Answer to RC1:

Reviewer comments are given in black and author answers are in blue. Changes in the revised manuscript are marked in red.

Following a request from Referee 2 the structure of the paper has been modified. Red is used for corrections/additions, while green is used for indicating pieces of text having been moved in the paper.

Summary

Validation of satellite NO2 measurements via aircraft or ground-station instruments is hampered by disparities in the spatial resolutions of the different measurement types. Strong localized pollution sources can be effectively sampled at a ground station but are diluted by smearing within the satellite footprint. The authors demonstrate a method to correct for this dilution, using long-term high-resolutionNO2datasets to give an estimate of the spatial variability within a satellite ground pixel. Related approaches, using proxies for NO_2 spatial distributions have been employed in the past for limited instruments over limited regions. However, this study is valuable, because it provides a more robust test of the correction algorithm with a large number of sites and combination of ground-based MAX-DOAS and direct-sun instruments. Results show significantly improved correlations with reduced bias between satellite (OMI and GOME-2A) and the validation data. As such, it makes an important contribution and can/should be employed in future validation studies. The paper is well-written, organized and referenced. In addition to presenting the dilution correction, it stands by itself as a good validation study. I have only a few comments and minor corrections. If addressed, I recommend acceptance in AMT.

Answer: *We thank the reviewer for his useful comments and suggestions. We answer to each point below.*

General comments

(1) The dilution factor is computed in a 50 km-radius circle centered on the measurement site. This approach is convenient, but the satellite FOVs are not circular and vary widely in size. It seems that systematic errors could be large and could be mitigated using the pixel-corner coordinates in the OMI and GOME-2A level-2 files. In the conclusions, the authors acknowledge this, but more discussion should be given earlier in the text. A comparison with a dilution correction based on physical pixel dimensions for perhaps one challenging case would be useful to show whether errors in the fixed-circle assumption are significant.

Answer: *Another dilution correction approach has been tested, to estimate an uncertainty for our dilution method. Starting from the same high resolution OMI QA4ECV grid used for the current dilution correction, a ratio between each grid cell and the cell containing the station has been calculated. Instead of calculating a radial correction for the pixel's center distance, a correction based on the weighted value of the intersection of the high resolution grid cells ratios and the polygon formed by the satellite pixels corners is calculated, and this value is used instead of Fdil(R) in eq (4).*

2 extreme cases and 2 normal cases are tested. For cases with strong over-estimation, the new method changes significantly the results, reducing the impact of the correction by about 8 to 65% (Beijing and UHMT). In the normal cases, the change is only of about 3 to 7% (Xianghe and Uccle). These changes are of about half the value of the current dilution correction for these stations. Considering the small number of extreme cases (4 over 39 cases), and the corresponding small number of comparison colocations, an uncertainty of 5% is estimated for the whole station ensemble, which is half the overall impact of the dilution (10 to 13%), as summarized in Table 6.

The outcome is discussed in Sect. 6.2 in P. 29, lines 36-39 (in addition to the existing discussion of P. 35) and in Sect.7, P.38, lines 14-15:

P. 29: An alternative dilution correction approach taking into account the geographical extent of the satellite pixel and its localisation in the NO2 field has been tested. In order to estimate an uncertainty on our correction method, we applied this modified scheme to two extreme urban cases (Beijng and UHMT), and two moderate cases (Xianghe and Uccle). Differences amounting to about half the value of the current dilution correction are obtained.

P. 38: the dilution correction improves the validation results for both sensors, by about 10 to 13% in total over the stations ensemble, with an overall uncertainty due to the method, estimated to about 5%.

(2) An error in the estimated stratospheric component of the satellite NO_2 is suggested as a reason for the non-zero y-intercepts in the scatter plots of Figure 9. Highly structured stratospheres –e.g. in the NASA OMI Standard Product or from assimilation (as in DOMINO)–may be locally more realistic on a given day than smoother stratospheres (e.g. from STREAM) but may also be prone to mean systematic biases that alias some tropospheric NO_2 into the stratosphere and vice versa. This is discussed in the STREAM and NASA v2 algorithm papers and elsewhere. Small stratospheric errors can be amplified by the AMFs.

a) An investigation of how stratospheric aliasing may affect validation is beyond the scope of this study, as stated. But a brief comment could be included, since the OMI and GOME-2A data used here are based on assimilated stratospheres. Future examination based on STREAM would be interesting.

Answer: *a comment on Compernolle et al. 2020 work, mentioning results with STREAM instead of a stratospheric assimilation approach for OMI, was already included in P. 31 (former page 28), lines 12-15. The following sentence is added at the end of the paragraph, according to the reviewer suggestion:*

P. 31: Investigation of the impact of the smoother STREAM stratosphere on the tropospheric validation results is out of the scope of this study, but would be interesting as the small stratospheric errors can be amplified by the AMFs.

b) Equation (2) states that satellite-derived stratospheric NO_2 is subtracted from the direct-sun measurements. If the same stratosphere has been subtracted from the satellite total columns, shouldn't any stratospheric errors at least partially cancel, leading to a ~0 intercept? Are there other factors that could cause the non-zero intercepts?

Answer: *The satellite's tropospheric VCD retrieval is not performed from the satellite total VCD columns to get the tropospheric VCD, but from the total SCD, following equation (1). An error on VCDstrato would therefore be normalized by the satellite AMFstrato/AMFtropo ratio.*

Systematic errors on the satellite slant column would lead to an additive offset in the comparisons. E.g., a wavelength calibration misfit, as shown for OMI, can have effects of NO_2 slant column by 0.85 × 10^{15} molec cm^{-2}, independent of latitude, solar zenith angle and NO_2 value (van Geffen et al. 2015). This additive positive SCD offset would be present in both satellite's stratospheric SCD and in total SCD, thus cancelling itself in the satellite's tropospheric VCDtropo calculation, while it would affect the direct sun tropospheric VCD through VCDstrato substraction in equation (2).
Another factor leading to a non-zero intercept is related to the linear regression used. When the comparison points do not strictly follow a linear relationship, i.e., situations where data tend to agree well for small and intermediate values, but show large discrepancies for large values, the regression is strongly influenced by the high columns. This situation leads to a shifting of the regression line to positive y-intercepts. This is why the filtering on percentile 75 was introduced.

c) Minor point: For the DS data, the slopes in Figure 9 show best agreement with GOME-2A for filtered, dilution-corrected data. Table 5 suggests no filtering gives better overall agreement. Is this again an effect related to the y-intercepts?

Answer: *It is unclear to us what in Table 5 suggests that "no filtering gives better overall agreement". Based on Table 5 we conclude, for the DS data, that:*
- the best correlation (0.91 and 0.83), slope (1.1 for both GOME-2 and OMI) and intercept is obtained for filtered data with dilution correction (last line),
- a smaller relative bias is observed for all the data with dilution correction (and no filtering), but with large positive intercepts (between 3.2 and 3.6).
The RMS parameter has been added to table 5, following comments of referee 2. This variable also tends to be slightly larger for direct-sun dilution corrected data, confirming the over-correction. A sentence has been added:

P. 31, line 5: In the case of direct sun data, however, we note that the dilution correction tends to over-correct satellite measurements (see also Fig. 9), also resulting in slightly larger RMS values for the dilution corrected cases.

Specific and minor comments

(1) Page 2, lines 19-20: "...Since the mid-1990s, NO2 has been measured from space..." *-->done*
(2) Page 2, line 23: "...afternoon have also been made by the OMI..." *-->done*
(3) Page 4, line 26: "...from slant (SCD) to vertical (VCD) column densities." *-->done*
(4) Page 4, line 37: "...2018). SCD structural uncertainties..." *-->done*
(5) Page 6, line 6: "...Satellite-to-satellite comparisons...have been performed..." *-->done*
(6) Page 6, line 21: "...crossing the Equator around 13:45 LT (in ascending node)." *-->done*
(7) Page 6, line 29: "...the GOME-2A product..." *-->done*
(8) Page 7, line 14: "...). For 18 cloud-free..." *-->done*
(9) Page 8, Table 1: The table would be easier to read if the two satellite instrument columns were better delineated (e.g. a vertical divider). GDP4.8 and Q4ECV v1.1 should be grouped with GOME-2A and clearly marked as the first and second column headings applied to the entire table below the instrument information box. Similarly, DOMINOv2.0 and QA4ECV v1.1, grouped with OMI, should be clearly marked as the third and fourth column headings throughout. *-->done*
(10) Page 14, line 33: Define DS since it is used later. "Direct-sun(DS) observations are routinely..." *-->done*

Technically “direct-sun” should be hyphenated throughout, but this may be at the discretion of AMT. *-->done*

(11) Page 15, line 14: “Those account for...” *-->done*

(12) Page 15, line 17: “...estimated using satellite data (SAT) (alone or within assimilation...” *-->done*

(13) Page 15, line 36: “...and only OMI pixels centered...” *-->done*

(14) Page 16, line 4: “Ground-based (GB) MAX-DOAS date were interpolated...” *-->done*

(15) Page 17, line 18: “...compared to early afternoon (13:30 hrs)...” Are LTs in this paragraph mean values for the stations (given 13:45 equator crossing)? *--> These numbers are the solar local time at equator crossing*

(16) Page19, lines 6 and 24: “...and GOME-2A overpass...”, “...and GOME-2A overpasses...” *-->done*

(17) Page 20, lines 3 and 24: “...their median difference at OMI and GOME-2A overpass are 5.7 and ...”, “...than for GOME-2A...” *-->done*

(18) Page 25, line 15: “...the outer extent of any 40 x 40 km2 GOME-2A pixels whose centers are within the 50 km radius.” *-->done*

(19) Page 27, lines 3, 4: “...GOME-2A...” *-->done*

(20) Page 27, line 18: “...scatter plots of GOME-2A and ground-based data...” *-->done*

(21) Page 30, line 11: “...and GOME-2A GDP...” *-->done*

(22) Page 31, lines 3-4: “...but should have relatively little systematic effect on regression slopes.” *-->done*

(23) Page 31, line 8: “...morning GOME-2A overpass...” *-->done*

(24) Page 32, Figure 11: Please define seasonal colors in the caption, or preferably as a legend on the figures. *-->done (in the new figure 12, following the reorganization of the paper suggested by reviewer 2)*

(25) Page 32, line 2:“...GOME-2A...” *-->done*

(26) Page 33, line 19: “...they found a complex spatial distribution...” *-->done*

(27) Page 35, line 4: “The number of comparison points for each case is shown in the corresponding color.” *-->done*

(28) Page 38, line 9, Figure 16 caption: Please explicitly define the colors *-->an explanation of the colors has been added in the figure caption, as follows:*

Red color is used for the dilution corrected data, while black is used for the previously presented products (OMI DOMINO and GOME-2A GDP) and grey is used for the QA4ECV products.