sensing of Environment, 120, 70–83, https://doi.org/10.1016/j.rse.2011.09.027, 2012.

Williams, J. E., Boersma, K. F., Le Sager, P., and Verstraeten, W. W.: The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation, Geoscientific Model Development, 10, 721–750, https://doi.org/10.5194/gmd-10-721-2017, 2017.

Wu, C. and Yu, J. Z.: Evaluation of Linear Regression Techniques for Atmospheric Applications: The Importance of Appropriate Weighting, Atmospheric Measurement Techniques, 11, 1233–1250, https://doi.org/10.5194/amt-11-1233-2018, 2018.

York, D., Evensen, N. M., Martínez, M. L., and Delgado, J. D. B.: Unified Equations for the Slope, Intercept, and Standard Errors of the Best Straight Line, American Journal of Physics, 72, 367–375, https://doi.org/10.1119/1.1632486, 2004.

Zhao, X., Griffin, D., Fioletov, V., McLinden, C., Davies, J., Ogyu, A., Lee, S. C., Lupu, A., Moran, M. D., Cede, A., Tiefengraber, M., and Müller, M.: Retrieval of total column and surface NO_2 from Pandora zenith-sky measurements, Atmospheric Chemistry and Physics, 19, 10 619–10 642, https://doi.org/10.5194/acp-19-10619-2019, 2019.