We would like to thank reviewer 3 for their comments. In the following, we respond to each of the comments individually. In order to facilitate the tracking, we use the following color coding:

Color coding:

reviewer comment

our answer

proposed change in manuscript

General comments:

1/I think it would help the reader to include a map (that could be Figure no 1 of the manuscript) of the Mexico City area showing the locations of the different instruments (MAX-DOAS, AERONET, in-situ). Moreover, indicating the pointing directions of the MAX-DOAS instrument involved in this study could maybe also give some insights on the interpretation of the discrepancy between MAX-DOAS and in-situ NO2 surface concentration values. For instance, part of the underestimation of the in-situ values by the MAX-DOAS could be related to the fact that the MAX-DOAS instrument points towards a part of Mexico City which is less polluted than the location of the in-situ instrument.

In addition to the map, the location (latitude, longitude, altitude) of all the instruments should appear in the text.

The AERONET site and the in-situ measurement site are at the same position (give a nd take a few tens of meters) as the MAXDOAS station. The site for the ballon launch for the sounding is at the airport (a few kilometers to the north). However, since the position of the ballon quickly changes, we do not think that the actual launching position is very important. We will however state the coordinates in brackets below the first mentioning of the measurements. As for a map for the MAXDOAS instrument and its orientation, such a map is included in Arellano et al. 2016. We will refer to that in the manuscript specifically.

2/AERONET data are used as input in the retrieval but also as ancillary data for the sky conditions screening. Was there any attempt to compare the retrieved AODs with those from AERONET ? It can be a good check for the aerosol retrieval part of the profiling. Also related: it seems that the availability of AERONET observations has been used as a quality control (QC) flagging for the MAXDOAS retrievals. Was there any attempt to apply a QC flagging which is more specific to the MAX-DOAS retrievals, e.g. using parameters like DOF and the RMS of the differences between measured and calculated dSCDs ?

We first clarify which aeronet data is used in which way (a) and then answer the question about the filtering (b):

(a) We use the AERONET data of omega and g as input for the forward model (time interpolated). We do not attempt to retrieve those values. Further, the extrapolated (at the aerosol retrieval wavelength, and time interpolated) aod value is used as a-priori. As input for the NO2 retrieval, an interpolation between the retrieved aod and the nearest AERONET wavelength (time interpolated) was performed, see also answer to question “specific 2” below. In case of failure of the aerosol retrieval (non-convergence, or a bad fit in terms of the average absolute value of the difference in measured and simulated dscd in units of dscd error [sum(abs(DSCD_sim-DSCD_meas)/DSCD_meas)/number_of_elevation_angles] but no filtering on DOF because we designed the scaling of the Thikonov constrained in such a way to have a DOF of just above 1), an extrapolation of the two nearest (both to the long-wavelength side) AERONET values (time interpolated) was performed. Hence we do not see too much sense to compare to AERONET.

(b) In order to ensure a small forward model error it is important to have good estimations on the aerosol parameters g and omega. Without AERONET data available, we use an interpolation of the nearest available data in time. Hence the forward model error is expected to be smaller if AERONET data is available close in time. We also use the presence of AERONET data as a proxy for cloud free conditions. See also the answer to question 4a from reviewer 2.
Regarding the NO2 retrieval we do currently not use any filtering, not on RMS not on DOF. We looked at the distribution of DOF. We show a histogram in Fig. 1

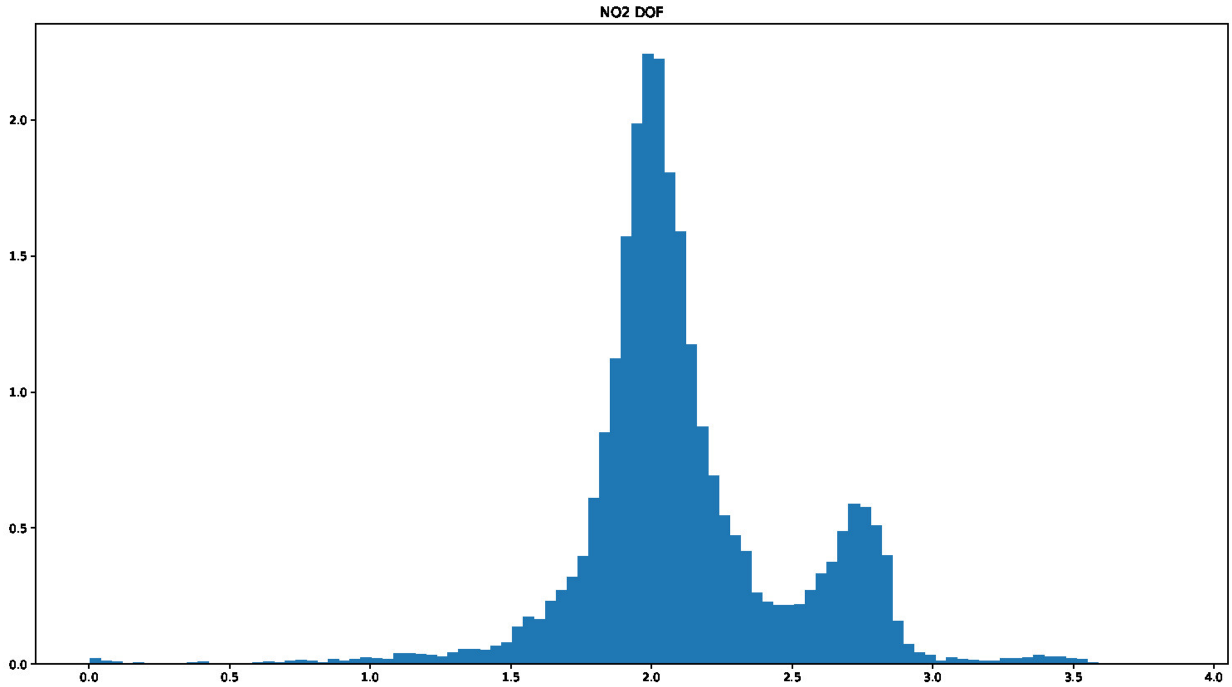

Fig.1 Distribution of DOF for NO2 retrievals.

If we were to include a filter on DOF, we would likely choose a limit of 1.5, hence the impact had been rather small.

Specific comments:

1/Page 4, beginning of Section 2.3: It would be good to list the exact elevation angle values of a typical scan.
We will include this in the last line on page 4, just before mentioning the likely change in measurement sequence.

2/According to Section 3, it seems that aerosol profile retrieval is done in the UV range and then retrieved profiles are used as input for the NO2 profile retrieval in the visible range. Has a correction been applied to the retrieved extinction/AOD for taking into account for the wavelength dependence of the AOD/extinction ? If not, then this approximation should be included as an additional error source in Table 1.
We perform a linear interpolation at the NO2 retrieval wavelength using the value retrieved value in the UV and the closest AERONET value. We plan to use an O4 retrieval window closer to our NO2 retrieval window in the future. The error arising from this would be one contribution to the estimated algorithm error which is already in the table (The error in NO2 from errors in the aerosol profile). The easiest way to get this error contribution would be perhaps to use PANDORA

instruments (calibrated-direct sun measurements) collocated to AERONET sites. But up to now there are not yet sufficient coincident measurements in Mexico City

3/Page 6, lines 8-11: Maybe you could add a couple of sentences about the performance of MMF in these profile comparison exercises. Please note that in the meantime, Friess et al. is now published in AMTD.
Thank you for the note that Friess et al. Is now published. This will be added to the references. The inclusion of the changes in MMF is moved to an appendix (see comment to question 2b from reviewer 2), a note on the time performance is included.

4/Page 8, line 7: Is it IO or HONO (cf page 6, end of Sect. 4) ?
It is HONO, we corrected this and also added the missing reference to the list of references.

5/Page 8, Sect. 4.2: You should add a paragraph on the SCIATRAN RTM, which has been also used in past MAX-DOAS profiling studies (see e.g. Friess et al., AMTD, 2018).
We will add SCIATRAN RTM in the list of examples for radiative transfer models used as forward models for profile retrieval with MAXDOAS.

6/Page 10, line 1: what type of interpolation is done for the pressure, temperature profiles ?
We use simple linear interpolation. We know that this can be improved upon for P. However we expect the effect to be rather small since the grid for the T and P profiles is of a similar resolution than the internal retrieval grid.

7/Page 10, line 14: a correlation length of 500m is used. Did you perform sensitivity tests on this parameter in order to estimate its impact on the retrieved profiles and on the level of agreement with in-situ measurements ?
The correlation length was only used for the error calculation, not for the retrieval. For the retrieval, no Sa matrix was constructed, but a Thikonov constrained used.

8/Page 19, Figure 6b: A priori profile should be also included in this Figure in order to see how far the retrieved profile differs from the a priori one.
We will include the a-priori profile in the plot. Also, as response to reviewer 1, we change one of the orange line for easier distinction. A proposed new Figure 6 (in the manuscript) is reproduced here as Fig. 2 (we will adjust the axis labels and tick labels to a more readable font size)

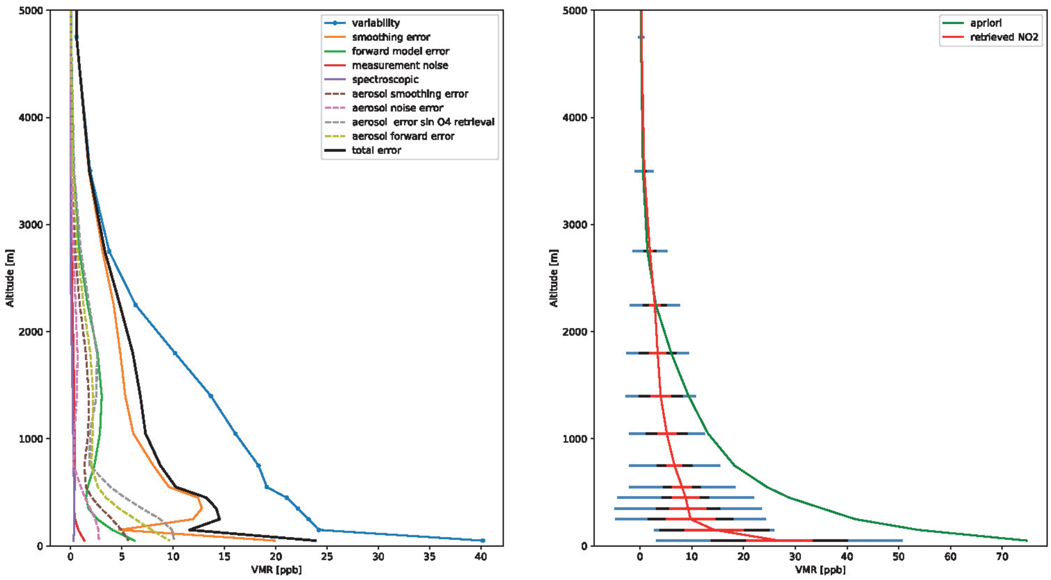

Fig. 2: as Figure 6 in the manuscript but including the a-priori (right) and a better distinction of the curves (left).

9/Page 26, lines 5-10: According to the authors, a possible reason for the underestimation of the in-situ surface concentration by the MAX-DOAS is the fact that the MAX-DOAS instrument has a maximum sensitivity around 1km and less sensitivity close to the ground. This feature is quite unusual since normally lowest elevation angles have a higher weight in the retrieval due to higher AMFs, and therefore MAX-DOAS measurements close to the ground. Could the authors elaborate on that ? Another possible explanation for the discrepancy is that, related to the horizontal extent and the pointing direction of the MAX-DOAS measurements, both instruments probe different air masses. I think this point should be also added in the discussion.

The averaging kernel, for a typical AK see Fig.4(a), shows that we expect an underestimation at the surface. It peaks at around 1km. The main reason for the lower sensitivity is the measurement angle distribution (too few low elevation angles) and the rather huge dscd errors at these low elevation angles, see Fig.2. AK=(K^T Se-1 K+Sa-1)-1 (K^T Se-1 K). While Sa^-1 is constant, a full covariance matrix from model averages, Se-1, the measurement error and K, the Jacobian, mainly dependent on the aerosol content, are variable. Therefore, the AK is only an example.

We can actually estimate the slope and underestimation theoretically using a typical Averaging Kernel using the variability of NO2 in the Mixing layer described by the Sa. However, since Sa is only an estimation, we tried two Strategies a) either using the Sa calculated from profiles of the model run and another taken from the Literature Wang et al., 2017. The first one is used as constraint in the OET-retrieval and described in the Manuscript. The latter uses a 100% variability of the a priori an the diagonal and a 500m exponential correlation length for off-axis elements as in Wang et al. (2017).

The Slope between a retrieved quantity, either the total column or the average of some layers, is calculated, respectively, by applying an operator g=(1 1 1 1 1 1 ....) on the profile in units of partial columns or g6=(1/6, 1/6 ,1/6, 1/6, 1/6, 1/6,0,0,0....) in VMR. "g6" is the operator which calculates the average of the lowest 6 layers. To get the in situ value we apply g1=(1 0 0 0 0...) on the profile.

The linear relation between retrieved values (e.g. averages of 6 layers) and in situ values depend on the correlation between all layers and is theoretically described by the following expression:

$< g_6\, AKV_{VMR} | SA_{VMR} |\, g_1 >$ which assumes that the profile variability is described by a normal distribution P(x)=1/sqrt(Pi Det(SA)) Exp(-(x-xa)^t SA-1 (x-xa)) (Rodgers 2000) of the Variability.

The theoretically calculated slope can be compared to the experimental obtained slope (Fig. 8) and so the Sa- matrix can be tested for, how plausible the estimation was.

In Fig. 3 here, we show the slopes (y-axis of Maxdoas v.s. insitu for the average of a different numbers of the lowest layers indicated by the x-axis (just as in Fig. 8 in the paper).

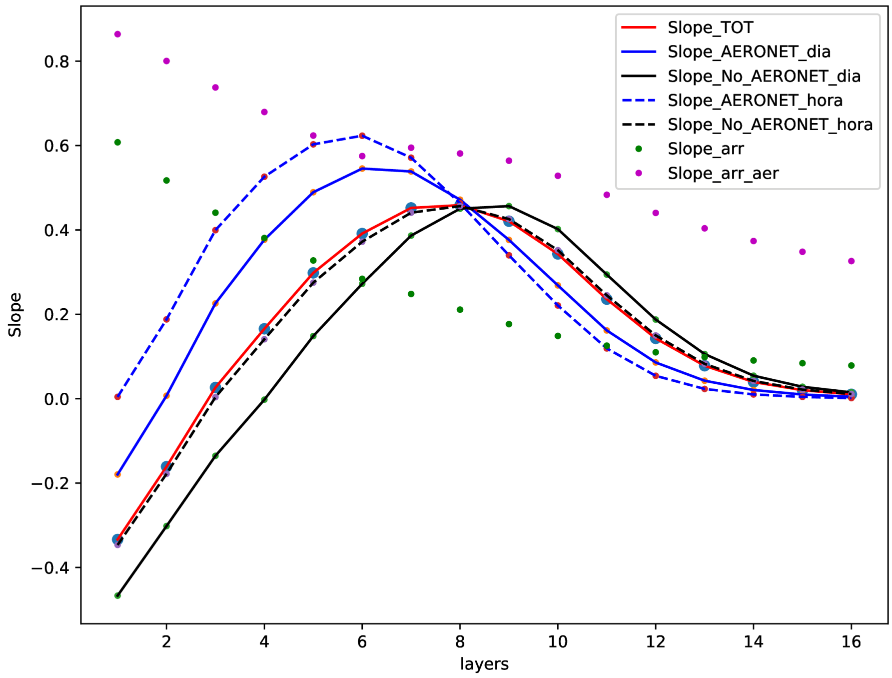

Fig3: As Figure 8 in the manuscript, just with a few more lines with respect to different filters related with the aeronet data and we added the two theoretical calculations shown by dots, green dots using the SA described in this work and the pink dots using the SA-Matrix described by Wang et al, 2017. Details see text.

The points are calculated theoretically either using the SA we constructed from the ensemble of modeled profiles (green points, Slope_arr) or taken from Wang et al., 2017 (pink points, Slope_arr_aer).

The graph explains even quantitatively the underestimation and no other arguments are needed. Just for the average of very few layers, there seems to be a discrepancy, which might indicate that the SA-Matrix do not describe the variability and correlation of the lowest level correctly.

We learned a lot by this exercise, but we would prefer not to complicate the manuscript too much and suggest just to add

"and variability Sa" behind the averaging kernel and maybe add as well the two calculated values

10/Acknowledgements: Depending on the conditions of use, the sources of ancillary data included in your study should be acknowledged here.

Thank you for pointing this out, we will include appropriate acknowledgements.

Technical corrections:
1/Page 1, line 1: '. . .to retrieve profiles. . .' -> '. . .to retrieve vertical profiles. . .'
corrected
2/Page 1, line 10: '. . .at the Universidad Nacional Autónoma de México (UNAM) campus.'
corrected
3/Page 1, line 20: '. . .The Multi-AXis Differential Optical Absorption Spectroscopy

(MAX-DOAS) technique. . ..'
corrected
4/Page 2, line 30: '. . .at the Universidad Nacional Autónoma de México (UNAM).'
corrected
5/Page 6, line 5: 'Sect.4.1' -> 'Sect. 4.1'. This typo should be corrected throughout the manuscript; similar corrections also needed for 'Fig.' and 'Eq.'.
Sect, Fig. And Eq. checked for space behind.
6/Page 6, line 9: 'CINDI2' -> 'CINDI-2'
corrected
7/Page 10, line 14: '. . .Eq. 4:'; same on Page 11, line 4
corrected
8/Page 11, line 20: 'extincion' -> 'extinction'
corrected
9/Page 16, legend of Fig. 4: (b,right)' -> '(b, right)'; should be also corrected for (a, left).
corrected
10/Page 19, legend of Fig. 6: 'total error'
corrected
11/Page 19, line 3: 'algoritm' -> 'algorithm'
corrected
12/Page 20, line 4: 'aprox.' -> 'approx.'
corrected
13/Page 21, line 11: 'Curretly' -> 'Currently'
corrected
14/Page 24, end of line 23: A reference could be added here.
We will add García-Franco et al. 2018