We thank the reviewer for the careful reading and the critical thinking, especially about the tropospheric AMF calculation and the MAXDOAS validation. Our response to the comments and detailed changes made to the manuscript are described below in black and blue, respectively, including the reviewer's text in red.

Major comments

1) The paper includes updates on surface albedo (from TOMS-based to GOME2-based) and NO_2 profile (from monthly climatology to daily varying). The importance of these updates are well known (e.g., see OMNO2, DOMINO, QA4ECV and POMINO for OMI), thus the significance of the finding here should be presented in a way that it confirms (or contrasts against) the findings of previous studies, aided with sufficient relevant citations in Sect. 1 and 6.3.

We have added more references in Sect. 1 to recent studies on surface albedo, NO_2 profiles, cloud and aerosol for OMI in page 3 line 16:

The AMFs are determined with a radiative transfer model (RTM) and stored in a look-up table (LUT) requiring ancillary information such as surface albedo, vertical shape of the a priori NO_2 profile, clouds and aerosols. Improvements in the RTM and LUT interpolation scheme, the ancillary parameters, and the cloud and aerosol correction approach have been reported for OMI instrument (e.g., Boersma et al., 2011; Lorente et al., 2017; Vasilkov et al., 2017; Krotkov et al., 2017; Veefkind et al., 2016; Lin et al., 2014; Castellanos et al., 2015; Laughner et al., 2018), which in principle are beneficial for similar satellite instruments like GOME-2.

Most of the above surface albedo studies use the MODIS BRDF product to improve the AMF calculation. This can however introduce additional biases due to the use of different instruments. Therefore, the significance of the finding here is related to the use of the latest improved version of the GOME-2 LER climatology that is consistent with the NO_2 retrieval from the same instrument, as introduced in Sect. 6.2.

We have added more references in Sect. 6.3 focusing on the updates of the a priori NO_2 profile, mainly the use of higher spatial resolution profiles and the use of daily profiles instead of monthly profiles, in page 20 line 9:

Increasing the spatial and/or temporal resolution of the a priori profiles have shown to produce a more accurate NO_2 retrieval (e.g., Russell et al., 2011; Heckel et al., 2011; McLinden et al., 2014; Nüß et al., 2006; Laughner et al., 2016).

We have related our findings about the use of daily vs. monthly profiles, which partly explains the large difference between the IMAGES used in GDP 4.8 and the

TM5-MP in GDP 4.9, to previous works in page 22 line 3:

In agreement with Nüß et al. (2006) and Laughner et al. (2016), the use of monthly profiles changes the tropospheric NO_2 columns by up to 3×10^{15} molec/cm^2 depending on the wind speed and wind direction, in particular for regions affected by transport (not shown).

Also, since the surface albedo climatology is used, what would be the implication of ignoring interannual variability and trends of albedo (as shown in many land cover change studies) on the AMF?

As analyzed by Sütterlin et al. (2016) for Europe using the AVHRR/BRDF albedo product for the year 1990-2014, the land surface albedo (in the visible wavelength) decreases by ∼0.004 per decade, probably driven by the change of land cover, vegetation, snow or ice. The interannual variability of the land surface albedo is generally smaller than 0.01 for snow-free vegetation cover but possibly larger than 0.06 for regions affected by snow. For our NO_2 retrieval, ignoring the decreasing trend of surface albedo leaves a small impact on the tropospheric NO_2 columns by up to 2×10^{14} molec/cm^2 (absolute) and 3% (relative). However, ignoring the interannual variability can introduce large errors in the AMF calculation for the varying snow and ice situation. Possible corrections for the surface albedo climatology include the use of external information about the actual snow and ice conditions, e.g., from Near-real-time Ice and Snow Extent (NISE) dataset. It is worth noting that reducing the error introduced by the interannual variability and trends was also a main motivation of using more recent observations for the LER climatology from GOME-2.

2) For tropospheric AMF, the most important sources of errors come from cloud retrievals and aerosols (e.g., Lorente et al., 2017 and references therein). However, the description of cloud retrieval and (especially) how it is affected by other ancillary parameters (e.g., from TOMS to GOME2 surface albedo) is not clear (e.g., in Sect. 6.4).

Lorente et al. (2017) estimated that the tropospheric AMF error is substantial related to the cloud correction approach (IPA vs. clear-sky AMF) by 5-40%, the aerosol correction approach (implicit vs. explicit) by ∼50%, and the choice of surface albedo, surface height, a priori NO_2 profiles, and cloud parameters. Boersma et al. (2018) concluded that the largest cloud parameters-related uncertainty is introduced by the surface albedo-cloud fraction error correlation by 10-20% using OMCLDO2 cloud parameters. For the OCRA/ROCINN cloud algorithm, the surface albedo-cloud fraction error correlation is relatively small. Unlike cloud algorithm that retrieves the cloud parameters from the TOA reflectance in the O_2-O_2 absorption band (OMCLDO2) or O_2-A absorption band (FRESCO), OCRA calculates the cloud fraction by separating a spectral scene into cloudy contribution and cloud-free background

using spectral information from the UV-VIS-NIR band. The surface albedo is not directly needed as an input in OCRA's cloud fraction retrieval, but it will affect the measured clear-sky TOA reflectances and thus the cloud-free map construction, which mainly influences the cases with rapidly varying surface conditions (e.g. fresh snow or melting) that happen on faster timescales than the monthly resolution of the clear-sky maps. In order to minimize this effect, OCRA linearly interpolates between two monthly maps to "daily" values. Therefore, the surface albedo leaves a relatively smaller impact on our cloud fraction retrieval, compared to other cloud algorithms like OMCLDO2 or FRESCO that require a surface albedo climatology. We have added the discussion about the cloud-related uncertainty in a new section describing the combined effect of algorithm changes and potential uncertainties (the new section is added as a response to comment 4):

The largest cloud-related uncertainty in NO_2 retrieval is introduced by the surface albedo-cloud fraction error correlation, as analysed by Boersma et al. (2018) for OMI using OMCLDO2 cloud product, which requires a surface albedo climatology as input in the cloud fraction retrieval. But this uncertainty is likely smaller for OCRA/ROCINN cloud algorithms, since the surface albedo is treated differently in OCRA's cloud fraction calculation. Retrieved by separating a spectral scene into cloudy contribution and cloud-free background, the cloud fraction from OCRA is affected by surface albedo through the cloud-free map construction with a larger impact over bright surfaces like snow or ice cover, in particular during snowfall (higher background) or melting (lower background), which has been corrected by interpolating towards a daily value between two monthly cloud-free map in OCRA (Lutz et al., 2016).

Additionally, we have removed Sect. 6.4 and moved the description of cloud parameters to the general introduction of GDP 4.8 in Sect. 3, since the cloud products are not changed between GDP 4.8 and 4.9.

3) The crucial role of aerosol treatment (in polluted cases) is only vaguely described, and should thus be better analyzed. If explicitly representing aerosols in the algorithm is not possible, a better description of implied errors should be presented, aided with relevant citations.

We have added more references and discussions about aerosol correction in the new section introducing the potential uncertainties (the new section is added as a response to comment 4):

The uncertainty introduced by aerosol in GDP 4.9 is ∼50% for high aerosol loading, in agreement with Lorente et al. (2017). With direct impact on NO_2 AMF calculation and indirect impact via cloud parameters retrieval, the aerosol effect has been considered for OMI implicitly through the cloud correction (Boersma et al.,

2004, 2011) or explicitly with additional aerosol information for regional studies (Lin et al., 2014, 2015; Kuhlmann et al., 2015; Castellanos et al., 2015; Chimot et al., 2018), leading to an increase or decrease of NO_2 AMF by up to 40% depending on NO_2 distribution and aerosol properties and distribution. Since aerosol is highly variable in space and time due to the dependency on emission sources, transports, and atmospheric processes (Holben et al., 1991), explicit aerosol correction will be applied in our AMF calculation when reliable observations or model outputs of aerosol optical properties and vertical distributions are available.

We agree with the importance of aerosol treatment in AMF calculation, and we will investigate the explicit correction for the aerosol effect in a following paper, with a case study over China using ground-based measurements from AERONET, Lidar, and MAXDOAS.

4) A global map of tropospheric NO_2 VCD (e.g., monthly climatology, not just one specific month) and how these algorithm updates/uncertainties affect the VCD should be given and discussed.

We have added an additional section named "6.4 Examples of GOME-2 tropospheric NO_2" to address the combined algorithm changes and uncertainties, with monthly climatology for Feb. and Aug. as examples:

Figure I shows the tropospheric NO_2 columns from the improved GDP 4.9 algorithm for February and August averaged for the year 2007-2016. Figure II shows the difference in tropospheric NO_2 columns from the GDP 4.9 and GDP 4.8 product. The tropospheric NO_2 columns increase globally by $\sim 1 \times 10^{14}$ molec/cm^2 due to the improved DOAS slant column fitting and increase further by $\sim 3 \times 10^{14}$ molec/cm^2 around moderately polluted regions beneficial from the use of new stratosphere-troposphere separation algorithm STREAM. A stronger change by more than 1×10^{15} molec/cm^2 is found mainly over polluted continents, as a result of the improvements to the AMF calculation, primarily the surface albedo (which also affects the snow or ice area, e.g., southern Canada and northeastern Europe) and/or the a priori NO_2 profiles (which also affects the polluted ocean, e.g., shipping lanes in southeastern Asia).

Over central northern Europe, the tropospheric NO_2 columns are reduced by $\sim 1 \times 10^{15}$ molec/cm^2 for GDP 4.9 in winter and $\sim 3 \times 10^{14}$ molec/cm^2 in summer. A larger number of negative values in GDP 4.8, possibly related to the overestimated stratospheric NO_2 around polar vortex areas, is largely corrected in GDP 4.9 by improving the stratosphere-troposphere separation algorithm. Over eastern China and eastern US, the seasonal variation is consistent between GDP 4.8 and 4.9, with reduced values in winter (by more than 1×10^{15} molec/cm^2) and enlarged values in summer (by more than 1×10^{15} molec/cm^2 for eastern China and 5×10^{14} molec/cm^2 for eastern US) for GDP 4.9 due to the combined impact of the algorithm changes, mainly the AMF calculation. Over India and its surrounding areas, a systematic

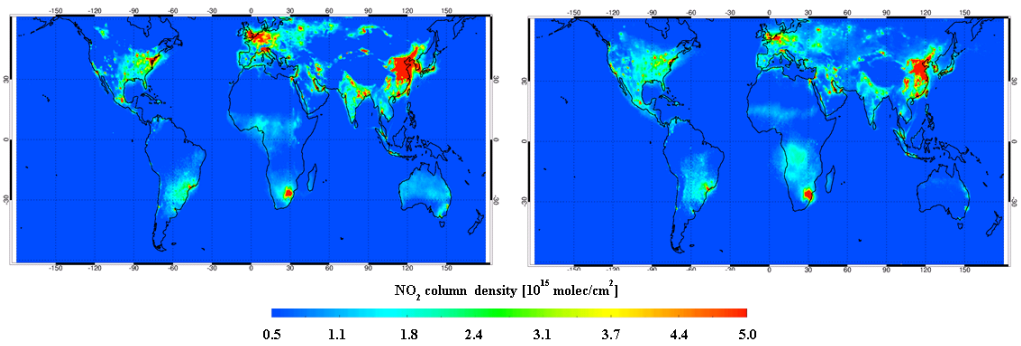

Figure I: Monthly average tropospheric NO_2 columns from GDP 4.9 for clear-sky conditions (cloud radiance fraction smaller than 0.5), measured by GOME-2A in February (left) and August (right) 2007-2016.

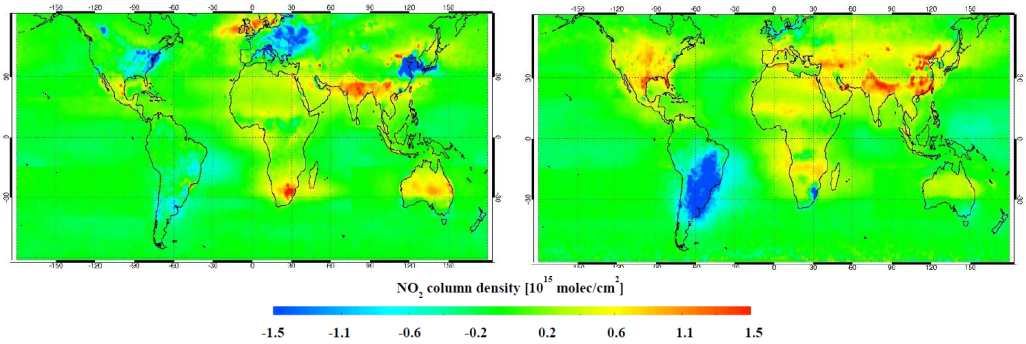

Figure II: Difference in tropospheric NO_2 columns from GDP 4.9 and GDP 4.8 for clear-sky conditions (cloud radiance fraction smaller than 0.5) in February (left) and August (right) 2007-2016 for GOME-2A.

increase in tropospheric NO_2 columns by $\sim 7 \times 10^{14}$ molec/cm^2 for GDP 4.9 benefits from the use of STREAM.

The uncertainty in our GDP 4.9 NO_2 slant columns is 4.4×10^{14} molec/cm^2, calculated from the average slant column error using a statistical method described in Sect. 4.5. The uncertainty in the GOME-2 stratospheric columns is $\sim 4-5 \times 10^{14}$ molec/cm^2 for polluted conditions based on the daily synthetic GOME-2 data and $\sim 1-2 \times 10^{14}$ molec/cm^2 for monthly averages. The uncertainty in the GDP 4.9 AMF calculation is likely reduced, considering the improved surface albedo climatology and a priori NO_2 profiles, which are the main causes of AMF structural uncertainty (Lorente et al., 2017). In addition, the AMF uncertainty is substantially driven by the cloud parameters and aerosol correction...

...To conclude, the uncertainty in the AMF calculation is estimated to be in the 10-45% range for polluted conditions, leading to a total uncertainty in the tropo-

spheric NO_2 columns likely in the range of 30−70%.

At its current form, the comparison with MAX-DOAS data does not tell much information regarding the causes of satellite errors (especially slope and bias), thus providing little new knowledge beyond previous findings.

Since the main goal of this work is the algorithm improvement rather than a validation study, we intended to present a new algorithm and show how it is an improvement wrt the previous version. We agree with the need of a better description, in particular in terms of the slope and bias, leading to the update of validation section below.

The discussion on sampling difference and retrieval algorithm should be aided with quantitative tests (of the algorithm assumptions/parameters, within the context of this study), particularly on how the data sampling (50 km is a relatively loose criterion; the temporal matching is not clear), cloud treatment and aerosol treatment each affects the comparison.

The choice of 50 km indeed seems loose at first glance, but considering the coarse resolution of GOME-2 pixels (80×40/40×40 km^2), it is not extreme. As explained above, the main goal of this section is not to do the best validation possible (this criterion has been tested in the past by e.g. Wang et al. (2017); Pinardi et al. (2015) and is included in a validation paper in preparation (Pinardi et al., 2018/2019)) but to compare the validation results for the 2 GDP versions for a fixed set of settings.

We have added the description of temporal matching in page 25 line 3:

The original ground-based MAXDOAS data usually retrieves NO_2 columns all day long every 20 to 30 minutes, and these values are linearly interpolated to the GOME-2 overpass time (9:30 local time), if original data exist within +/-1 hours.

The cloud treatment and aerosol treatment are the same between GDP 4.8 and 4.9 and thus leave almost no different impact during the comparison (Sect. 6.4 has been removed to avoid confusion, see response to the major comment 2).

Also, the main text only emphasizes the comparison at Xianghe (Figs. 14-16), which gives best consistency. However, the SI shows much poorer comparisons (in terms of bias and slope) at other sites. A better description in the main text of these comparisons (e.g., by better reference to SI, or moving some figures from SI to the main text) is needed to accurately present the overall quality of the satellite product.

We agree with the need to emphasize the results from other stations and the discussions in terms of bias and slope. Therefore, we have added more discussions of

the figures in the supplement and have rewritten page 26-28 with better comments at the other sites with a more quantitative explanation of bias and slope. We have updated page 26 line 1-page 27 line 3:

The daily differences are also reported through the histogram panel, where the reduction in the spread of the daily comparison points is clearly visible for GDP 4.9. The reduction of the bias, which is smaller and more stable in time, is seen in the absolute and relative monthly mean bias time-series. Three years show a standard deviation of the monthly biases larger for GDP 4.9 than for GDP 4.8 ($\pm$12% instead of $\pm$8% in 2010, $\pm$12% instead of $\pm$8% in 2013, and $\pm$41% instead of $\pm$27% in 2014) but with a strongly reduced mean bias (-4% instead of -20%, -8% instead of -34%, and -1% instead of -44%).

Similar figures as Fig. 14 and 16 for all the stations are gathered in Fig. S1 to S4 in the supplement, and all the statistics are summarized in Table 4 and Table 5 for GOME-2A and GOME-2B, respectively. Fig. S1 and S2 present the time-series and scatter plots for GDP 4.9, while Fig. S3 and S4 present the differences for both GDP 4.9 and GDP 4.8 comparisons. As discussed in Pinardi et al. (2015), for background stations (here Bujumbura, Reunion and OHP), the mean bias is considered as the best indicator of the validation results, due to the relatively small variability in the measured NO_2. In urban (Beijing and Uccle) and suburban (Xianghe) situations, the NO_2 variability is large enough and in this case, the correlation coefficient is a good indication of the linearity or coherence of the satellite and ground-based dataset, although larger difference in term of slope (closer to 0.5 than to 1 for urban cases) and of mean bias can be expected because satellite measurements (and especially GOME-2 80×40/40×40 km^2 pixels) smooth out the local NO_2 hot spots. This can be seen e.g. in the cases of Beijing and Xianghe for GOME-2A (see Fig. S1a and Fig. 14, respectively), where very high correlation (R=0.94 and 0.91, respectively) are obtained from GDP 4.9, showing the very consistent behavior of both datasets for small and large NO_2 columns, while their slopes (S=0.4 and 0.72, respectively) show almost a factor 2 of difference, with a smaller slope in the Beijing case, where the MAXDOAS instrument is in the city center and thus much more subject to local emission smeared out by the GOME-2 large pixel. This last effect is also seen through the biases values (RD=-47% and -5.8%, respectively) that are strongly reduced when moving the MAXDOAS outside the city in a suburban location like Xianghe. Slope of 0.47 (similar to the 0.4 of Beijing) is also obtained in Uccle, another urban site, where the MAXDOAS is affected by local emissions. In remote cases such as OHP, Bujumbura or Reunion Island, as discussed above, the variation of the NO_2 columns are small and the statistical analysis on the regression are not very representative of the situation, with a cloud of points giving small slopes and low correlation coefficients (see e.g. Fig. S1b to d and Table 4 for GOME-2A). In those cases, GOME-2 is lower than the ground-based, with sometime almost no seasonal variation, e.g. Bujumbura and Reunion, and in other cases, like OHP, some of the daily peaks are captured by GOME-2 (as days in the winter of 2014 and 2015), and the seasonal patterns

and the orders of magnitudes of both datasets are similar. In these cases, it is best to look at the absolute biases (as relative biases are large due to the division with small ground-based columns), as presented in e.g. Fig. S3b to d and Table 4. Mean absolute differences for GDP 4.9 are about -3.6 $\times$ 10^{15} molec/cm^2 for Bujumbura, -8.5 $\times$ 10^{14} molec/cm^2 for OHP, and -1.5 $\times$ 10^{15} molec/cm^2 for Reunion, which are all smaller than their respective GDP 4.8 values. The daily differences presented in the histograms of those figures also show reduced spread of GDP 4.9 comparisons when superposed to the GDP 4.8 results. Similar differences are also found for GOME-2B.

To conclude, although the Xianghe case presented in Fig. 14 to 16 is the best case (due to its suburban location and its long time-series), a better seasonal agreement between GDP 4.9 and MAXDOAS data is found for urban and suburban cases like Beijing, Uccle, and Xianghe, compared to results with GDP 4.8. In remote locations such as OHP, which is occasionally influenced by polluted air masses transported from neighboring cities, the comparison is also meaningful (e.g. with a mean bias reduced from -45% for GDP 4.8 to -25% for GDP 4.9 for GOME-2A), while cases such as Bujumbura and Reunion are quite challenging for satellite validation, with specific local conditions (Bujumbura is in a valley on the side of the Tanganyika lake, while the MAXDOAS at Reunion is in St-Denis, on the coast of the 65 km long and 50 km wide island in the Indian ocean, containing a mountain massif with summits above 2740 m asl). In both cases the MAXDOAS instrument is located in small cities surrounded by specific orography, difficult for satellite retrievals and challenging for validation. The absolute and relative differences show, however, a clear improvement for all the stations, when comparing to GDP 4.8 results for both daily and monthly mean biases. The daily biases and spreads are all reduced. To summarize, the impact of the improvement of the algorithm (as seen in Table 4 and 5 and in Fig. S3 and S4) leads to a decrease of the relative differences in urban conditions such as in Beijing or Uccle from [-52,-60]% for GDP 4.8 to [-43,-47]% for GDP 4.9 for GOME-2A and from -54% to -40% for GOME-2B. In suburban conditions such as in Xianghe, the differences go from -30% to -6% for GOME-2A and from -26% to -2% for GOME-2B. In remote (difficult) cases such as in Bujumbura or Reunion, the differences go from [-89,-90]% to [-64,-76]% for GOME-2A and from [-86,-87]% to [-47,-74]% for GOME-2B, while in background case such as in OHP, the differences decrease from -45% to -25% for GOME-2A and from -42% to -17% for GOME-2B. The differences in numbers for GOME-2A and GOME-2B are due to the different time-series length of both comparisons (e.g. March 2010-November 2016 for GOME-2A and December 2012-November 2016 for GOME-2B in Xianghe), the different sampling of the atmosphere by GOME-2A and GOME-2B (slight time-delay between both overpasses and reduced swath pixels for GOME-2A since July 2013), and the impact of the decreasing quality of the satellite in time, i.e. the GOME-2A degradation (Dikty et al., 2011; Munro et al., 2016). This lead, e.g. for Xianghe, to -2% bias and 0.49 slope for GOME-2B compared to -6% and 0.72 for GOME-2A for GDP 4.9.

These comparisons results aim at showing how the final GDP 4.9 product is im-

proved compared to its predecessor, and not to summarize the improvements of each of the changes discussed in previous sections. In addition, the specific validation method could be improved or at least better characterized (including results uncertainties), by e.g. changing the colocation method (averaging the MAXDOAS within an hour of the satellite overpass or selecting the closest satellite pixel, or only considering the pixels containing the station, etc.), but this is out of the scope of the present manuscript that wants to compare to standard validation results performed routinely on GDP 4.8 (and publicly available on http://cdop.aeronomie.be/validation/valid-results).

We have updated page 28 line 15-21:

When the average kernels are used to remove the contribution of a priori NO_2 profile shape, as seen in Table 4 and 5 and in Fig. S5 and S6, the relative differences in urban conditions such as in Beijing or Uccle decrease from [-52,-57]% for GDP 4.8 to [-34,-37]% for GDP 4.9 for GOME-2A and from -56% to -29% for GOME-2B. The differences go from -32% to -13% for GOME-2A and from -27% to -11% for GOME-2B for suburban conditions such as in Xianghe and go from -77% to -31% for GOME-2A and from -64% to -7% for GOME-2B for remote conditions such as in Reunion. The results...

...near the surface, the gradient-smoothing effect, and the aerosol shielding effect. These effects are often inherent to the different measurements types or the specific conditions of the validation sites (as seen for the different results for Beijing and Xianghe sites in this manuscript), but also to the remaining impact of structural uncertainties (Boersma et al., 2016), such as the impact of the choices of the a priori NO_2 profiles and/or the albedo database assumed for the satellite AMF calculations (see Sect. 6). Lorente et al. (2017) estimated e.g. the AMF structural uncertainty to be on average 42% over polluted regions and 31% over unpolluted regions, mostly driven by substantial differences in the a priori trace gas profiles, surface albedo and cloud parameters used to represent the state of the atmosphere.

Specific comments

Title: change "total" to "slant". The term "total column" is misleading.

As defined by EUMETSAT AC SAF product navigator (https://acsaf.org/products/nto_no2.html), the name "total and tropospheric NO_2 column" is used in the previous retrieval algorithm descriptions (Valks et al., 2011, 2017) and therefore is preferable for this manuscript as a heritage algorithm.

P1, L18: Please give the full name of VLIDORT and clarify which version

Done.

P1, L20: "a large effect" is somehow ambiguous. Does it mean decrease or increase?

We have updated the sentence to:

A large effect (mainly enhancement in summer and reduction in winter) on the retrieved tropospheric NO_2 columns by more than 10% is found over polluted regions.

P2,L6 VOC is not necessary for aerosol formation from NOx.

We have updated to:

serve as a precursor of zone in the presence of volatile organic compounds (VOC) and of secondary aerosol through gas-to-particle conversion (Seinfeld et al., 1998).

The title and content of Sect. 3 should be clarified such that the reader understands the section is describing the old algorithm.

We have updated to:

Total and tropospheric NO_2 retrieval for GDP 4.8

P3,L16-17 the treatment of aerosol is crucial especially for polluted cases.

See response to major comment 1.

Table 1 check the formulas of intensity offset for QA4ECV

To make the comparison less confusing, we have removed the formulas from the table and classified the intensity offset in general as "constant" vs. "linear" in wavelength.

Figure 2, right panel discuss the discontinuity in 2015

Unfortunately we have not found any document that explicitly states a processor update, key-data change, or in-orbit operation during this period. One possible explanation would be the occurrence of a solar eclipse event on 20 March 2015, which might disturb the thermal environment of the instrument/platform.

Figure 6 the error appears to be stabilized after 2010. Please discuss.

We have added discussions in page 12 line 4:

...instrument degradation (Dikty et al., 2011; Munro et al., 2016) until the major throughput test in September 2009 (see Sect. 4.3.1) and stabilize afterwards.

P15,L10 "latitudinal" should be "longitudinal"

Done.

P19, the paragraph "Figure 11 illustrates..." what is the impact on cloud retrieval, and how would this affect the sampling (with criterion of CRF <0.5)?

See response to major comment 2 and 4.

P22, L8: All discussion in this section only shows the improvement of cloud retrieval itself. How about its influence on NO_2 retrieval?

See response to major comment 2.

P25,L3 test the choice of 50 km, which is relatively loose. Also, how is the temporal interpolation done?
P26, first paragraph need a better description of comparisons at other sites (see my major comment)
P26,L8-11 The explanation is qualitative. A more quantitative explanation of slope and bias is needed. For example, to what extent the slope and bias can be explained by sampling difference and algorithm limitations? See my major comment.

See response to major comment about validation.

P28,L9-10 the sentence is not clear.

We have removed page 28 line 7-10, as it will be discussed in the following paragraph.

P28,L17-21 again, the statement is too general.

See response to major comment about validation.

P29,L32 the uncertainty in tropospheric NO_2 AMF appears underestimated, especially given the large error in cloud and aerosol treatment (e.g., Lorente et al., 2017).

As described by Valks et al. (2011) for GDP 4.4, the estimated uncertainties in tropospheric NO_2 AMF is in the 15-50% range for polluted conditions with an average of ∼33%. Considering the values 12-42% suggested by Lorente et al. (2017)

and all the improvements in the AMF calculation in GDP 4.9, we have updated the improved uncertainties in AMF to the 10-45% range and thus the total uncertainties in tropospheric NO_2 columns to the 30-70% range.

P30, first paragraph the consistency at other sites is much poorer. This point should be presented here.

We have updated page 30 line 4-7:

Taking Xianghe station as an example, the GDP 4.9 dataset shows a similar seasonal variation in the tropospheric NO_2 columns as the MAXDOAS measurements with a relative difference of -5.8% (i.e. -2.7×10^{15} molec/cm^2 in absolute) and a correlation coefficient of 0.91 for GOME-2A, indicating a good agreement. The Xianghe site, by its suburban nature, is the best site for validation. At the other sites, mean biases ranges from [-47%; -16 $\times 10^{15}$ molec/cm^2] for Beijing, [-76%,-74%; -3.6 $\times 10^{15}$ molec/cm^2,-2.8 $\times 10^{15}$ molec/cm^2] for Bujumbura, [-25%,-17%; -0.9 $\times 10^{15}$ molec/cm^2,-0.5 $\times 10^{15}$ molec/cm^2] for OHP, [-64%,-47%; -1.5 $\times 10^{15}$ molec/cm^2,-0.8 $\times 10^{15}$ molec/cm^2] for Reunion, and [-43%,-40%; -5 $\times 10^{15}$ molec/cm^2,-4.2 $\times 10^{15}$ molec/cm^2] for Uccle. Reunion and Bujumbura are difficult sites for validation, due to their valley/mountain nature, while urban sites Beijing and Uccle show similar relative results. The smaller absolute bias is found at the rural OHP station.

References

Boersma, K., Eskes, H., and Brinksma, E.: Error analysis for tropospheric NO2 retrieval from space, J. Geophys. Res. Atmos., 109, 2004.

Boersma, K., Eskes, H., Dirksen, R., Veefkind, J., Stammes, P., Huijnen, V., Kleipool, Q., Sneep, M., Claas, J., Leitão, J., et al.: An improved tropospheric NO2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmos. Meas. Tech., 4, 1905, 2011.

Boersma, K., Vinken, G., and Eskes, H.: Representativeness errors in comparing chemistry transport and chemistry climate models with satellite UV–Vis tropospheric column retrievals, Geosci. Model Dev., 9, 875, 2016.

Boersma, K., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Zara, M., Peters, E., Van Roozendael, M., Wagner, T., Maasakkers, J., van der A, R., Nighttingale, J., De Rudder, A., Irie, H., Pinardi, G., Lambert, J.-C., and Compernolle: Improving algorithms and uncertainty estimates for satellite NO2 retrievals: Results from the Quality Assurance for Essential Climate Variables (QA4ECV) project, submitted, 2018.

Castellanos, P., Boersma, K., Torres, O., and De Haan, J.: OMI tropospheric NO2 air mass factors over South America: effects of biomass burning aerosols, Atmos. Meas. Tech., 8, 3831–3849, 2015.

Chimot, J., Veefkind, J. P., de Haan, J. F., Stammes, P., and Levelt, P. F.: Minimizing aerosol effects on the OMI tropospheric NO2 retrieval – An improved use of the 477 nm O2-O2 band and an estimation of the aerosol correction uncertainty, Atmospheric Measurement Techniques Discussions, 2018.

Dikty, S., Richter, A., Weber, M., Noël, S., Bovensmann, H., Wittrock, F., and Burrows, J.: GOME-2 on MetOp-A Support for Analysis of GOME-2 In-Orbit Degradation and Impacts on Level 2 Data Products–Final Report, Tech. rep., ITT 09/10000262, IUP University of Bremen, 2011.

Heckel, A., Kim, S.-W., Frost, G., Richter, A., Trainer, M., and Burrows, J.: Influence of low spatial resolution a priori data on tropospheric NO2 satellite retrievals, Atmos. Meas. Tech., 4, 1805, 2011.

Holben, B., Eck, T., and Fraser, R.: Temporal and spatial variability of aerosol optical depth in the Sahel region in relation to vegetation remote sensing, International Journal of Remote Sensing, 12, 1147–1163, 1991.

Krotkov, N. A., Lamsal, L. N., Celarier, E. A., Swartz, W. H., Marchenko, S. V., Bucsela, E. J., Chan, K. L., Wenig, M., and Zara, M.: The version 3 OMI NO2 standard product, Atmos. Meas. Tech., 10, 3133–3149, 2017.

Kuhlmann, G., Lam, Y., Cheung, H., Hartl, A., Fung, J. C. H., Chan, P., and Wenig, M. O.: Development of a custom OMI NO2 data product for evaluating biases in a regional chemistry transport model, Atmos. Chem. Phys., 15, 5627–5644, 2015.

Laughner, J. L., Zare, A., and Cohen, R. C.: Effects of daily meteorology on the interpretation of space-based remote sensing of NO2, Atmos. Chem. Phys., 16, 15 247–15 264, 2016.

Laughner, J. L., Zhu, Q., and Cohen, R. C.: The Berkeley High Resolution Tropospheric NO_2 Product, Earth System Science Data Discussions, 2018, 1–33, doi:10.5194/essd-2018-66, URL `https://www.earth-syst-sci-data-discuss.net/essd-2018-66/`, 2018.

Lin, J., Martin, R., Boersma, K., Sneep, M., Stammes, P., Spurr, R., Wang, P., Van Roozendael, M., Clémer, K., and Irie, H.: Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide, Atmos. Chem. Phys., 14, 1441–1461, 2014.

Lin, J., Liu, M., Xin, J., Boersma, K., Spurr, R., Martin, R., and Zhang, Q.: Influence of aerosols and surface reflectance on satellite NO2 retrieval: seasonal and spatial

characteristics and implications for NOx emission constraints, Atmos. Chem. Phys., 15, 11 217, 2015.

Lorente, A., Boersma, K. F., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., et al.: Structural uncertainty in air mass factor calculation for NO2 and HCHO satellite retrievals, Atmos. Meas. Tech., 10, 759, 2017.

Loyola, D., Thomas, W., Livschitz, Y., Ruppert, T., Albert, P., and Hollmann, R.: Cloud properties derived from GOME/ERS-2 backscatter data for trace gas retrieval, IEEE Trans. Geosci. Remote Sens., 45, 2747–2758, 2007.

Loyola, D., Koukouli, M., Valks, P., Balis, D., Hao, N., Van Roozendael, M., Spurr, R., Zimmer, W., Kiemle, S., Lerot, C., et al.: The GOME-2 total column ozone product: Retrieval algorithm and ground-based validation, J. Geophys. Res. Atmos., 116, 2011.

Loyola, D. G., García, S. G., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R. J., Pedergnana, M., Doicu, A., García, V. M., and Schüssler, O.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, Atmos. Meas. Tech., 11, 409, 2018.

Lutz, R., Loyola, D., Gimeno García, S., and Romahn, F.: OCRA radiometric cloud fractions for GOME-2 on MetOp-A/B, Atmos. Meas. Tech., 9, 2357–2379, 2016.

McLinden, C., Fioletov, V., Boersma, K., Kharol, S., Krotkov, N., Lamsal, L., Makar, P., Martin, R., Veefkind, J., and Yang, K.: Improved satellite retrievals of NO2 and SO2 over the Canadian oil sands and comparisons with surface measurements, Atmos. Meas. Tech., 14, 3637–3656, 2014.

Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., Huckle, R., Lacan, A., Grzegorski, M., Holdak, A., et al.: The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing–an overview, Atmos. Meas. Tech., 9, 1279–1301, 2016.

Nüß, H., Richter, A., Valks, P., and Burrows, J.: Improvement of the NO2 total column retrieval for GOME-2, O3M SAF Visiting Scientist Activity, Final Report, IUP University of Bremen, 2006.

Pinardi, G., Lambert, J.-C., Granville, J., Yu, H., De Smedt, I., van Roozendael, M., and Valks, P.: O3M-SAF validation report, Tech. rep., SAF/O3M/IASB/VR/NO2/TN-IASB-GOME2-O3MSAF-NO2-2015, Issue 1/1, 2015.

Russell, A., Perring, A., Valin, L., Bucsela, E., Browne, E., Wooldridge, P., and Cohen, R.: A high spatial resolution retrieval of NO2 column densities from OMI: method and evaluation, Atmos. Chem. Phys., 11, 8543–8554, 2011.

Seinfeld, J. H., Pandis, S., and Noone, K.: Atmospheric chemistry and physics: from air pollution to climate change, 1998.

Sütterlin, M., Stöckli, R., Schaaf, C., and Wunderle, S.: Albedo climatology for European land surfaces retrieved from AVHRR data (1990–2014) and its spatial and temporal analysis from green-up to vegetation senescence, Journal of Geophysical Research: Atmospheres, 121, 8156–8171, 2016.

Valks, P., Pinardi, G., Richter, A., Lambert, J.-C., Hao, N., Loyola, D., Van Roozendael, M., and Emmadi, S.: Operational total and tropospheric NO2 column retrieval for GOME-2, Atmos. Meas. Tech., 4, 1491, 2011.

Valks, P., Loyola, D., Hao, N., Hedelt, P., Slijkhuis, S., Grossi, M., Begoin, M., Gimeno Garcia, S., and Lutz, R.: Algorithm Theoretical Basis Document for GOME-2 Total Column Products of Ozone, NO2, BrO, SO2, H2O, HCHO and Cloud Properties (GDP 4.8 for AC SAF OTO and NTO), Tech. rep., SAF/AC/DLR/ATBD/01, Iss./Rev.: 3/A/2, 2017.

Vasilkov, A., Qin, W., Krotkov, N., Lamsal, L., Spurr, R., Haffner, D., Joiner, J., Eun-Su, Y., and Marchenko, S.: Accounting for the effects of surface BRDF on satellite cloud and trace-gas retrievals: a new approach based on geometry-dependent Lambertian equivalent reflectivity applied to OMI algorithms, Atmos. Meas. Tech., 10, 333, 2017.

Veefkind, J. P., De Haan, J. F., Sneep, M., and Levelt, P. F.: Improvements to the OMI O 2-O 2 operational cloud algorithm and comparisons with ground-based radar-lidar observations., Atmos. Meas. Tech., 9, 2016.

Wang, Y., Beirle, S., Lampel, J., Koukouli, M., De Smedt, I., Theys, N., Ang, L., Wu, D., Xie, P., Liu, C., et al.: Validation of OMI, GOME-2A and GOME-2B tropospheric NO2, SO2 and HCHO products using MAX-DOAS observations from 2011 to 2014 in Wuxi, China: investigation of the effects of priori profiles and aerosols on the satellite products, Atmos. Chem. Phys., 17, 5007, 2017.