We thank the reviewer for the careful reading and the critical thinking, especially about the wavelength-dependency in 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. Validation

The validation using MAXDOAS suggests that GPD4.9 agrees better with this independent correlative measurements than GDP4.8, because the bias reductions (shown in Tables 4 and 5) achieved with the newer version GDP. But this overemphasizes the bias reduction. The equally important measure is the standard deviation, which shows little or no improvement. In fact, there are many cases (Tables 4 and 5) that show large standard deviation or lower correlation with MAXDOAS for the new GDP, indicating that the agreement becomes worse with the algorithm changes. One could argue that biases in satellite measurements may be easier to remove, often achieved by offset adjustments. Therefore lower biases likely do not say much about retrieval improvements. Keeping in mind that the coincidence of the ground-based MAXDOAS and GOME-2 may be limited, the agreements (as measured by standard deviation of difference or correlation coefficient) have its limitation as well. I recommend revise section 7, and add some discussions on the agreement based on standard deviation and correlation.

We agree with the need of emphasizing the standard deviation and correlation, leading to the revision of Sect. 7 page 26-28 with more discussions. As a response to the second reviewer, we have also added more discussions in terms of slope and bias as well as more descriptions at other stations besides Xianghe (which gives the best validation results). Please refer to the second response letter page 6.

2. Larger Fitting Window

While it is true that the noise level of slant NO_2 column is lower for DOAS retrieval from a larger fitting window in general, there is a downside as well. The key assumption of DOAS approach is that the AMF is (nearly) independent of wavelength. However as the fitting window becomes larger, the spectral dependence of tropospheric AMF becomes more prominent. For instance, the measurement sensitivity (called the box AMF ml, Eq.8) is 20% higher at 490 nm than at 425 nm at 1 km above the surface, for a spectral invariant surface reflectivity at 0.1. Closer to the surface at 0.1 km, ml is 30% higher at 490 than at 425 nm. Furthermore, surface reflectivity depends on wavelength as well, introducing additional spectral variation the tropospheric AMF. This spectral variation in AMF implies that the absorption signal would be 20% to 30% stronger at 490 nm than at 425 nm for the same amount tropospheric NO_2. Considering that many improvements described in the paper are on the order of a few percent, perhaps it is a good idea to discuss how the AMF spectral variation affects the retrieval accuracy. My rough estimate indicates that it may have up to 15% error for the larger fitting window when neglecting the spectral variation.

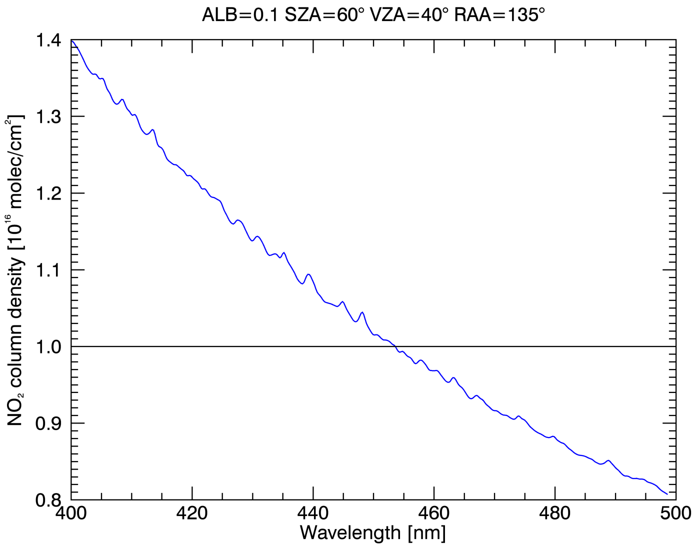

Figure I: The retrieved tropospheric NO_2 columns as a function of wavelength for a polluted boundary layer with 1×10^{16} molec/cm^2 (horizontal line), 1 km boundary layer height, surface albedo of 0.1, SZA of 60°, VZA of 40°, RAA of 135°, and without aerosol.

The different fitting window results in different sensitivities to boundary layer pollution, especially when the surface albedo varies with wavelength at the same time. To evaluate the error introduced by the wavelength-dependency of AMF to our NO_2 retrieval, we have implemented a sensitivity study using the LIDORT radiative transfer model. We simulated clear-sky TOA reflectances (and also tropospheric AMFs) between 400-500 nm with a 0.22 nm step for a typical polluted scenario with a "true" tropospheric NO_2 column 1×10^{16} molec/cm^2. Given the slant column of 2.2×10^{16} molec/cm^2 that was fitted using the 425-497 nm fitting window, the retrieved tropospheric NO_2 column as a function of wavelength (using the wavelength-dependent tropospheric AMFs) is shown in Fig. I. The tropospheric NO_2 column decreases (i.e. the tropospheric AMF increases) between 425-497 nm by 30.9% due to the stronger Rayleigh scattering, similar to the values suggested by the reviewer. However, the retrieval accuracy may not be affected that much, since the tropospheric NO_2 column calculated at the mid-point wavelength of fitting window (461 nm) is close to the true value (with a difference of 3.7%). Similar results are also obtained for other geometries and surface albedo. Since the discussion for surface albedo has been included in Sect. 6.2 with an impact on the tropospheric NO_2 column by up to 5%, we have extended the discussion for AMF calculation in page 18 line 10:

The box-AMFs m_l for each layer are calculated for the mid-point wavelength of fitting window, i.e., 461 nm in our NO_2 retrieval, which is representative of the window-average box-AMFs. Compared to the tropospheric AMFs at 440 nm (mid-point wavelength in GDP 4.8), the ones calculated at 461 nm are higher by up to 10% for polluted situations, due to the wavelength-dependency of Rayleigh scattering, in agreement with Boersma et al. (2018) (see Fig 7 therein). Note that the uncertainty related to the wavelength-dependency of the AMF is much smaller than the uncertainties introduced by surface albedo, a priori NO_2 profile, cloud and aerosol (see Sect. 6.4).

3. Intensity Offset
* Based on Eq. 1, the offset listed in table 1, should simply be 'a' for GDP4.8, and 'a + bλ' for GDP4.9, not ln(I+a) and ln(I+a+bλ), respectively.

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

* Please add in the paper a description of how the parameters 'a' and 'b' are determined.

The offset correction is modelled using a polynomial and fitted as a non-linear parameter in QDOAS. We have added the method for GDP 4.8 in page 4 line 20:

The intensity offset, which describes the additional contributions such as stray light in the spectrometer to the measured intensity, is modelled using a zero order polynomial with polynomial coefficient as fitting parameter.

We have added the updates for GDP 4.9 in page 7 line 27:

To correct for this drift, an intensity offset correction with a linear wavelength dependency (i.e., polynomial degree of 1) is applied for the large fitting window in this study.

The additive intensity offset looks similar to stray light contribution. It would be helpful add some plots to accompany the difference plots in Fig. 1. Specifically, include a map of LER to show the scene reflectivity, and maps of the parameters 'a' and 'b'. If the offset has something to with stray lights, 'a' would be high for low reflectivity scene, and vice versa, assuming that L1B data are not yet stray-light corrected. In any case, these additional plots would reveal information about L1B, and lend credibility of the intensity-offset correction.

Figure II shows the parameter 'a' on 05 Feb 2009 (instead of the given day in March in the manuscript so that the LER map in Fig. 10 can be referable). The intensity offset depends on the stray light in the instrument, but the dependency of 'a' (and also 'b', not shown) on LER seems not dominant. Given the combined impact of several factors on the intensity offset, this is not surprising. The dominating structures in 'a' are related to the ocean-land contrast (vibrational raman scattering), the orbital swath (instrument calibration issue related to the scanning mirror), and the cloud pattern (e.g., cloud albedo or fraction, possibly related to the stray light or incomplete removal of Ring effect). Since a stray light correction has been included during the GOME-2 level 0 to 1b processing, the residual stray light might have a relatively small impact, compared to the factors described above.

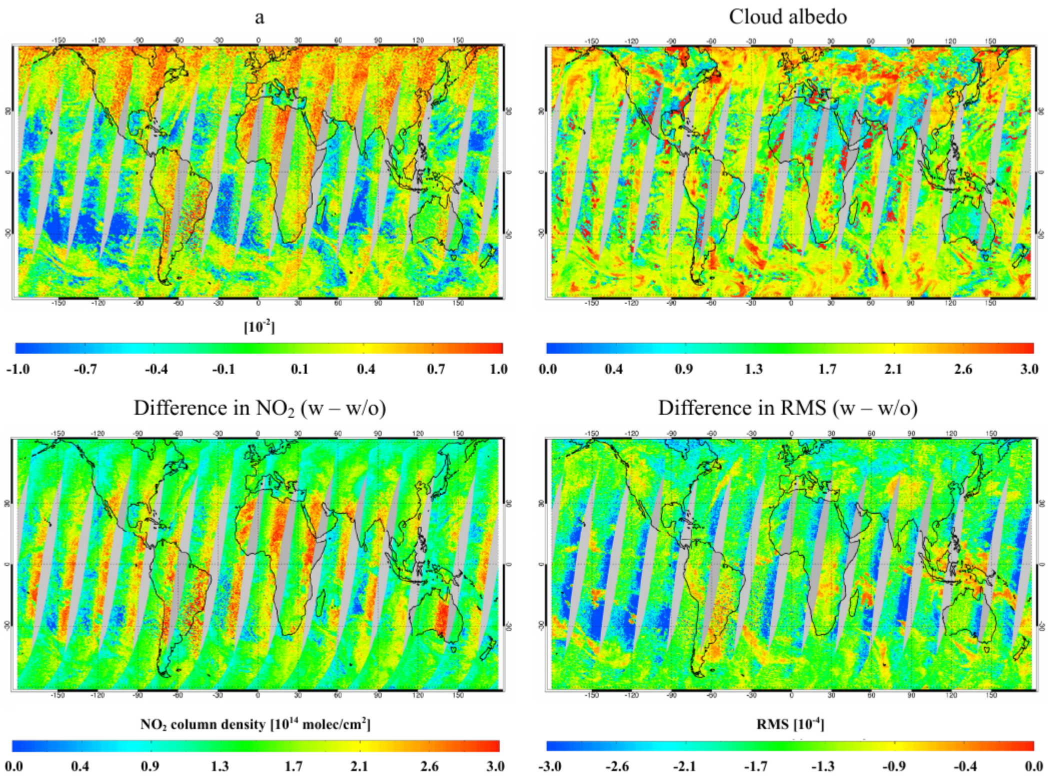

Figure II: The fitted (constant) polynomial coefficient for intensity offset correction (top left), the cloud albedo (top right), difference in NO_2 columns (slant columns scaled by geometric AMFs) (bottom left) and retrieval RMS (bottom right) estimated with and without intensity offset correction for GOME-2A on 5 February 2009.

4. Minor edits

* Page 2, line 9: "a strong growth of NO_2 since two decades has caused", 'since two decades' is not clear, revise please.

We have changed to:

a strong growth of NO_2 during the past two decades has caused

* Page 2, line 33: "measured GOME-2 (ir)radiances", should be measured GOME-2 radiances, or measured GOME-2 sun-normalized radiances.

We have changed to:

GOME-2 sun-normalized radiances

* Page 22, line 10: "and photon path", not specific or meaningful here. May be removed.

Done.

References

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.

Müller, J.-P., Kharbouche, S., Gobron, N., Scanlon, T., Govaerts, Y., Danne, O., Schultz, J., Lattanzio, A., Peters, E., De Smedt, I., Beirle, S., Lorente, A., Coheur, P., George, M., Wagner, T., Hilboll, A., Richter, A., Van Roozendael, M., and Boersma, K.: Recommendations (scientific) on best practices for retrievals for Land and Atmosphere ECVs, Tech. rep., Deliverable 4.2 - version 1.0, 2016.

Richter, A. and verification team, S.-P.: S5P/TROPOMI Science Verification Report, Tech. rep., S5P-IUP-L2-ScVR-RP issue 2.1, 2015.