Comment on NO_2 vertical profiles and column densities from MAX-DOAS measurements in Mexico City

The authors describe a new MAX-DOAS profile retrieval code, the Mexican Maxdoas Fit (MMF), including an examination of the measurements, the algorithmic approach, and the error budget. The code is then applied to an 18 month dataset from Mexico City and compared to an in situ sensor. When averaged over the boundary layer MAX-DOAS measurements are found to capture diurnal and seasonal trends, but are systematically lower than the in situ measurements. The patterns in the mismatch can be understood to in part reflect meteorology.

At its most basic, the manuscript does two separate but related things: it describes MMF and it examines the 18 months of data in Mexico City using MMF. The fact that MMF has been updated since it was applied to the 18 month data set presents something of a challenge in presenting both the code and the data set clearly. Some restructuring of the content would better provide transparency and clarity to the reader. Further, some minor additions to the figures and text could make the message of the manuscript more coherent and relevant. I offer specific comments below.

Major comments:

1) Regarding the dSCD retrieval and Section 3:

a) The authors use a zenith measurement prior to the scan to analyze the scan. If the upper atmospheric contribution to the dSCDs changes during the scan this can lead to signal in the measurements which is from the upper atmosphere being falsely attributed to lower altitudes especially in low eastward elevation angles at the end of the scan. The effect would be expected to lead to lower VCDs in the morning and higher VCDs in the evening, especially in winter. Because of the short 7 minute scan time this effect would likely only be significant at twilight.

Does the instrument acquire data at twilight which are included in the analysis? Can the authors bound the impact of such an effect and compare it to the magnitude of their error budget?

b) For the fitting setting in the retrieval the authors use older cross-sections where newer cross-sections for the same gases are increasingly standard in the community e.g. (Damadeo et al., 2013; Peters et al., 2017). For O_3 they use (Burrows et al., 1999) rather than (Bogumil et al., 2003) or (Serdyuchenko et al., 2014) and for O_4 they use (Hermans et al., 1999) rather than (Thalman & Volkamer, 2013). Was there any particular reason for these choices? Is it based on Orphal, 2002 cited later?

2) Regarding the profile retrieval method description and Section 4:

a) At present the aerosol and trace gas inversions in Fig. 1 are presented as the same, whereas the former uses Tikhonov regularization and the latter optimal estimation. This should be reflected in the figure as it is in the text.

b) The code used to analyze the 18 month data set presented in the work utilized a Gauss-Newton (GN) iteration scheme for inversion, however, MMF has since been updated to utilize a Levenberg Marquardt (LM) iteration scheme, as well as other more minor updates. At present both schemes are described somewhat in parallel, and the authors are diligent in describing which scheme they are discussing. Nonetheless, equations for the GN scheme, which was used, are sometimes left out in favor of the more current LM scheme equivalents, leaving the methods applied not fully transparent to the reader.

I would recommend describing the GN scheme as default as it is most relevant to the titular topic of the work and collecting and describing the changes for the LM scheme either all together in a dedicated section or within the relevant subsections.

c) At the top of page 8 is the following paragraph :

"For the aerosol retrieval used in this study, we use the L1 operator ($R = L1^T \alpha L1$) where the scaling parameter α is supplied via an input script to limit the degrees of freedom (DOF) to just slightly above 1. Different scalings for the upper layers and lower layers can be supplied, as well as a complete regularization matrix R."

I understand the latter sentence to describe a capability of MMF, but how was the regularization conducted for the analysis presented later? Was a constant α determined such that the DOF was just over 1 or was something else done?

d) Discussing the advantages of MC RTM codes, the ability to model statistically rare photons and output information of the distribution of photons is also useful. In particular the statistics are worth mentioning as they quite intuitively play into the time trade-off.

e) For the aerosol retrieval on page 11, line 13-14 "The average sing scattering albedo ω and asymmetry parameter g are not subject to retrieval and are constant in all layers". What values are used?

f) The necessary inputs for VLIDORT are normalized as the authors state e.g. page 9 line 4-5, but this should be made clear more consistently. For instance the listed elements 4-6 on page 9 lines 9-11 should be "normalized rate of change ….". Similarly on page 12, line 10 "… what needs to be done is to calculate the normalized derivatives …" as this is what is presented in Eq. 11,12

3) Regarding Section 5 and error analysis:

a) In the description of the averaging kernels and degrees of freedom it should be noted that both are relative to the a priori information. This is especially important for the aerosol retrieval which uses Tikhonov regularization which yields an unbiased estimator contingent upon the a priori. E.g. on page 14 line 21 language similar to "DOF, the number of pieces of information independent of the a priori in the profile retrieval, …" should be used.

b) Section 5.3.1 is difficult to parse, particularly the first sentence: "The error originating from the cross-section is estimated by assuming that the column amount regarding to the used cross-section has a uncertainty of 3% (Wang et al., 2017)". I assume Eq. 25 has an error and should have 3% or 3.0% rather than 0.3%, otherwise I am misunderstanding. A clearer distinction in the language regarding errors in the measurements (y) as opposed to in the column or partial columns (x).

c) The error budget is composed in a number of different ways with some common terminology describing similar errors in the aerosol and NO_2 retrievals. This is relatively clear and transparent in Table 1, but can be difficult to follow in the text. For instance the measurement of error in NO_2 is 2.4% first quoted on page 16 line 6. Later on page 17 line 27 "measurement of noise" of 2.2% is quoted, this latter number is measurement noise in O_4 propagated to the NO_2 retrieval, a different quantity, nonetheless it can seem inconsistent. Earlier and more frequent reference to Table 1 would be useful I offer a key example:

The language at the end of Section 5.3.2 should be revised, it is difficult to understand precisely. Starting at page 17 line 27: “The propagation of the smoothing (4.6%) and measurement noise (2.2%) errors of the O4 retrieval into the NO2-retrieval results in a 5.1% error in the NO2 VCD” this appears to refer to Table 1 line 9 and is reasonably clear perhaps end the sentence here. Continuing, “while if no O_4-retrieval is performed successfully the error would be in our example 9.8%”, here as I understand it line 7 of Table 1 is now substituted without reference to other errors, this should be stated explicitly. Finally, “In case we would include the algorithm error (7.8%) introduced by Wang et al. (2017) the error when a O_4-retrieval is performed successfully would be 9.4%.” This is reasonably clear but there appears to be a discrepancy with line 10 of Table 1.

In Fig. 2 the NO_2 dSCD errors are shown, is the variability largely a reflection of the relative magnitude of the underlying dSCDs? Are the proportional errors reasonably constant areound the 2.4% value quoted in Table 1, or do they vary with viewing angle also?

4) Regarding the Results and Conclusions

a) For the limited degrees of the aerosol retrieval, the authors state (page 21 lines 2-3) that “Currently, the integration times in the spectra from which the O_4 dSCDs are calculated, are not long enough to ensure an O_4 dSCD error resulting in DOF larger than 1 for the aerosol retrieval.” However, based on the error budget presented in Table 1, the measurement noise in O_4 is the smallest component. Should increased integration times be expected to yield significant improvement? In the next sentence: “Since we use a Tihkonov regularization for aerosol retrieval, this means that we can basically retrieve the total aerosol extinction.” Based on Fig. 4(b) the retrieved DOF is approximately a total column below ~5.5km, very likely similar to AOD under most circumstances, but not necessarily the same.

b) Regarding the comparison with in situ NO_2 measurements, the authors highlight the impact of clouds on the comparison in Figs. 8 and 9 and examine the diurnal and seasonal components of the comparison in Figs. 11 and 12 respectively. Figure 10 to some degree combines all these aspects in the context of case studies. I wonder whether it is possible to build on this further. For instance, the slope of a MAX-DOAS – in situ comparison can be to some degree inferred from the information presented in Figs. 11 and 12, are there sufficient statistics to present Pearson’s R on these graphs also? If so it might bring greater precision to some of the discussion. Similarly, the results in Figure 8 should have some diurnal and seasonal variation which would help point to the representativeness of the effects highlighted in the Fig. 10 case studies and accompanying discussion.

The caption to Fig. 8 says the slopes where forced to zero, while in Fig. 9 the fits have non-zero intercepts. Why the inconsistency? Does this have any significant impact?

At present it is difficult to make much of the point cloud in Fig. 9, are the correlations reasonably linear across the space? Binning data and presenting statistics might provide better insight than the present graph.

c) The authors conclude that the MAX-DOAS “systematically underestimates the ground level concentrations…”, however, this is relative to a single in situ sensor and could in part reflect systemic persistent horizontal inhomogeneity. Such effects have been observed before e.g. (Dunlea et al., 2007; Oetjen et al., 2013; Ortega et al., 2015; Shaiganfar et al., 2011). Particularly at UNAM and in Mexico City, (Rivera et al., 2013) highlights that the MAX-DOAS at UNAM is likely to sample across a significant horizontal gradient. This is relevant to the later discussion of future plans to compare with more sites and with satellites, especially as the Acatlán and Vallejo sites should have overlap in their sampling (Arellano et al., 2016).

Minor comments:

Page 1, line 14: "… the total error is considerably large …" large relative to what? The errors do not seem atypical, further they are quantified immediately thereafter.

Page 1, line 19-20: "it is indispensable to have the proper tools to measure them not only at ground level but also throughout the boundary layer." Consider including a citation to support that contention that boundary layer measurements are indispensable as this is quite a strong statement.

Page 2, line 5: insert "been" to have: "applications of this technique have been demonstrated to"

Page 2, line 11: change "in" to "on" for "restrictions on the usage"

Page 2, lines 12-14: this is a long sentence, consider breaking. Also consider changing "and" to "which" at the end otherwise to clarify relation of clauses, i.e. "… (MCMA) which has been …"

Page 5 line 4: "and average" here should be "an average"

Page 5 line 5: Multiple errors, homogeneity vs inhomogeneity, something can be true or untrue, consider rephrasing e.g. "… since the assumption of horizontal homogeneity likely holds less well." or " … since this likely deviates further from the assumption of horizontal homogeneity."

Page 6 line 13: eliminate double negative, perhaps replace "… is not too non-linear so …" with "… is sufficiently linear such …"

Page 7 line 6: should "unlinear" here be "non-linear"?

Page 17 line 17: "equals" should be "equal"

Page 8 line 7: References to manuscripts in preparation do not appear in the reference section. Here there is a reference to a manuscript by Wang et al. on IO whereas previously on page 6 line 10 there is a reference to a manuscript by Wang et al. on HONO. Are these two difference manuscripts or is one instance a typo?

Page 9 line 4: Here "Jacobians" is capitalized whereas it was not previously, check consistency.

Page 9 line 13: "enclosed" here should replace "inclosed" which is no longer standard.

Page 10 line 15: As described above, the temperature dependence of the cross-section is not presently implemented, as such it should likely be eliminated from Eq. 4.

Page 12 lines 14-16: There are a number of formatting errors in equations specifically, Eq. 2, 16, 22, 25. Here three equations appear but only two are numbered, specifically the normalized derivative of ω is not assigned an equation number. In Eq. 12 unlike the previous equations only the simplified expression is given, not an intermediate step in the derivation.

Page 13 lines 4,6: The logarithm in Eq. 16 is base e, since ln(x) specifically appears in the text below, these should probably match to avoid the potentially for an apparent difference.

Page 13 line 21: “produce” here should be “produces”

Page 14 line 9: “constraint” here should be “constrained”

Page 14 line 12: Should “not symmetrical” here be “asymmetric”?

Page 18 line 7: The word “most rigorous” is probably not the best choice. Depending on what the authors wish to communicate, most imposing, or least supported might be alternatives.

Page 18 line 9: eliminate the before VLIDORT

Page 19 line 7: eliminate “relatively” it is not needed.

Page 20 line 3: Here “*in situ*” appears as two word in italics which I believe is the Copernicus standard for such phrases derived from Latin; “*a priori*” should I believe appear the same way.

Page 21 line 12: eliminate “in” to get “about half of”, it is not necessary

Page 23 lines 16-17: consider rephrasing sentence for clarity, perhaps “When all the coincident data is considered, regardless of if the retrieval had data available from the AERONET instrument on that day or not, the R and slope values are 0.62 and 0.39, respectively.”

Page 24 line 16: “relatively” is not needed; change “despite that there are more” to “despite there being more”

Page 26 line 8: Based on Fig. 8 and the prior text, the MAX-DOAS results are on average 0.4 (or 40%) of the ground level in situ measurement. The underestimate then is the difference, namely 0.6 or (60%) is this not the case?

Page 27 line 1: I don’t think the ‘s is needed after NO_2

Appendix A equations: Some of the numbers in these equations given with decimal precision are numeric factors and I don’t think require the decimal precision. Some instances are the leading 1’s in A8, A9, and A11, and I think all whole numbers in A16 and A20.

References: There are some formatting oddities in the references. Many but not all papers appear with both a DOI code and also a url which in many instances are redundant. The Bates citation includes a citation statistic.

References:

Arellano, J., Krüger, A., Rivera, C., Stremme, W., Friedrich, M. M., Bezanilla, A., & Grutter, M. (2016). The MAX-DOAS network in Mexico City to measure atmospheric pollutants. https://doi.org/10.20937/ATM.2016.29.02.05

Bogumil, K., Orphal, J., Homann, T., Voigt, S., Spietz, P., Fleischmann, O. ., et al. (2003). Measurements of molecular absorption spectra with the SCIAMACHY pre-flight model: instrument characterization and reference data for atmospheric remote-sensing in the 230–2380 nm region. *Journal of Photochemistry and Photobiology A: Chemistry*, *157*(2–3), 167–184. https://doi.org/10.1016/S1010-6030(03)00062-5

Burrows, J. P., Richter, A., Dehn, A., Deters, B., Himmelmann, S., Voigt, S., & Orphal, J. (1999). Atmospheric remote-sensing reference data from GOME-2. temperature-dependent absorption cross sections of O3 in the 231–794 nm range. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *61*(4), 509–517. https://doi.org/10.1016/S0022-4073(98)00037-5

Damadeo, R. P., Zawodny, J. M., Thomason, L. W., & Iyer, N. (2013). Atmospheric Measurement Techniques SAGE version 7.0 algorithm: application to SAGE II. *Atmos. Meas. Tech*, *6*, 3539–3561. https://doi.org/10.5194/amt-6-3539-2013

Dunlea, E. J., Herndon, S. C., Nelson, D. D., Volkamer, R. M., San Martini, F., Sheehy, P. M., et al. (2007). Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment. *Atmospheric Chemistry and Physics*, *7*(10), 2691–2704. https://doi.org/10.5194/acp-7-2691-2007

Hermans, C., Vandaele, A. C., Carleer, M., Fally, S., Colin, R., Jenouvrier, A., et al. (1999). Absorption cross-sections of atmospheric constituents: NO2, O2, and H2O. *Environmental Science and Pollution Research*, *6*(3), 151–158. https://doi.org/10.1007/BF02987620

Oetjen, H., Baidar, S., Krotkov, N. A., Lamsal, L. N., Lechner, M., & Volkamer, R. (2013). Airborne MAX-DOAS measurements over California: Testing the NASA OMI tropospheric NO_2 product. *Journal of Geophysical Research: Atmospheres*, *118*(13), 7400–7413. https://doi.org/10.1002/jgrd.50550

Ortega, I., Koenig, T., Sinreich, R., Thomson, D., & Volkamer, R. (2015). The CU 2-D-MAX-DOAS instrument – Part 1: Retrieval of 3-D distributions of NO₂ and azimuth-dependent OVOC ratios. *Atmospheric Measurement Techniques*, *8*(6), 2371–2395. https://doi.org/10.5194/amt-8-2371-2015

Peters, E., Pinardi, G., Seyler, A., Richter, A., Wittrock, F., Bösch, T., et al. (2017). Investigating differences in DOAS retrieval codes using MAD-CAT campaign data. *Atmos. Meas. Tech*, *10*, 955–978. https://doi.org/10.5194/amt-10-955-2017

Rivera, C., Stremme, W., & Grutter, M. (2013). Nitrogen dioxide DOAS measurements from ground and space: comparison of zenith scattered sunlight ground-based measurements and OMI data in Central Mexico. *Atmósfera*, *26*(3), 401–414. https://doi.org/10.1016/S0187-6236(13)71085-3

Serdyuchenko, A., Gorshelev, V., Weber, M., Chehade, W., & Burrows, J. P. (2014). High spectral resolution ozone absorption cross-sections – Part 2: Temperature dependence. *Atmos. Meas. Tech*, *7*, 625–636. https://doi.org/10.5194/amt-7-625-2014

Shaiganfar, R., Beirle, S., Sharma, M., Chauhan, A., Singh, R. P., & Wagner, T. (2011). Estimation of NO x emissions from Delhi using Car MAX-DOAS observations and comparison with OMI satellite data. *Atmos. Chem. Phys*, *11*, 10871–10887. https://doi.org/10.5194/acp-11-10871-2011

Thalman, R., & Volkamer, R. (2013). Temperature dependent absorption cross-sections of O_2-O_2 collision pairs between 340 and 630 nm and at atmospherically relevant pressure. *Physical Chemistry Chemical Physics*, *15*(37), 15371–81. https://doi.org/10.1039/c3cp50968k