Review of: "Is a scaling factor required to obtain closure between measured and modelled atmospheric O_4 absorptions? – A case study for two days during the MADCAT campaign (amt-2018-238)" Wagner et al., 2018

Wagner et al., 2018 address a very important topic of the need of scaling factor to bring MAX-DOAS measured differential slant column densities (dSCD) of oxygen collision complex (O_4) retrieved from 352 – 387 nm in agreement with the radiative transfer modeled dSCD at 360 nm. An extensive and very thorough evaluation of the error sources in the DOAS analysis and RT modeling is presented. The authors analyzed data from two time periods (18 June and 8 July 2013) during MADCAT campaign in Mainz, Germany, when time and location coincident MAX-DOAS, aerosol (AERONET, Ceilometer) profile measurements were conducted with a support of additional surface observations ($PM_{2.5}$, PM_{10}, temperature, pressure and relative humidity). They identified "standard" cases for DOAS fitting and for RT model simulations, and a number of potential scenarios deviating from the standard cases. The authors concluded that the agreement between the measured and modeled O_4 dAMF is almost perfect 1.01 (±0.16) on 18 June 2018. On the other hand the "measured" O_4 dAMF had to be scaled by 0.71 (±0.12) to bring in agreement with the modeled absorption for standard case DOAS fitting and RT modeling scenarios. The cause of the discrepancy was not identified.
This work is very important and is well suited for AMT publication. However, I think the article will benefit from some reorganization.

Major comments:
I think that there are two main topics that the authors are trying to address (I would say each of them is worth a separate publication):

(1) *Is a scaling factor required to obtain closure between measured and modeled atmospheric O_4 absorptions – Part A: identifying best-case scenarios based on auxiliary measurements and best practices.*

In this part the best case DOAS fitting scenario and best case RT modeling scenario should be identified based on the best available data to describe atmospheric conditions during the selected periods. Potential sources of errors for ***these particular*** scenarios should be evaluated. For example, for RT modeling:

- Mie scattering phase functions using AERONET inversion data results for size distribution and refractive index real and imaginary parts extrapolated to 360 nm from longer wavelengths (440, 675 nm). Evaluating errors associated with these particular inputs to the RT (e.g using 440 nm inversion results directly?). Please also note that the AERONET level 2.0 inversions are not available during some of the selected periods, potentially due to presence of clouds. Available dates/time are listed below:
  6/18/13 07:24:51
  6/18/13 15:34:32
  6/18/13 16:12:07
  7/8/13 05:16:20
  7/8/13 05:48:33

7/8/13 06:54:34
7/8/13 07:32:12
7/8/13 15:38:04
7/8/13 16:12:13
7/8/13 17:18:13
7/8/13 17:50:24

- Ceilometer backscatter profiles corrected by AERONET CIMEL AOD, and their errors (backscatter to aerosol extinction coefficient profiles conversion, wavelength differences, extrapolation to the surface)
- Radiosonde temperature, pressure and relative humidity measured profiles at fine grid with ECMWF ERA-Interim reanalysis above and their errors (e.g. different groups extraction of the data, usage of MERRA-2 profiles available at better than 1 km resolution near ground and every 3 hours)
- Accounting for polarization and RRS in the RT calculations and their errors (e.g. different models)
- If we consider O_4 cross section by Thalman and Volkamer (2013) accurate at all temperatures use T-dependent O_4 cross sections for RT calculations.
- Surface albedo from satellite measurement or AERONET inversion at 440 nm (which varies from 2.7 to 4% during the selected times).
- Effect of instrument FOV and pointing error, especially under shallow aerosol layer presence (the fact that measured dSCD at several low VEA are close to each other does not exclude potential error in pointing that has to be accounted for in modeling).

DOAS fitting scenario selected for the standard case can be considered best practice. The only things I would probably recommend changing is the offset from polynomial order 2 to 1 and not applying polynomial at all to the O_4 cross section due to its broad band wavelength dependency. In calculating the errors due to the fitting, I would not go to the extreme case of no offset. At low elevation angles the effective O_4 temperature is around 270K, I would suggest using O_4 cross section at 273K as one of the sensitivity cases.

There is another change I would recommend here – what quantity is actually compared.

Since the actual measurements are ground-based hyperspectral sky radiances the derived variable directly from the measurements without any assumptions about the atmosphere (accept for species effective temperatures) is the differential slant column density (dSCD). There are no passive measurements at the bottom of atmosphere that do not contain O_4 absorption, including the reference used in this study (zenith direction). From Beer's law, ignoring wavelength shift, offset and other corrections:

$$\left(\frac{\ln\left(I_{90}^{measured} - I_{VEA}^{measured}\right)}{\sigma_{O4}(T)}\right)_{\lambda\, window} = dSCD_{VEA}^{measured} =$$

$$= \underbrace{SCD_{VEA}^{total} - SCD_{90^o}}_{individual\ components\ are\ not\ measured\ directly}$$

$$dAMF_{VEA} = \frac{dSCD_{VEA}^{measured}}{VCD} = \frac{SCD_{VEA}^{total} - SCD_{90^o}}{VCD} = AMF_{VEA} - AMF_{90^o}$$

From the above discussion AMF and dAMF are quantities derived based on the assumptions made about AMF_{90} and VCD:

$$AMF_{VEA} = dAMF_{VEA} + AMF_{90^o} = \frac{dSCD_{VEA}^{measured}}{VCD} + AMF_{90^o}$$

I believe the paper will benefit if dSCD are compared directly with the RT modeled dSCD in the first section of the paper.

At the end of this section the reader should clearly see based on the best DOAS fitting and relevant to it errors and best atmosphere modeling (with its relevant errors) whether the measured and modeled dSCDs agree and to what extent.

(2) *Is a scaling factor required to obtain closure between measured and modeled atmospheric O_4 absorptions – Part B: error analysis to explain potential causes of SF (varying the parameters outside of (1).*

This section can include all the other cases for (d)AMF comparisons. Its main purpose could be to make recommendations and identifying problems with using less realistic atmospheric scenarios in the MAX-DOAS data inversions and DOAS fitting limitations.

Minor comments:

1. The paper is very long and difficult to read due to constant references to the appendices and main body figures and tables. Some of the figures and tables can be consolidated or eliminated.
2. Clear days are probably more appropriate to call cloud-free?
3. L 49 … agree within 1% with the corresponding radiative transfer simulations at 360 nm
4. L246: which version of LIDORT is used in this study?
5. L277: rephrase to make clear that the comparison is done between hyperspectral fitting DOAS analysis vs. singe wavelength
6. What is the source of extraterrestrial irradiance used for synthetic spectra simulation?
7. L293: Level 2 data are available now. It will be good to comment how it compares to level 1.5.
8. L306: Link from the pdf does not work, URL is valid.
9. Abstract refers to the campaign MAD-CAT, other places MADCAT
10. L348: Intensity Offset polynomial of order 2 is quite large. Can you please explain why it was chosen?
11. L903: Can you please explain how 1.01±0.16 and 0.71±0.12 are calculated? Is this for the entire two days and all observation geometries?

Time scale on Fig. 1 for the top panel (A) is unclear.