Dear Jochen,

We have uploaded our replies to the reviewer comments and the revised version of our paper. In addition, we also uploaded an offline comment from Rainer Volkamer, Ted Koenig and Ivan Ortega from 02.01.2019. The email was sent directly to me after the official discussion phase was closed. It was thus not automatically stored in the discussion forum.
The email contained a number of suggestions, from which especially one turned out to be very important for the interpretation of the results in the paper (for details see below). After the manuscript was updated with the new results, it was sent to Rainer Volkamer, Ted Koenig and Ivan Ortega, and they were invited to become co-authors. In the following weeks, a long sequence of email exchange started. Unfortunately, this email discussion eventually turned into a self-repeating, complicated and controversial one. Finally I came to the conclusion that no agreement could be reached. But still, the paper benefited a lot from this discussion.

Below, this document contains four parts:
-the reply to reviewer #1
-the reply to reviewer #2
-the email discussion with Rainer Volkamer, Ted Koegig and Ivan Ortega (which was also put to the discussion page)
-the revised paper with track changes activated

Best regards,

Thomas

Reply to reviewer #1

The replies to the reviewer comments are marked in blue

General comments

This manuscript discusses the statistical significance of the gap between observed and simulated AMFs of O4 on selected two clear-sky days during MADCAT campaign. Thorough and detailed analysis of various factors producing uncertainties in the observed and simulated AMFs was made. The authors pointed out the importance of proper usage of temperature and pressure for the condition, proper account of aerosol optical parameters (phase function, aerosol profile extraction) in the simulation, and standardization of DOAS settings (spectral range, degree of polynomial etc) for observations. Considering these factors altogether, the authors conclude that the gap was insignificant on one day (June 18) but was significant on other day (July 8), supporting conclusion from some previous works. Recognizing that there is a hot debate in the community if the scaling factor is necessary, the manuscript is valuable since it provides as thorough analyses as ever provided.

We thank the reviewer for the positive assessment of our paper and for the good suggestions. We addressed them as described in detail below.

Nonetheless, I would like to request revision on the following points. First, I find the studied uncertainties could be classified into two types: those from apparently ill treatment (i.e., 203K O4 cross section, US standard atmosphere without temperature correction, no offset in the

DOAS analysis etc) and those unavoidable even with the stateofthe-art analysis. For the purpose of evaluating spread of results from multiple groups and of determining best practice to avoid potential hazard during the analysis, determination of the former type uncertainty helps. But when discussing the significance of the gap between observed and simulated AMFs of O4 critically, only latter type uncertainties should be used. In such a way better control of the determined uncertainties is recommended.

We agree that such a separation of different types of uncertainties would be helpful. Therefore we added two columns to tables 9 and 10 in which we quantify the uncertainties if optimum settings were used and sufficient independent information was available. For the radiative transfer simulations of the O4 dAMFs the uncertaintes for these optimum settings are about ±4% compared to ±(6 – 9)% for two days of the MAD-CAT campaign. For the spectral analysis the uncertaintes for the optimom settings are about ±6% compared to ±(11-13)% for the two selected days of the MAD-CAT campaign.
These findings indicate that for future campaigns the comparison of measured and simulated O_4 absorptions can ptobably be carried out with much better accuracy (if these optimum settings were used). Here it should, however, be noted that the optimum settings for the radiative ransfer simulations will require LIDAR measurements at the same wavelengths as the MAX-DOAS measurements and without a sensitivity gap close to the surface. Such measurements are currently hardly available. This information was added to the new section 4.4.

Secondly, it should be more clarified in Abstract that the precise determination of the uncertainties (+/- 0.16 and +/-0.12 here) is the main point. Careless readers may not realize the importance.

We agree and modified the abstract to make this point more clear. We also changed the title to: 'Is a scaling factor required to obtain closure between measured and modelled atmospheric O4 absorptions? An assessment of uncertainties of measurements and radiative transfer simulations for two days during the MAD-CAT campaign'.

Thirdly, possible influence of horizontal heterogeneity of aerosol optical parameters should be mentioned. When the aerosol abundance over the line of sight is becoming less with distance (which may be likely when instrument is located in a city looking out of it), the observed higher O4 dAMFs might be better explained by considering such inhomogeneity even on July 8. I understand that with 1-D radiative transfer models homogeneity needs to be assumed and detailed discussion would be beyond the scope. However, some simple analysis such as that on spatial distribution of AOD from satellite with a fine resolution maybe possible.

We agree that this is a potentially important aspect. However, for the two selected periods the wind direction and wind speed were rather constant. On 18 June the wind direction was between 80° and 150° wrt North, and the wind speed was about 2 m/s. On 8 July the wind direction was between 70° and 90° wrt North, and the wind speed was about 3 m/s. Thus on 8 July the wind came from almost the same direction at which the instruments were looking. Taking the wind data into account, during the 4 hours of the selected period on 8 July, the air masses moved along a distance of about 40 km. During the 3 hours of the selected period on 18 June, the air masses moved along a distance of about 20 km. These distances are larger than the distances for which the MAX-DOAS observations are sensitive. Since also the AOD and the aerosol extinction profiles were rather constant during both selected periods, we conclude that for the measurements considered here horizontal gradients can not explain the discrepancies between measurements and observations. It should also be noted that the

discrepancies were simultaneously observed at all 4 azimuth directions. We added this information to section 4.2.1.

Lastly, conciseness should be attained during revision. I would suggest shortening section 4.1 and section 5 (paragraphs before section 5.1).

We moved several parts of section 4.1 to the appendix. We also shortened the paragraphs before section 5.1.

Overall, I would suggest minor revisions on the general comments above and some specific comments listed below.

Specific comments

1. Line 359. Probably appendix A2?

Corrected

2. Line 526. US standard atmosphere

Corrected

3. Figure 10. What are the differences of the first three series, with same legend "HG AP 0.6?"

The correct labels (0.60, 0.68, and 0.75) were added.

4. Figure 11. Although the panel is for showing noise influence, the gap related to the main conclusion of this study is well represented as the difference in the O4 optical depths in the first two panels. Such discussion should be added in section 4.3.1.

We added the following sentence to section 4.3.1:
'Here it is interesting to note that the ratios of the results for the measured spectrum and the simulated spectra are between 0.68 and 0.74, similar to ratio for the dAMFs on 8 July shown in Table 8.'

5. Table A12 in line 1922 is mislabeled. (Table A10)

Corrected

6. Table A11. MCARTIM

Corrected

7. Lines 846-848. Second and third points should be exchanged, considering the order of Fig. 14b and c and the following discussion.

The order was changed

8. Line 906. Overall uncertainty calculation deriving 0.12 is not clear. When considering 3% uncertainties for VCD, 6.1% from radiative transfer simulation, and 10.8% from spectral analysis, the overall uncertainty may be 13%. When it is around 0.71, it can be 0.09?

Many thanks for this hint! We agree and updated the calculations accordingly (with slightly modified uncertainties, see tables 9 and 10.

9. Line 944. 8 July

Corrected

Reply to reviewer #2

The replies to the reviewer comments are marked in blue

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 (O4) 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 (PM2.5, PM10, 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 O4 dAMF is almost perfect 1.01 (±0.16) on 18 June 2018. On the other hand the "measured" O4 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.

We thank the reviewer for the positive assessment of our paper and for the good suggestions. We addressed most of them as described in detail below.

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 O4 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 O4 cross section by Thalman and Volkamer (2013) accurate at all temperatures use T-dependent O4 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 O4 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 O4 temperature is around 270K, I would suggest using O4 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 O4 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 AMF90 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 O4 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.

We thank the reviewer for this suggestion. We understand the intention, but we decided not to split the paper into two parts. The main reason is that both suggested parts are closely linked and it would thus be difficult for the readers to follow them when split into separate papers. In addition, the suggested part 2 would be rather short and mostly speculative, because the reason for a scaling factor is still not known.
Thus we addressed the suggestion of the reviewer by including a new section (section 5.2 'Which conditions would be needed to bring measurements and simulations on 8 July into agreement?'). In that section changes of the measurement conditions are discussed which could bring measurements and simulations into agreement.

The detailed suggestions of the reviewer given above (for part 1) are addressed below:

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?).

In our opinion, we already selected scenarios for the quantitative comparison which are (at least close to) the optimum choice. On both days we selected periods around noon, for which the measured intensities are high and the variation of the SZA is small. Moreover, during the selected periods, the variation of the ceilometer profiles is relatively small compared to before and after. We added this information to section 3.2.

Many thanks for the information about the available phase functions! We performed sensitivity studies to quantify the effect of the extrapolation of the phase functions. We found that the O4 (d)AMFs hardly change (<1%) if either the phase functions at 440 nm or extrapolated to 360 nm are used. Similar small changes are found if the phase functions before or after the selected periods are used.

- Ceilometer backscatter profiles corrected by AERONET CIMEL AOD, and their errors (backscatter to aerosol extinction coefficient profiles conversion, wavelength differences, extrapolation to the surface)

As stated above, the variation of the ceilometer backscatter profiles was relatively small during the selected periods.

- 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)

In principle one could use fine grid ECMWF ERA-Interim reanalysis data, but since the uncertainties related to the temperature and pressure profiles are rather small compared to other uncertainties, we did not use additional meteorological data.

- Accounting for polarization and RRS in the RT calculations and their errors (e.g. different models)

As shown in our study, the effects of polarization in the RT are negligible. RRS was taken into account for the synthetic spectra and almost perfect agreement with the simulated O4 (d)AMFs was found. Thus we conclude that the effects of polarization and RRS can be neglected.

- If we consider O4 cross section by Thalman and Volkamer (2013) accurate at all temperatures use T-dependent O4 cross sections for RT calculations.

This is in principle a good idea. However, the effect is probably very small as indicated by the very good agreement of the results from the synthetic spectra and the simulated O4 (d)AMFs.

- Surface albedo from satellite measurement or AERONET inversion at 440 nm (which varies from 2.7 to 4% during the selected times).

The variation of the surface albedo could also be taken into account, especially if it deviates strongly from the 'standard settings'. However, as shown in our study, the influence of small changes (e.g. from 5% to 3%) on the O4 (d)AMFs is rather small (below 1%).

- 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).

We agree with the reviewer and performed additional sensitivity studies varying the FOV and also systematically distorting the elevation calibration by ±0.5°. The changes of the O4 (d)AMFs were below 1%. We added this information to the text (section 3.2).

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 O4 cross section due to its broad band wavelength dependency.

In our opinion there might be good reasons for increasing the degree of the fitted intensity offset. For example, the relative contribution of spectral stray light could cause an intensity offset in the measured spectra, which changes non-linearily with wavelength. Thus we think it is difficult to give a clear recommendation on the degree of the intensity offset.
We added the following text in section 4.3.2: 'Higher order intensity offsets might compensate for wavelength dependent offsets (e.g. spectral straylight), which can be important for real measurements, while the synthetic spectra do not contain such contributions.'

Concerning the application of the polynomial, there might be a misunderstanding. We included the O4 cross section without any previous high or low pass filtering. Concerning the degree of the DOAS polynomial we see good reaons to use such a polynomial, e.g. that the broad band wavelength dependence of the measured spectra are different for the different elevation angles, and also change with time. The very good agreement between the results of the synthetic spectra and the simulated O4 (d)AMFs indicates that the chosen polynomial degree is not problematic.

In calculating the errors due to the fitting, I would not go to the extreme case of no offset.

We agree. Note that the case without intensity offset was already ignored for calculating the errors in the discussion version of our paper.

At low elevation angles the effective O4 temperature is around 270K, I would suggest using O4 cross section at 273K as one of the sensitivity cases.

In principle we agree with the reviewer here. However, we did not change the O4 cross section because of two reasons:
a) the effect of such small temperature changes is rather small.
b) in most existing studies, O4 cross sections for room temperature were used. Thus we prefer to stay consistent with those studies.

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 O4 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 AMF90 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.

In our opinion, the only difference to your suggestion is that we divide the O4 (d)SCDs by the O4 VCD. Both choices are equivalent. To make the interpretation of the results in units of (d)SCDs easier, we added second y-axes in Figures 2 and 3 in (d)SCD units.

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.

We understand this concern. However, one important part of the study deals with the quantification of the uncertainties of the spectral analysis and radiative tarnsfer simulations. For readers with interest in the details of the sensitivity studies the figures and tables in the appendix will be important. In contrast, for readers who are mostly interested in the general

findings the figures and tables in the main part should be sufficient. We therefore decided not to remove any figures or tables.

2. Clear days are probably more appropriate to call cloud-free?

Changed

3. L 49 … agree within 1% with the corresponding radiative transfer simulations at 360 nm

'at 360 nm' was added

4. L246: which version of LIDORT is used in this study?

Version 3.3. The version is 3.3. This information was added to the text.

5. L277: rephrase to make clear that the comparison is done between hyperspectral fitting DOAS analysis vs. singe wavelength

We added the following text: 'at one wavelength (here: 360 nm)'

6. What is the source of extraterrestrial irradiance used for synthetic spectra simulation?

We used the high resolution solar spectrum from Chance and Kurucz (2010). We added this information and the corresponding reference in section 2.4.

7. L293: Level 2 data are available now. It will be good to comment how it compares to level 1.5.

Many thanks for this hint! The Level-2 data are exactly the same as the Level-1.5 data. We removed the corresponding sentence about the Level 1.5 data from the text.

8. L306: Link from the pdf does not work, URL is valid.

Many thanks for his hint! This link should work after the final copy-editing.

9. Abstract refers to the campaign MAD-CAT, other places MADCAT

Now consistently 'MAD-CAT' is used.

10. L348: Intensity Offset polynomial of order 2 is quite large. Can you please explain why it was chosen?

The following text was added in section 4.3.2:

'Higher order intensity offsets might compensate for wavelength dependent offsets (e.g. spectral straylight), which can be important for real measurements, while the synthetic spectra do not contain such contributions.'

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?

The information was added that the ratio was calculated for the middle period of that day.

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

The corresponding labels were added.

Email discussion with Rainer Volkamer, Ted Koegig and Ivan Ortega

This author comment does not refer to the comments of the 'official' reviewers. It refers to an offline comment from Rainer Volkamer, Ted Koenig and Ivan Ortega on 02.01.2019, after the official discussion phase was closed. The email was directly sent to Thomas Wagner and is thus not stored in the discussion forum. Below this email together with the subsequent email discussion is listed.

The email contained a number of suggestions, from which especially one turned out to be very important for the interpretation of the results in the paper. Rainer Volkamer, Ted Koegig and Ivan Ortega argued that the relative backscatter profile derived from the ceilometer measurements at 1064 nm are probably not representative for the aerosol extinction at 360 nm, because the sensitivity to coarse and fine aerosols at both wavelengths is in general different. This important argument led to extensive additional calculations mainly based on the information available in the AERONET inversion products (mainly the phase functions and optical depths of the coarse and fine mode aerosols at different wavelengths). The new calculations showed that the extracted aerosol extinction profile at 360 nm had to be modified compared to the profile described in the original version of the manuscript. This modification decreased the difference between the simulated and measured O_4 (d)AMFs. However, still no agreement between measurements and simulations was found for one of both days (08 July 2013).
After the manuscript was updated with the new results, it was sent to Rainer Volkamer, Ted Koegig and Ivan Ortega. They were invited to become co-authors. In the following weeks, a long sequence of email exchange started. Unfortunately, this email discussion eventually turned into a self-repeating, complicated and controversial one. Finally Thomas Wagner came to the conclusion that no agreement could be reached, and the revised version was sent to the other co-authors. After their feedback was received, the updated paper was again sent to Rainer Volkamer, Ted Koegig and Ivan Ortega and co-authorship was again offered (but no response was received).Below, the whole email discussion is chronologically listed. Part of the discussion refers to intermediate versions of the paper, which are thus also made available.

Comments from Rainer, Ted, Ivan in black
Comments from Thomas in blue
Red: names of pdf-files

Email from Rainer Volkamer, 02.01.2019:

Dear Thomas,
Sorry for the slow response, which are due to a hectic and eventful summer.
The attached our view on the CFO4 paper. I hope you will find these comments useful.

Happy New Year!
-Rainer

Reply to Rainer, Ted, Ivan, email 27.01.2019

Dear Rainer, Ivan, Ted,

many thanks for your valuable comments!
Attached I send you the current version of the revised manuscript together with the replies to both reviewers and to your comments.
Your comments were very helpful, especially with respect to the question of the representativity of the ceilometer measurements at 1020 nm for the MAX-DOAS measurements at 360 nm.
The revised version addresses this aspect, and also the effect of elevated aerosol layers.
I want to invite you to become co-author of the paper. If you agree, please send me the correct details of your affiliation(s). In any case, your feedback would be welcome.
I already had asked for an extension of the deadline for the submission of the revised manuscript until 4 February, and I will ask the editorial office again for a further extension. Nevertheless, I would appreciate receiving your feedback within one week from now, because I also have to iterate the manuscript with the other co-authors before re-submission.

Many thanks and best regards,

Thomas

The attached pdf file is O4_scaling_factor_27012019.pdf

The detailed comments were provided in an attachment (Comments on Wagner etal CU-Boulder.docx). The content and the replies are given below:

Comments on Wagner et al. 2018 "Is a scaling factor required to obtain closure between measured and modelled atmospheric O4 absorptions? - A case study for two days during the MADCAT campaign"
By Rainer Volkamer, Ivan Ortega, and Ted Koenig

Dear Rainer, Ivan, and Ted,
Many thanks for your valuable comments. Please find our detailed answers below.

Dear Thomas,
This is a significant body of work, and valuable albeit somewhat inconclusive. We agree the present solution can be viewed as consistent with the measurements. But we also think that there is significant evidence that supports an elevated aerosol layer as a plausible explanation for CFO4 on 8 July.
You mention that Ortega et al. 2016 used a similar approach, but we were missing a connection of the present work with the findings in that study (and other our related papers). We have tried to make it easy to establish this connection by suggesting specific text that could easily be added in the introduction and discussion sections here. Feel free to modify it as you see fit, or let us know if we are missing something in suggesting these changes.

We added part of the suggested text and also sensitivity studies of the effect of elevated layers on the O4 (d)SCDs in the revised version of the manuscript (new appendix A6)

Since we did not really contribute to your work until this late point, we do not feel that we need to be added as co-authors. We would like to see our O4 data that we had submitted compared here, if we were to be added as co-authors. Alternatively, you could just add our names in the acknowledgements.

Your comments on the different sensitivities of ceilometer measurements at 1020 nm to fine and coarse mode aerosols led to important additions to and changes of the manuscript (for details see below). Thus we would like to invite you to become co-authors of the paper.

Sorry for the slow response, which are due to a hectic and eventful summer.
Best wishes,
-Rainer

Specific comments:
1) Abstract, line 35: "many studies, in particular based on direct sun light measurements…". Most studies that concluded on the lack of a need for a correction factor in Table 1 are actually scattered sun light measurements, so this is misleading. Revise language to reflect scattered and direct sun light.

We changed 'in particular' to 'including such'

2) Please add the following studies in Table 1, among "studies that did not apply a scaling factor":
- Volkamer et al., 2015, AMAX-DOAS, 360, 477nm – see Figs. 3 + 4. doi:10.5194/amt-8-2121-2015
- Thalman and Volkamer 2010, CE-DOAS, 477nm – see Figs. 8 + 9. doi: 10.5194/amt-3-1797-2010

Both references were added

3) Abstract, line 37: "Up to now, there is no explanation for the observed discrepancies between measurements and simulations." change to "no broad consensus for an explanation", or eliminate entirely. Note that Ortega et al. 2016 provides the following explanation (quote from the abstract): "However, if in the calculations the aerosol is confined to the surface layer (while keeping AOD constant) we find 0.53<CFO4<0.75, similar to previously reported CFO4. Our results suggest that elevated aerosol layers, unless accounted for, can cause negative bias in the simulated O4 dSCDs that can explain CFO4."

The text was changed as suggested.

Fig. 6 and Table 3 of that paper demonstrate, that - surprisingly - elevated aerosol layers mostly modify the O4 SCD in the lower elevation angles. This is somewhat counterintuitive, but warrants a sensitivity study in your paper in our opinion.

Elevated layers (even at higher altitudes than in Ortega et al. (2016)) were already considered in our paper and could not explain the observed discrepancies on 8 July (see also below). Nevertheless, we added sensitivity studies about the effect of increased aerosol extinction at elevated layers to the paper (new appendix A6). We found that the effect of elevated layers on the O_4 (d)AMF is rather small.

4) We agree the present solution is consistent with the measurements. But we also note that the information available (see below) is indicative of an aerosol layer aloft being missed by the ceilometer on 8 July. Could an aerosol layer aloft explain the need for CFO4 on 8 July?
Summary of results in Ortega et al 2016:

- there is no issue with the O4 measurements.
- neglecting layers aloft biases the RTM low for low elevation angles.
- overestimating layers aloft biases the RTM high for low elevation angles.
- there is no bias if the correct profile is represented as input to the RTM.

By extension, we expect there to be a layer aloft on 8 July, but not on 18 June. The higher Angstroem exponent on 8 July is consistent with this expectation, and warrants further discussion in the paper. The below points 7-9 further elaborate on this point.

We now discussed the effect of elevated layers in detail in several parts of the paper, see also the detailed response below.

5) The abstract highlights the "need for more detailed independent aerosol measurements" (line 62). We agree. Consider adding the following text to summarize existing literature on this point:
Introduction: "Optical closure studies based on measured O4 SCDs show excellent agreement also in presence of aerosols, when detailed information from independent aerosol measurements is available. Specifically, the aerosol extinction profile inferred from altitude resolved AMAX-DOAS O4 observations agrees very well, and quantitatively, with the aerosol extinction profile measured by collocated airborne High Spectral Resolution Lidar (HSRL) (Volkamer et al. 2015). The HSRL wavelength (532nm) in this study closely resembled that of the O4 wavelength (477nm). HSRL is highly sensitive to quantify aerosols even at the low aerosol extinction in the free troposphere. The existence of elevated aerosol layers has been suggested as a possible explanation for CFO4 (Ortega et al. 2016). Furthermore, when the size distribution and refractive index of aerosols are actively controlled, CE-DOAS O4 observations infer aerosol extinction values that agree very well with Mie calculations of the extinction constrained by these known aerosol properties (Thalman and Volkamer, 2010). Any scaling of the measured O4 SCD by CFO4 smaller unity would lead to systematic high bias in the inferred extinction in either of these studies, which is not observed. These studies thus provide strong evidence from field studies and laboratory experiments, that there is no fundamental limitation to use O4 SCDs to infer aerosol extinction."
Some language on the relation between microphysical properties and macroscopic scattering and extinction is needed to make this point more clearly. Of particular relevance is the aerosol size distribution. At the moment, the only mentioning of the aerosol size distribution is on line 964 in relation to g on 18 June. However, the size distribution strongly impacts g and the Angstroem exponent. In fact, the aerosol size distribution as the underlying property that controls and probably explains differences in the Angstroem exponent on 8 July seems very relevant. This is currently missing, and worth mentioning.

See detailed reply below.

6) line 116: "… similar to Ortega et al. 2016… "– given a similar approach is used, how do the results compare? Some discussion seems appropriate, and is currently missing in Sections 4.1 and 4.2. For example, Ortega et al was the first study to my knowledge that systematically looked at the issue of using realistic density profiles vs a US Standard atmosphere.

Consider adding in Section 4.1: "Ultimately, the accuracy with which O4 concentrations can be calculated is limited by the assumption that O2-O2 is pure collision induced absorption (CIA). If the oxygen concentration profile is well known, the uncertainty due to bound O4 is smaller 0.14% in Earth's atmosphere (Thalman and Volkamer, 2013). By comparison, deviations of air density from the US standard atmosphere (pressure, temperature and humidity) can lead to errors of 15-18% in estimated O4 concentrations (Ortega et al. 2016). Here we investigate different extraction methods… "

At the end of section 4.1 we added the reference to Ortega et al., 2016.
We also added the following text: Ultimately, the accuracy with which O_4 concentrations can be calculated is limited by the assumption that O_4 (O_2-O_2) is pure collision induced absorption. If the oxygen concentration profile is well known, the uncertainty due to bound O_4 is smaller 0.14% in Earth's atmosphere (Thalman and Volkamer, 2013).

7) line 527 ff: This is a key paragraph in my opinion. We agree with the listed assumptions, but am missing discussion of the uncertain wavelength scaling. In other words, how certain can you be that the profile shape measured at 1064nm actually resembles the profile shape that drives AOD at 360nm?
The ceilometer wavelength is sensitive primarily to larger particles, has limited sensitivity aloft, and could miss smaller particles that are expected more abundant aloft. At the same time, smaller particles contribute more effectively to AOD at 360nm (Q factor due to Mie resonances). In this context, it is interesting to note that the Angstroem exponent on 8 July is rather large (Fig. 1C). We suspect that the ceilometer data is less representative of the actual aerosol profile shape at 360nm on 8 July. And that the profile shape from the ceilometer is actually a better proxy at 360nm on 18 June due to the lower Angstroem exponent on that day. Uncertain wavelength scaling being more important on 8 July seems relevant to the discussion of Section 4.2, and also in Section 5.1, b) aerosol properties.

This is a very good and important remark. We investigated this effect and found that – surprisingly - the ceilometer measurements are actually a better proxy for the aerosol extinction profile on 8 July than on 18 June. The corresponding calculations are added at the end of appendix A5.

8) Section 5.1, and Fig. 1: The aerosol backscatter from the ceilometer indicates lower boundary layer height on July 8 (the day with low PM load). In our opinion, the discrepancy in in-situ PM between both days is indicative that small particles need to be above the boundary layer in order to explain similar AOD [If the boundary layer is shallower and the PM load measured at the surface is lower, than the total aerosol load in the boundary layer is lower by extension. Despite this, July 8 has a similar total AOD. Without aerosol aloft, this would require a greater extinction per unit mass for the boundary layer.]
We suspect the non-zero aerosol extinction retrieved consistently at high altitudes on July 8 (in contrast to June 18), coupled with the higher Angstroem exponent that day, provide two important clues to resolve the apparently inconsistent conclusions regarding CF04 on both days. In our opinion the origin for the significant CFO4 on July 8 lies in the shape of the aerosol profile from the ceilometer being propagated as seen in Fig. 8. Note that the ceilometer is not sensitive to smaller particles.

Again we refer to the detailed calculations which were added to appendix A5 of the revised version of the manuscript, which is also copied at the end of this file.
We also added the following text at the end of section 4.2.1:

Finally, we investigated the effect of changing aerosol optical properties with altitude (changing LIDAR ratio). Such effects are in partivular important if the wavelength of the ceilomter measurements (1020 nm) differs largely from that of the MAX-DOAS observations (360 nm). Based on the partitioning in fine and coarse mode aerosols derived from the sun photometer observations, as well as the corresponding phase functions and optical depths, the sensitivity of the ceilometer to fine mode aerosols can be estimated (for details see appendix A5). While for 18 June the contribution of the fine mode to the ceilometer signal is about 32% on 8 July it is much larger (about 82 %). Thus it can be concluded that the aerosol extinction profile derived rrom the ceilometer is largely representative for the fine mode aerosols on that day. Nevertheless, the remaining uncertainties of the aerosol extinction profile at 360 nm together with the assumption that the coarse aerosols are probably located close to the surface led to a repartitioning of parts of the aerosol extinction profile (extracted assuming a constant LIDAR ratio). This repartitioning led to a decrease of the aerosol extinction close to the surface which is balanced by an increase at higher altitudes (see Fig. A34). The O4 dAMFs calculated for the modified profile are by about 15 % larger than those for the standard settings (for details see appendix A5).

Ortega et al emphasizes the importance of the aerosol profile shape for a given AOD. We concluded in that paper that profile shape uncertainty is actually more important than air density uncertainty as drivers for CFO4 (Section 3.4 in Ortega, and their Fig 5 and Fig 7). In our opinion, a similar sensitivity study that varies the profile shape at constant AOD is needed here, essentially redistributing a fraction of the AOD to the layer that is visible in the ceilometer data near 7km on 8 July.

We added such sensitivity studies to the revised version of the manuscript (new appendix A6). The results indicate that the effect of redistributing large parts of the aerosol extinction profile leads to rather small changes of the O_4 dAMFs (between –7 and +7 %) for rather large changes of the aerosol extinction (+40%) at different layers.

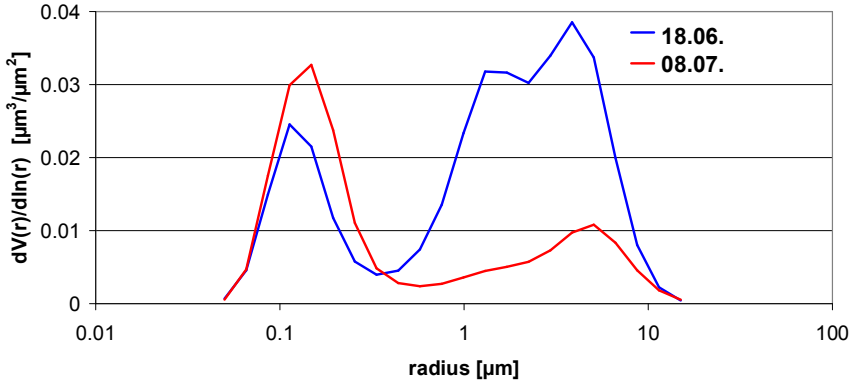
9) In Section 5.1, b) aerosol properties:
- Please add the fraction (%) of the AOD that resides below 1km, between 1-2km, and above 2km for both days.
- It would be interesting to add a few sentences how the aerosol profile shapes compare with the climatology for elevated aerosol layers - see section 3.3 in Ortega for context. How do the days during MADCAT compare with the days studied during TCAP?

Since for the measurements selected in this study elevated layers can not explain the discrepancy between measurements and simulations, we think it is not important to add this information. Nevertheless, we compared the aerorosl profiles for the selected days with those in Ortega et al. (2016).
At the end of section 4.2.1 the following text was added:
The effect of elevated aerosol layers (see Ortega et al., 2016) was further investigated by systematic sensitivity studies (appendix A6). On both selected days enhanced aerosol extinction was found at elevated layers (Fig. 9). Compared to those reported by Ortega et al. (2016) the profiles extracted in this study reach even up to higher altitudes. For the investigation of the effect of changes of the aerosol extinction at different altitudes, the aerosol extinction profile on 8 July was subdivided into 3 layers (0-1.7 km; 1.7 – 4.9 km; 4.9 – 7 km), and the extinction in the individual layers was increased by +20% or + 40 %. It was found that even a strong increase of the aerosol extinction at high altitudes by 40% leads only

to an increase of the O4 dAMFs by 7 %. Here it should be noted that on 8 July no indications for such a strong underestimation of the aerosol extinction at high altitudes are found.

D) Influence of a changing LIDAR ratio with altitude

For the extraction of the aerosol profiles described above, a fixed LIDAR ratio was assumed, which implies that the aerosol properties are independent from altitude. However, this is a rather strong assumption, because it can be expected that the aerosol properties (e.g. the size) change with altitude. With the available limited information, it is impossible to derive detailed information about the altitude dependence of the aerosol properties, but it can be how representative the ceilometer measurements at 1020 nm are for the aerosol extinction profiles at 360 nm. For these investigations we again focus on the middle periods of both selected days. From the AERONET Almucantar observations information on the size distribution for these periods is available (see Fig. A32). On both days two pronounced modes (fine and coarse mode) are found with a much larger coarse mode fraction on 18 June compared to 8 July. From the AERONET observations, also separate phase functions for the fine and coarse mode as well as the relative contribution of both modes to the total aerosol optical depth at 500 nm are available. On 18 June and 8 July the relative contributions to the total AOD at 500 nm are 40 % and 5 %, respectively. Assuming that the AOD of the coarse mode fraction is independent on wavelength, the relative contributions of the coarse mode at 360 nm and 1040 nm can be derived (see Table A27).



Fg. A32 Size distributions derived from AERONET Almucatar observations on 18 June (07:24 & 15:34) and 08 July (07:32 & 15:38).

Table A27 Contribution of the coarse mode to the total AOD at different wavelengths

Date	Total AOD 360 nm	Total AOD 1020 nm	Relative contribution of coarse mode 360 nm	Relative contribution of coarse mode 1020 nm
18 June, 11:00 – 14:00	0.37	0.12	24.9%	77.7%
08 July, 07:00 – 11:00	0.33	0.055	3.0%	18.1%

It is found that on 18 June the coarse mode clearly dominates the AOD at 1020 nm, whereas on 8 July it only contributes about 20 % to the total AOD. As expected the relative contributions of the coarse mode to the AOD at 360 nm are much smaller (25 % and 3%).

In the last step the probability of aerosol scattering in backward direction is considered, because the ceilometer receives scattered light from that direction. For that purpose the ratios of the optical depths are multiplied by the corresponding values of the normalised phase functions at 180° and in this way the relative contributions to the backscattered signals from the coarse mode for both wavelenghs and both days are calculated (Table A28). Interestingly, on 8 July the contributions of the coarse mode to the backscattered signal at both wavelengths differs only by about 10%. In contrast, on 18 June the difference is much larger.

Table A28 Ratio of phase functions (coarse / fine) in backward direction and relative contribution of coarse mode to the backscattered signal at both wavelengths

Date	Ratio phase function at 360 nm	Ratio phase function at 1020 nm	Relative contribution of coarse mode at 360 nm	Relative contribution of coarse mode at 1020 nm
18 June, 11:00 – 14:00	1.13	0.61	27.3%	68.0%
08 July, 07:00 – 11:00	2.7	0.99	7.8%	18.0%

For 8 July, the results can be interpreted in the following way: at 360 nm the aerosol profiles extracted as described above overestimate the contribution from the coarse mode by about 10%. To estimate the effect of this overestimation we construct modified aerosol extinction profiles, in which 10% of the total AOD is relocated. Since we expect that the coarse mode aerosols are usually located at low altitude, we construct 4 different modified profiles (see Fig. A33) with different altitudes (1.5 km, 1 km, 0.75 km, or 0.5 km), below which 10% of the aerosol extinction is relocated to altitudes above (assuming that the coarse mode aerosol is only located below these altitudes). Of course, such a sharp boundary is not very realistic, but it allows to quantify the overall effect of the relocation. Here it should be noted that we selected the aerosol profile for 8 July extracted by INTA which reached up to 7 km (see Fig. 9).

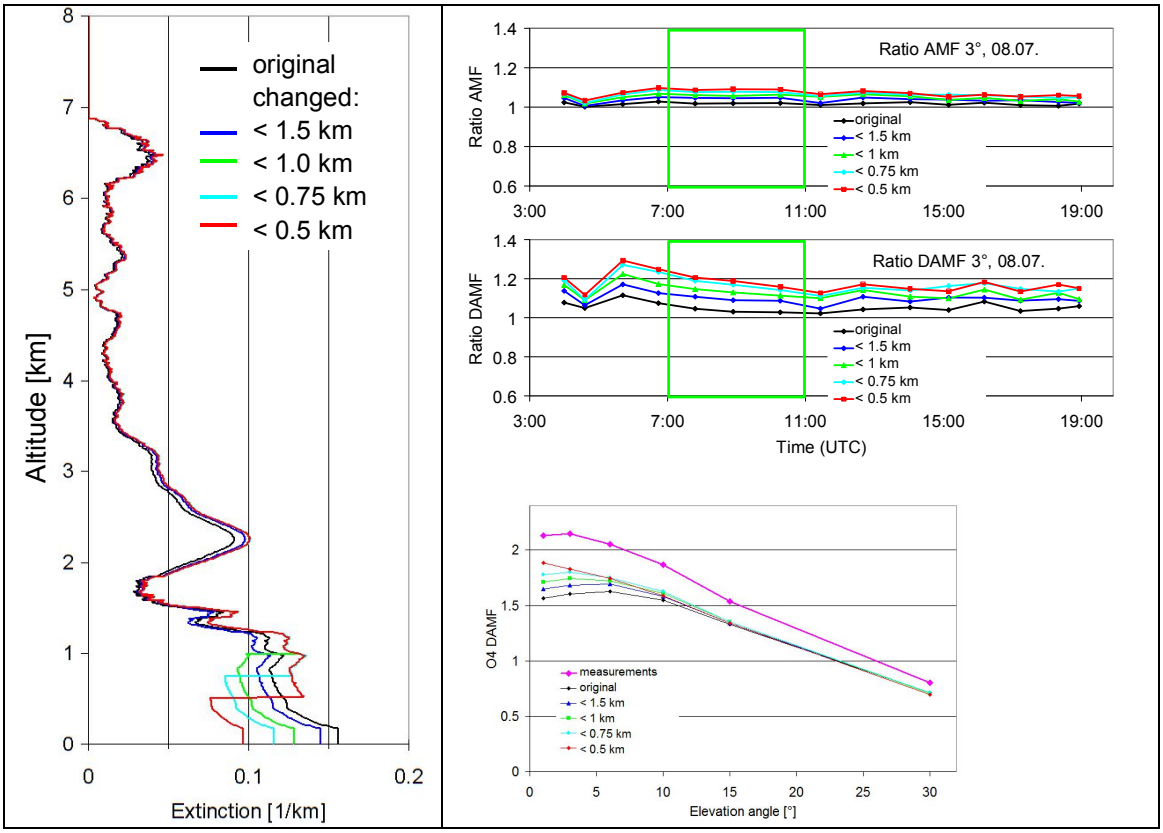


Fig. A33 Left: Modified aerosol profiles for 08 July assuming that the coarse mode aerosol is only located in the lowest part of the atmosphere. Top right: ratios of the (d)AMFs calculated for the modified profiles compared to the dAMFs for the standard settings. With decreasing layer height the (d)AMFs increase systematically, because the aerosol extinction close to the surface decreases. Righ bottom: comparison of the measured elevation dependence of the O_4 dAMFs for the period 7:00 – 11:00 on 8 July and simulation results for the different profiles.

Table A29 Ratio of the (d)AMFs for the modified profiles versus those of the standard settings

	original INTA	coarse mode below 1.5 km	coarse mode below 1 km	coarse mode below 0.75 km	coarse mode below 0.5 km
AMF	1.02	1.04	1.05	1.06	1.08
dAMF	1.04	1.09	1.13	1.17	1.18

For all modified profiles, a systematic increase of the O_4 (d)AMFs compared to those for the standard settings is found. For the O_4 dAMFs this increase can be up to 18 % (see Table A29. From the comparison of the elevation dependence of the measured and simulated O_4 dAMFs (see Fig. A33), we conclude that the aerosol profile with the coarse mode aerosol below 0.75 km is probably the most realistic one. The main conclusion from this section ist that the dAMFs for 8 July derived from the standard settings probably underestimates the true dAMF by about 15 ±5 %.

For 18 June we did not perform similarly detailed calculations, because on that day the uncertainties of the aerosol extinction profile caused by the missing sensitivity of the ceilometer below 180 m are much larger than on 8 July. On 18 June also the magnitude of the relocation of the aerosol extinction between different altitudes would be much larger than on 8 July.

Email from Rainer and Ted, 02.02.2019

Lieber Thomas,

thank you very much for the revised manuscript. The additions seem promising, but in our opinion there is one additional sensitivity study that warrants to be added in Appendix 6. Can you please add a case, where you redistribute 10% of the AOD on 8 July from the lowest layer to the upper layer. Specifically, the partial column AOD would look as follows:

Alt_low	Alt_up	Ext_INTA	Ext_redist
0	1.68	0.186	0.1523
1.68	4.9	0.116	0.116
4.9	7	0.035	0.0687

Please also include the EA dependence of the O4 dAMF in Figure A34.

And since the Figure is already quite busy, it would be helpful to digest the results for the O4 AMFs, and O4 dAMFs in form of a new Table A31.

We believe that the addition of the size distribution from both days is a key improvement to the paper. I have discussed with Ted, and we are not sure that we agree with everything that is said in Appendix 5. But in essence we agree with the main conclusion, that the profile shape at 1020nm is a poor indicator for that at 360nm. Following your argument in Table A28, up to 40% of the AOD should be able to redistribute (18 June). This does not seem to be sensible in light of the size distributions in Fig. A32, which support the mismatch in ceilometer wavelengths is much less of an issue on 18 June than on 8 July. This does not yet appear to transpire from the discussion in Appendix 5. But I hope the above sensitivity study will help move that discussion along.

We will send more detailed comments in due time, but wanted to get back to you in a timely manner.

Thanks for all your efforts, and I hope this email finds you well.

Best wishes,
-Rainer & Ted

Email from Thomas, 03.02.2019

Dear Rainer and Ted,

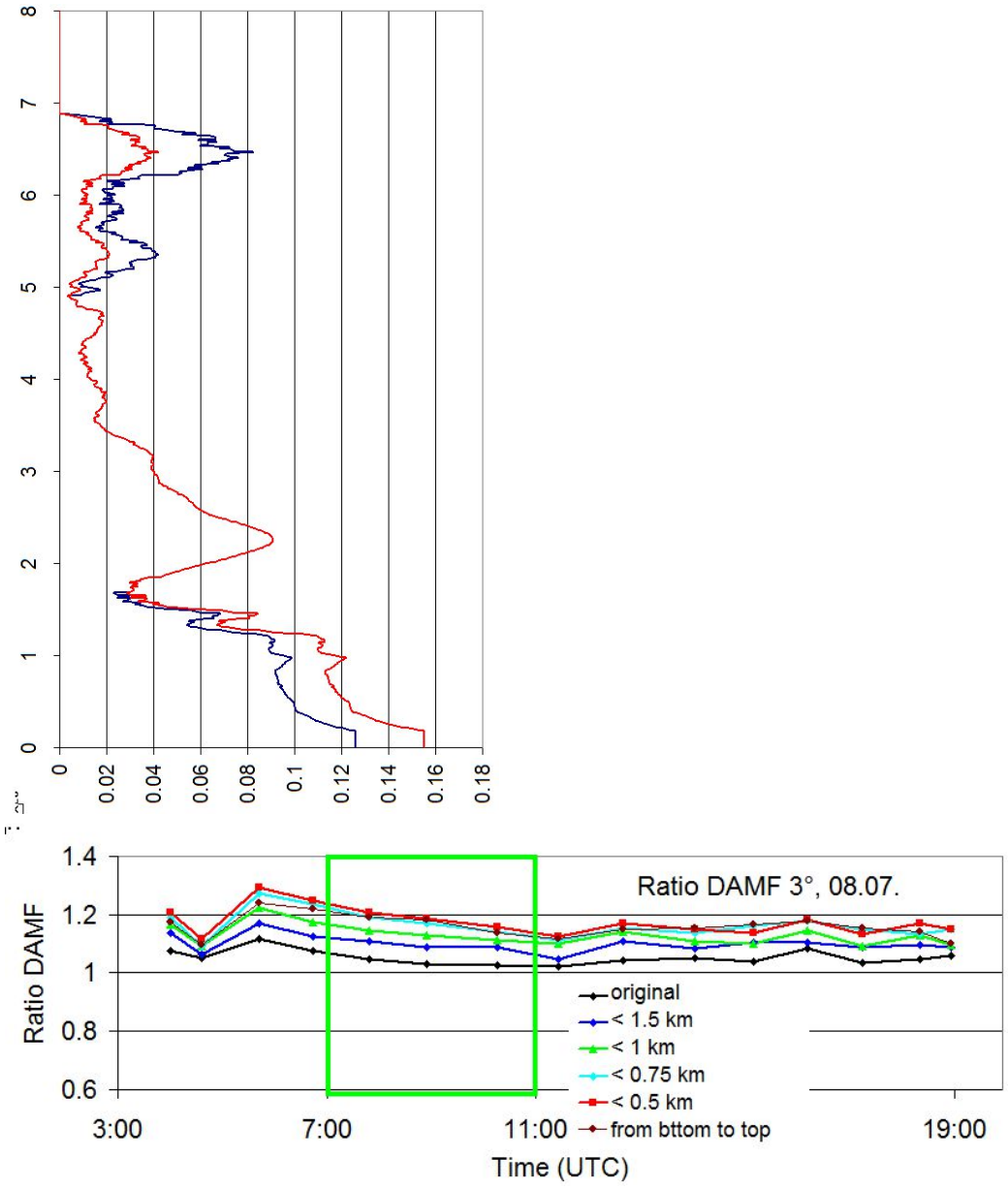
many thanks for your feedback and further suggestions. Please find my detailed replies below.

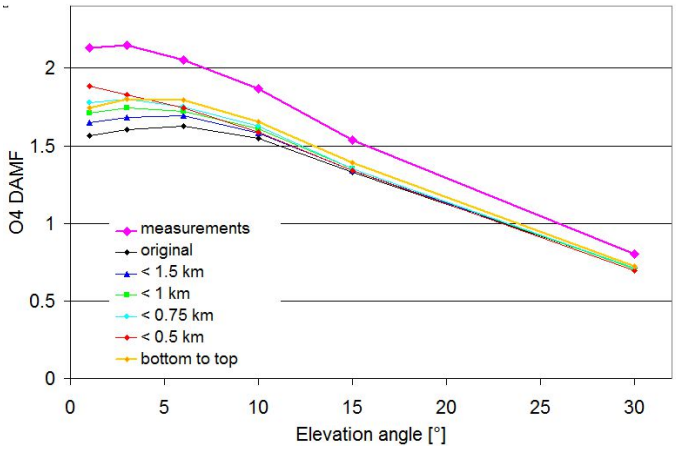
> thank you very much for the revised manuscript. The additions seem promising, but in our opinion there is one additional sensitivity study that warrants to be added in Appendix 6. Can you please add a case, where you

redistribute 10% of the AOD on 8 July from the lowest layer to the upper layer. Specifically, the partial column AOD would look as follows:

Alt_low	Alt_up	Ext_INTA	Ext_redist
0	1.68	0.186	0.1523
1.68	4.9	0.116	0.116
4.9	7	0.035	0.0687

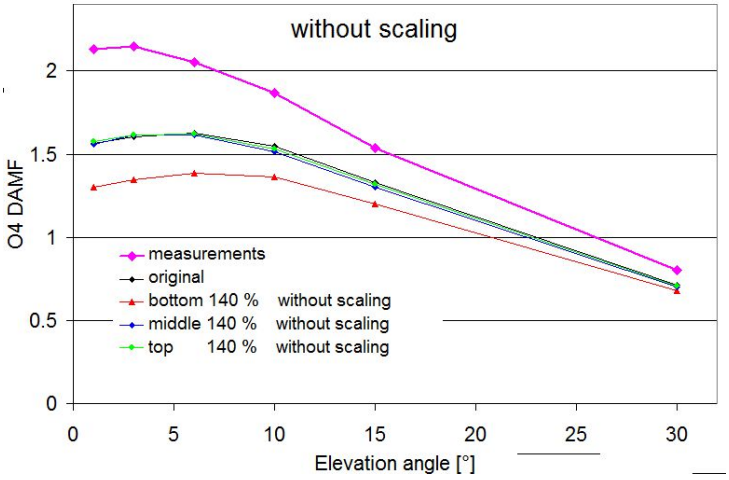
Good point! The simulations were done, please see the results below. The ratios of dAMFs for the new profile ('from bottom to top') are similar to those of the profile '< 0.75 km'. This information is now mentioned in the manuscript in appendix A5

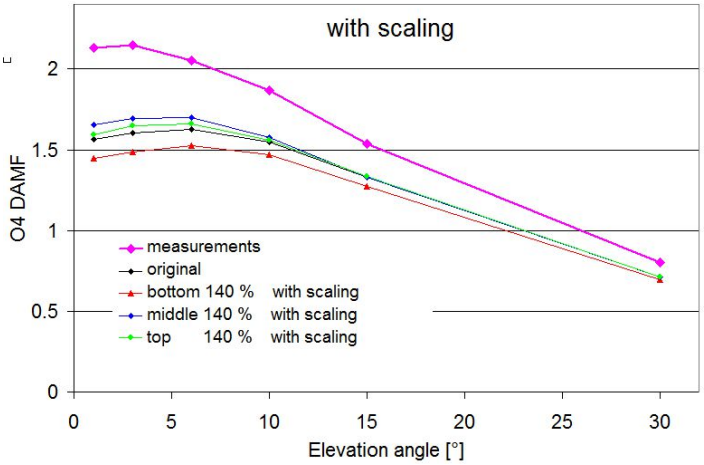




Please also include the EA dependence of the O4 dAMF in Figure A34.

The elevation dependence was added, see figures below:





And since the Figure is already quite busy, it would be helpful to digest the results for the O4 AMFs, and O4 dAMFs in form of a new Table A31.

The Table was added:

Table A31 Ratios of (d)AMFs for 8 July 2013 for the modified profiles with respect to the original profile

	low 140 %	middle 140 %	top 140 %
ratio AMF without scaling	0.95	1.03	1.03
ratio dAMF without scaling	0.85	1.02	1.02
ratio AMF with scaling	1.00	1.06	1.04
ratio dAMF with scaling	0.94	1.08	1.04

> We believe that the addition of the size distribution from both days is a key improvement to the paper. I have discussed with Ted, and we are not sure that we agree with everything that is said in Appendix 5. But in essence we agree with the main conclusion, that the profile shape at 1020nm is a poor indicator for that at 360nm. Following your argument in Table A28, up to 40% of the AOD should be able to redistribute (18 June). This does not seem to be sensible in light of the size distributions in Fig. A32, which support the mismatch in ceilometer wavelengths is much less of an issue on 18 June than on 8 July. This does not yet appear to transpire from the discussion in Appendix 5. But I hope the above sensitivity study will help move that discussion along.
>
> We will send more detailed comments in due time, but wanted to get back to you in a timely manner.

Could you already estimate when you will send me the more detailed comments? The current deadline is on 4 February. I will ask for a further extension today.

Many thanks!

Thomas

Email from Rainer, 03.02.2019

Thanks Thomas,

A quick clarification & request:
Please add the results from the new profile shape also on Table A31 and Fig A34 for direct comparison with those data. That would be most helpful.
Please also comment on our point regarding your argument supporting why redistribution of extinction should be limited to 10% (and not 40%? or 80%?)...
It seems that by redistributing only 10% of the partial AOD to higher altitudes, about half (or slightly more) of the correction factor is explained on 8 July. The effect of adding extinction aloft (while keeping AOD constant) increases the dAMF in the higher EAs, while the opposite is observed in the lower EAs if extinction is added near the surface. After scaling to normalize AOD, adding extinction aloft is relatively more efficient at closing the gap. The latest profile shape is in fact the closest of all dAMFs, better even than the 0.75km case for EA ~4.5 and larger.
This bears the question then: How much AOD would need to be redistributed on 8 July in order to obtain closure?

We will make it a priority to get back to you, but asking for a further extension seems a good idea.
Greetings on a sunny Sunday morning - it's Superbowl in the NFL today - Patriots vs LA RAMs. Meaning the city, trails and slopes should be empty this afternoon... ;)

-Rainer

Emai from Thomas, 04.02.2019

Dear Rainer,

from table A28 it is found that for 8 July the coarse mode fraction contributes 18 % to the ceilometer signal at 1020 nm, while it would contribute 8 % to a ceilometer signal at 360nm. That means there is a difference of 10% of the total AOD measured at 1020 nm, which would not be seen at 360 nm. These 10% of the coarse mode contribution could be anywhere in the atmospheric column, but most probable close to the surface. This is also supported by the elevation dependence. The elevation dependence for the profile where 10% from below 750 m is relocated to above 750 m fits best to the measurements. Results for relocations from below 500m to above 500m also those for relocation from below 1.68 km to above 4.9 km (your suggestions) don't fit. I hope this makes the argument more clear.
You are right that the results for the latest profile shape is in fact the closest of all dAMFs, better even than the 0.75km case for EA ~4.5 and larger. However, the complete elevation dependence does not fit to the measurements.
There is another point: you wrote: It seems that by redistributing only 10% of the partial AOD to higher altitudes, about half (or slightly more) of the correction factor is explained on 8 July. This is not really true. The ratio between measurements and simulations (without the relocation) is 0.71. If it is multiplied by 1.15 one gets: 0.82.
I will add the results for the new profile shape to Table A31 and Fig A34

Best regards,

Thomas

Email from Rainer, 05.02.2019

Hi Thomas,

can you send us the McArtim files and O4 measurements for both days?

We probably want to run some simulations here ourselves to have an effective conversation. McArtim3 is also what we usually use. Its a good exercise for Chris (cc here) to setup calculations for the 18.6. and 8.7 case study days and inform further discussions.

Please also send the size distribution files from Aeronet, and any info on the complex refractive index (and its variation with size if available).

Maybe I am missing something, but Table A28 discusses % units of coarse mode contributions to backscatter signal, which is not the same as % units AOD. I appreciate what you are trying to do here. But if your argument is applied to 18 June data, 40% of AOD can be redistributed at 360nm on 18 June (change from 27.3 to 68%). This is four times larger flexibility to redistribute extinction, and would be at odds with the primary message that I

take away from Tables A27 & A28, which is that 1) there is less of a need to extrapolate wavelength on 18 June than on 8 July, 2) aerosol profiles at 1020nm make a relatively larger contribution to control extinction also at 360nm on 18 June than 8 July, and 3) the fact that no correction factor is needed on 18 June.

It will help to have the data to play with it... thanks for you soon response.

Thanks,
-Rainer

Email from Thomas, 05.02.2019

Dear Rainer,

please find my response to your points below:

On 05.02.2019 01:30, Rainer Volkamer wrote:

> Hi Thomas,
> can you send us the McArtim files and O4 measurements for both days?
> We probably want to run some simulations here ourselves to have an effective conversation. McArtim3 is also what we usually use. Its a good exercise for Chris (cc here) to setup calculations for the 18.6. and 8.7 case study days and inform further discussions.
> Please also send the size distribution files from Aeronet, and any info on the complex refractive index (and its variation with size if available).

All input data of the first comparison round are available at the MADCAT web page, see http://joseba.mpch-mainz.mpg.de/Comparison.htm
Additional AERONET inversion data were provided in my email from 10 May 2017 (including you as an addressee). I will re-send this email again in the next minute. Please let me know if you need something else.
However, I don't want to wait for the results of these additional simulations for the paper to be submitted. Detailed comparison studies between different RTM were already performed and are an important part of the paper. Also many sensitivity studies covering a large variety of settings (including your recent suggestions) were performed and are an important part of the paper. You were always included in the respective emails, but I never got feedback from your group during the last two years. Additional RTM exercises will lead to further delays of the paper, but one can not expect significantly new findings.
The present study is not very conclusive, and the question about a scaling factor can not be answered. More future comparison studies will be needed to address this issue (as stated in the paper).
Thus I want to ask you if you (and Ivan and Ted) can agree to become co-authors of the paper in its current form. If not, I will mention your contributions to the paper in the acknowledgments. Please send me your feedback within the next days.

> Maybe I am missing something, but Table A28 discusses % units of coarse mode contributions to backscatter signal, which is not the same as % units AOD.

Yes, this is true. And this is the reason why the respective fractions are compared for both wavelengths. If the fractions were exactly the same, the ceilometer measurements at 1020 nm would be perfectly representative for the aerosol profile at 360 nm. If they were different by 100%, then from the ceilometer measurements at 1020 nm no information about that at 360 nm could be retrieved.

> I appreciate what you are trying to do here. But if your argument is applied to 18 June data, 40% of AOD can be redistributed at 360nm on 18 June (change from 27.3 to 68%). This is four times larger flexibility to redistribute extinction, and would be at odds with the primary message that I take away from Tables A27 & A28, which is that 1) there is less of a need to extrapolate wavelength on 18 June than on 8 July, 2) aerosol profiles at 1020nm make a relatively larger contribution to control extinction also at 360nm on 18 June than 8 July, and 3) the fact that no correction factor is needed on 18 June.

On 18 June, indeed the ceilometer measurement is much less representative for the aerosol extinction at 360 nm. This is a surprising finding, but by using the detailed AERONET inversion products, this it what comes out. The reason that no scaling factor is needed for 18 June is probably simply a coincidence.

Best regards,

Thomas

Email from Ivan, 05.02.2019

Dear Thomas,
I should check this email more often, I am sorry I have missed discussions about your analysis. I am still trying to catch up with everything. My very initial input was about what Rainer just suggested. From the very beginning I noticed that you use similar extinction values in the boundary layer for both days, even larger extinction on July 8, although it is clear that the surface mass loading is significantly lower on July 8. I also have notice that you use zero extinction above 6-7 km or so, but in reality there might be some extinction, maybe assuming 1x10-3 or so might be more realistic, maybe is not important?. I am still trying to understand why the correction factor is not needed in one day and needed on another day. I think either the state of the atmosphere is not well characterized yet when correction factor is needed or purely luck when the correction is not needed.
Thanks for all the hard work getting this manuscript out.
Best,
Ivan

Email from Rainer, 05.02.2019

Dear Thomas,

thanks for re-sending the link to the files. I understand your desire to wrap this up.

However, I still see a disconnect between the in-situ PM mass loadings and near surface extinction values. I see a strong motivation to digest this relevant information in form of a further sensitivity study. The settings are in the last column of the below table:

Alt_low	Alt_up	Ext_INTA	Ext_redist_10%	Ext_redist_30%
0	1.68	0.186	0.1523	0.085
1.68	4.9	0.116	0.116	0.116
4.9	7	0.035	0.0687	0.136

The issue is that PM10 and PM2.5 mass loadings near the surface in Fig. 1D are significantly (factor 2) lower on 8 July compared to 18 June, while the extinction profiles in Fig. 9 have very similar extinction near the surface. The additional sensitivity study is needed to make the interpretation of the overall dataset more coherent. At the same time, I expect an improvement in the dAMFs for all EAs, based on the results of the first new profile shape. This would be significant!

I was trying to help save you time by offering to involve Chris. But after giving it some thought, the above case is probably sufficient to finalize our thinking about this paper.

If you agree, please add the results into Table A31 and Fig A34, and also archive the AMFs and dAMFs for all data in Fig. A34 (all EAs) in form of a new Table (similar to Table A29). I liked to see the results before sending detailed comments, as I certainly see potential to "expect significantly new findings" from these RTM calculations.

I have a dental procedure tomorrow, but should be back online on Friday. If we see the revised manuscript with the above changes by then we should be able to send our detailed comments by early next week. Sound good?

Regards,
-Rainer

Email from Thomas, 06.02.2019

Dear Rainer,

the differences in the in situ pm measurements on both days are in fact an important point. However, since the ceilometer is blind below 180m it is difficult to make a direct connection between the ceilometer measurements and the in situ data. Here the AERONET inversion products become important. From the AERONET inversion products for 8 July it is found that 10% of the total integrated extinction (but not 30%) 'could' be redistributed.

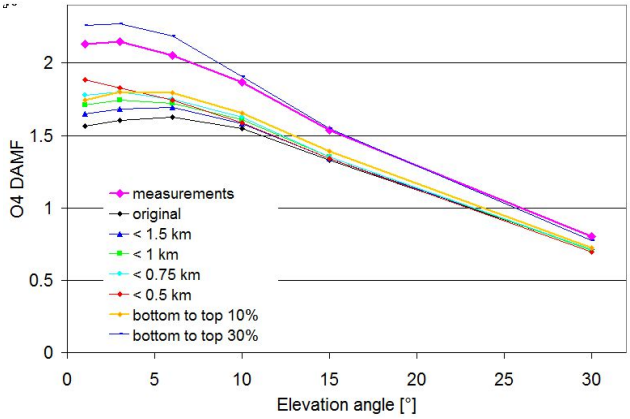
Your suggested case of a redistribution from the lowest layer to the upper layer of 30% can indeed bring measurements and simulations into closer agreement, see figure below (in fact a redistribution of about 28 % (not shown) would lead to an even better agreement. I think it makes sense to add this information to the paper.

Nevertheless, the AERONET inversion products don't support the assumption of a re-distribution of 30% of the total AOD from the lowest to the highest layer. Remember that only 18% of the total ceilometer signal is caused by the coarse mode aerosols.

Concerning the differences between both days, I think there is a simple explanation: since the ceilometer is blind below 180 m, it is very probable that on 18 June a much higher concentration of coarse aerosols exists below 180m. Note especially the large amount of pm10 on that day, which is unlikely to be lifted up to high altitudes. The assumption of large aerosol extinction below 180m on 18 June would also lead to an underestimation of the measured O4 (d)AMFs by the simulations (see Fig. A4 and Fig. A34). But this is of course a speculation and can not be further quantified based on the existing measurement data.
As stated above, for me it makes sense to include a) the results for the modified profile (relocation of 28%) to the paper, and b) mention the fact that such an assumption is not supported by the AERONET inversion results.
I hope you can agree to this procedure.
And don't forget, this paper will not explain the world. One important aim of the study is to provide guidelines to improve further comparison studies.

Best regards,

Thomas



Email from Rainer, 06.02.2019

Dear Thomas,

a potential temperature profile should be able to tell the height of the first inversion. A significant gradient below 200m would lead to a significant gradient in the measured O4 dAMFs between the 1 and 3EA, while in fact both angles show near identical dAMFs. I do not see any evidence to support a strong gradient below 180m, certainly not on 8 July (Fig 2), but also not really on 18 June. It is reasonably easy, and worth corroborating this point by calculating potential temperature profiles for both days.

I do not have an issue with the calculations shown in Appendix 5D. But I do not follow relevancy for redistribution of AOD at 360nm. Your argument equates a %signal contribution of the coarse mode with a %AOD redistribution. Why the focus on the coarse mode in the first place? Large particles are more likely near the surface (some may be aloft). Aloft its likely fine particles. If I was to take a guess on what size particles is responsible for "redistribution of AOD" its the fine particles. Quantifying detector signal from the coarse mode carries no information about how aerosols are distributed, or can be redistributed. In my opinion, Table A28 should be constructed from a perspective of fine particles, as they dominate optical properties at 360nm and AOD.

All that Appendix 5D is saying in my opinion, is that on either day small particles contribute to the ceilometer signal at 1064nm. And that the contribution from small particles to the AOD at 360nm is sufficient to justify redistribution of 30% AOD. In particular, the contribution of fine particles to the AOD at 360nm is 73% (18 June) and 92% (8 July) at 360nm (ignoring caveats from the lack of knowledge about wavelength dependent refractive index, questions about whether Mie theory applies, etc).

The wavelength extrapolation adds uncertainty to the profile shape at 360nm. This is significant, since the Mie resonances at 360nm are likely very different than at 1064nm. Meaning that a ceilometer measurement at 360nm would look very different. One would need to know the wavelength-, size- and altitude- resolved refractive index to relate an extinction profile from 1064nm to 360nm. In lack of that information, the profile shape measured at 1064nm is a crude guess on that at 360nm. I have to respectfully disagree that "From the AERONET inversion products for 8 July it is found that 10% of the total integrated extinction (but not 30%) 'could' be redistributed."

A minor point: All calculations in Appendix 5 should be done at 1064nm, with the AOD from 1020nm extrapolated to 1064nm. There is confusion also in other parts of the paper about the 1020/1064nm wavelengths pair (it certainly confused me at first). But profile information is extrapolated from 1064 to 360nm, not 1020 to 360nm. Not a biggie in the big picture, but there is some confusion here.

I wanted to get back to you before my dentist knocks me out for the rest of the day. And I look forward to seeing the potential temperature profiles, if available, and the revised manuscript. Ivan, Ted, Chris - feel free to add to this.

Cheers,
-Rainer

Email from Thomas, 06.02.2019

Dear Rainer,

please find my replies below.

On 06.02.2019 18:52, Rainer Volkamer wrote:

> Dear Thomas,
>
>
> a potential temperature profile should be able to tell the height of the first inversion. A significant gradient below 200m would lead to a significant gradient in the measured O4 dAMFs between the 1 and 3EA, while in fact both angles show near identical dAMFs. I do not see any evidence to support a strong gradient below 180m, certainly not on 8 July (Fig 2), but also not really on 18 June. It is reasonably easy, and worth corroborating this point by calculating potential temperature profiles for both days.

Unfortunately, there is no potential temperature profile data available. ECMWF data are at a rather coarse grid and are thus not representative for the local conditions.
On 18 June the dAMFs for 1° are smaller than for 3° and 6°. This indicates that high aerosol extinction is located close to the surface. But as mentioned yesterday, this is a speculation and can not further be quantified based on the available data.

> I do not have an issue with the calculations shown in Appendix 5D. But I do not follow relevancy for redistribution of AOD at 360nm. Your argument equates a %signal contribution of the coarse mode with a %AOD redistribution. Why the focus on the coarse mode in the first place? Large particles are more likely near the surface (some may be aloft). Aloft its likely fine particles. If I was to take a guess on what size particles is responsible for "redistribution of AOD" its the fine particles. Quantifying detector signal from the coarse mode carries no information about how aerosols are distributed, or can be redistributed. In my opinion, Table A28 should be constructed from a perspective of fine particles, as they dominate optical properties at 360nm and AOD.

Both perspectives (from coarse or fine mode aerosols) are equivalent. Taking the 'fine mode perspective', on 8 July the fine mode contributes 82% to 1064 nm and 92% to 360 nm. The difference of 10% (either 92% - 82% or 18 % - 8%) is what matters.

> All that Appendix 5D is saying in my opinion, is that on either day small particles contribute to the ceilometer signal at 1064nm. And that the contribution from small particles to the AOD at 360nm is sufficient to justify redistribution of 30% AOD. In particular, the contribution of fine particles to the AOD at 360nm is 73% (18 June) and 92% (8 July) at 360nm (ignoring caveats from the lack of knowledge about wavelength dependent refractive index, questions about whether Mie theory applies, etc).

Still, on 8 July the difference is 10%.

> The wavelength extrapolation adds uncertainty to the profile shape at 360nm. This is significant, since the Mie resonances at 360nm are likely very different

than at 1064nm. Meaning that a ceilometer measurement at 360nm would look very different. One would need to know the wavelength-, size- and altitude-resolved refractive index to relate an extinction profile from 1064nm to 360nm. In lack of that information, the profile shape measured at 1064nm is a crude guess on that at 360nm. I have to respectfully disagree that "From the AERONET inversion products for 8 July it is found that 10% of the total integrated extinction (but not 30%) 'could' be redistributed."

I still think my arguments are correct.
There is another indication that the '30% redistribution profile' does not fit to the measurements. I compared the measured diurnal variation of the O4 SCDs at 90° elevation to the simulations (see below). For the '30% redistribution profile' too much light is scattered from high altitudes leading to smaller O4 AMFs for high SZA.

A minor point: All calculations in Appendix 5 should be done at 1064nm, with the AOD from 1020nm extrapolated to 1064nm. There is confusion also in other parts of the paper about the 1020/1064nm wavelengths pair (it certainly confused me at first). But profile information is extrapolated from 1064 to 360nm, not 1020 to 360nm. Not a biggie in the big picture, but there is some confusion here.

This is correct, and I also discovered this slight inconsistency today. It will of course be corrected in the final version. However, it is a very small effect. The coarse mode contribution on 8 July changes from 18.1 to 18.3 %. On 18 June, it is completely negligible.

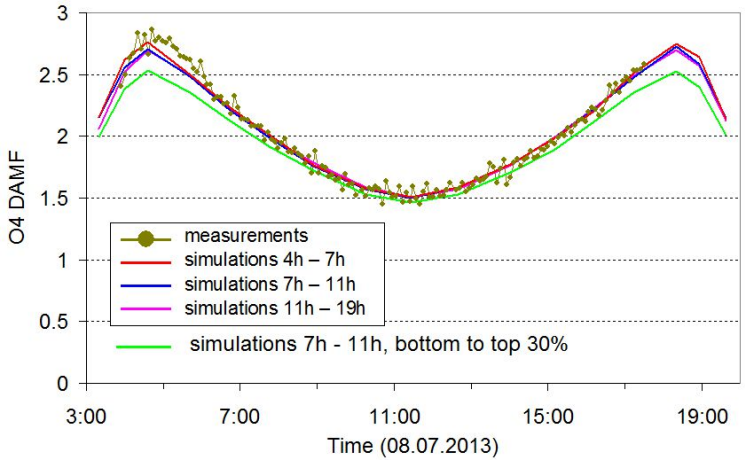
I wanted to get back to you before my dentist knocks me out for the rest of the day. And I look forward to seeing the potential temperature profiles, if available, and the revised manuscript. Ivan, Ted, Chris - feel free to add to this.

It seems to me that we will not come to an agreement about the aerosol profiles. Therefore I suggest that the editor and reviewers should decide. Of course your comments and my replies will be made available to them and also be made available for the public at the discussion site. If you still want to send me your detailed comments, they are still welcome.

Many thanks,

I hope the surgery at the dentist will not be too painful and you will recover soon.

Thomas



email from Ted, 07.02.2019

Hello Thomas,

I present my thinking at more length below, but skipping to the conclusions: the SZA dependence highlights the 30% redistribution is not ultimately fully consistent, and I further suspect that a fully consistent solution is not easily found. Rather, these sensitivity studies can be presented and framed to highlight that discrepancies between modeled and measured O4 can be explained by such changes which cannot be ruled out by available data. This points to the potential to leverage the angle specific O4 dAMFs and SZA dependent dAMFs in conjunction with certain assumptions to make adjustments and perhaps reach a fully consistent solution. That is beyond the scope of the paper. Still, I think the sensitivity studies can highlight that while such a exercise is challenging, at present poorly constrained, and perhaps impractical, it is not impossible.
I outline my thinking below, my apologies for the length:
As I understand the analysis in the manuscript, an angstom exponent was derived for a given point in time, and then applied to the entire ceiliometer profile. The extinction for monodisperse large particles is relatively flat with wavelength, whereas for smaller particles the extinction changes more rapidly with wavelength. However, the coarse fine dichotomy is not the only concern, see for instance Schuster et al., 2006. While the precise size distribution of coarse mode aerosol does not change the angstrom exponent, the specifics of particle distributions in the fine mode can act as a strong lever on the angstrom exponent.
The absolute contribution of fine mode aerosol provides some measure of the expected inaccuracy of adapting an extinction profile from a different wavelength. I don't think that the difference in the relative contribution of the aerosol modes to the different measurements is a relevant metric for this effect. Unfortunately I don't have a firm constraint to offer beyond the fact that it should be more important when more aerosol volume is in the fine mode.
In this context it is not surprising that June 18 has a correction factor closer to 1 than July 8, because overall the aerosol are larger and therefore a constant angstrom exponent with altitude is more likely to be closer to the truth. For July 8, while aerosol size distribution profiles in the atmosphere are complex they generally tend to get smaller and narrower rising through the troposphere. Both these effects increase the angstrom exponent and as such there is expected to be a general tendency that

when transferring an extinction profile from longer to shorter wavelengths that it will be relatively enhanced at higher altitudes. Atmospheric layering of course also plays a role.
The sensitivity studies in Appendix 6 are therefore consistent with expected effects in the absence of better constraints highlighting layers especially. I cannot offer a corollary to your ~10% bound, but I don't believe the ~30% effect can be completely rejected either. If another sensitivity study can further illustrate the principle while ignoring any layers, a naive smooth altitude dependence of scaling might serve i.e multiply the extinction profile by [(1-x) + (2x/7 km-1) * altitude], where x between between 0.1 and 0.3 should serve to illustrate.
Please let me know if there is anything compelling which I am overlooking. Perhaps we will not reach agreement. In any case, I would appreciate your thoughts on this perspective and framing, it would be useful in determining some of the specific comments. Thank you for bringing this extensive exercise together and for your responsiveness these last days.
Best Wishes,
Ted

Email from Thomas, 07.02.2019

Dear Ted,
many thanks for your feedback!
I want to clarify my general view: I think we all agree that there are uncertainties about the aerosol extinction profiles.
To decide which extinction profile might be the most probable, we can use the following information:
a) the ceilometer data and the AERONET inversion products
b) the elevation dependence
c) the SZA dependence
In addition to these observations, we can assume that coarse mode aerosol is probably located at lower altitudes than the fine mode aerosols.
All of these observations and assumptions have their uncertainties. Nevertheless, taking all information into account, I conclude that the scenario of a 10% redistribution is the most probable.
The results of the sensitivity studies for the different profiles and their compatibility with the above stated observations and assumptions will of course be provided in the paper.
Then not only the editor and the reviewers, but also the readers can reach their own conclusion on what they think is most probable.

I hope you can agree to that procedure.

Please find my response to the individual points below.

Best regards,

Thomas

On 07.02.2019 03:39, Theodore Konstantinos Koenig wrote:

> Hello Thomas,
>
> I present my thinking at more length below, but skipping to the conclusions: the SZA dependence highlights the 30% redistribution is not

ultimately fully consistent, and I further suspect that a fully consistent solution is not easily found. Rather, these sensitivity studies can be presented and framed to highlight that discrepancies between modeled and measured O4 can be explained by such changes which cannot be ruled out by available data. This points to the potential to leverage the angle specific O4 dAMFs and SZA dependent dAMFs in conjunction with certain assumptions to make adjustments and perhaps reach a fully consistent solution. That is beyond the scope of the paper. Still, I think the sensitivity studies can highlight that while such a exercise is challenging, at present poorly constrained, and perhaps impractical, it is not impossible.

see my general comments above

I outline my thinking below, my apologies for the length:

As I understand the analysis in the manuscript, an angstom exponent was derived for a given point in time, and then applied to the entire ceiliometer profile.

This is not exactly true. The angstrom exponent was determined for the selected period. Also it is not applied to the entire ceilometer profile. The altitude dependence of the size distribution (and thus the angstrom exponent) is implicitly accounted for by the re-distribution of 10% of the total extinction.

The extinction for monodisperse large particles is relatively flat with wavelength, whereas for smaller particles the extinction changes more rapidly with wavelength. However, the coarse fine dichotomy is not the only concern, see for instance Schuster et al., 2006. While the precise size distribution of coarse mode aerosol does not change the angstrom exponent, the specifics of particle distributions in the fine mode can act as a strong lever on the angstrom exponent.

Of course this is true. However, the wavelength dependencies are intrinsically taken into account by the use of the phase functions for fine and coarse mode aerosols derived from the AERONET inversion. This information is not perfect, but describes best the aerosol properties during that day.

The absolute contribution of fine mode aerosol provides some measure of the expected inaccuracy of adapting an extinction profile from a different wavelength. I don't think that the difference in the relative contribution of the aerosol modes to the different measurements is a relevant metric for this effect. Unfortunately I don't have a firm constraint to offer beyond the fact that it should be more important when more aerosol volume is in the fine mode.

Intuitively, I had the same expectations at the beginning. Nevertheless, by taking the optical depths and the phase functions of fine and coarse mode aerosols into account, it turns out that on 8 July even at the rather large wavelength of the ceilometer measurements the fine mode

dominates the ceilometer signal (82%). I think this is the key point and tells us that on 8 July the ceilometer measurements at 1064 nm are a very good proxy for the aerosol extinction profile shape at 360 nm.

> In this context it is not surprising that June 18 has a correction factor closer to 1 than July 8, because overall the aerosol are larger and therefore a constant angstrom exponent with altitude is more likely to be closer to the truth. For July 8, while aerosol size distribution profiles in the atmosphere are complex they generally tend to get smaller and narrower rising through the troposphere. Both these effects increase the angstrom exponent and as such there is expected to be a general tendency that when transferring an extinction profile from longer to shorter wavelengths that it will be relatively enhanced at higher altitudes. Atmospheric layering of course also plays a role.

Of course, I agree that in general the size distribution varies with altitude. This is what our whole discussion is about. But I think this is the case for both days: We should expect that the size of the aerosols in general decreases with altitude. The important difference is that on 8 July the relative contribution from the coarse mode to the ceilometer signal is much larger than on 18 June which complicates the quantitative interpretation.

> The sensitivity studies in Appendix 6 are therefore consistent with expected effects in the absence of better constraints highlighting layers especially. I cannot offer a corollary to your ~10% bound, but I don't believe the ~30% effect can be completely rejected either. If another sensitivity study can further illustrate the principle while ignoring any layers, a naive smooth altitude dependence of scaling might serve i.e multiply the extinction profile by [(1-x) + (2x/7 km-1) * altitude], where x between between 0.1 and 0.3 should serve to illustrate.

Initially, I also had this thought. Such a smooth altitude dependence is surely more realistic than a re-distribution between layers. However, I decided to use the more extreme re-distributions between layers for the sensitivity studies because of two reasons:
a) we have no information on the altitude dependence of the fine and coarse mode fractions. All assumed re-distributions are simply assumptions (of course with some plausibility)
b) from the extreme scenarios the overall magnitude of the effect can be estimated, and that is what matters.

> Please let me know if there is anything compelling which I am overlooking. Perhaps we will not reach agreement. In any case, I would appreciate your thoughts on this perspective and framing, it would be useful in determining some of the specific comments. Thank you for bringing this extensive exercise together and for your responsiveness these last days.

My current plan is to prepare an updated version of the manuscript in the next two days and send it to you. If your detailed feedback contains further fundamental points, it would be good to know these points before I prepare the updated version. I want to avoid too many iterations.

Many thanks,

Thomas

Email from Rainer, 08.02.2019

On 2/7/2019 5:52 PM, Rainer Volkamer wrote:

> Dear Thomas,
>
> sorry for another lengthy email. I have added my comments below your initial text to Ted, as well as below your responses to Teds comments:
>
> On 2/7/2019 5:33 AM, Thomas Wagner wrote:
>
> > Dear Ted,
> > many thanks for your feedback!
> > I want to clarify my general view: I think we all agree that there are uncertainties about the aerosol extinction profiles.
> > To decide which extinction profile might be the most probable, we can use the following information:
> > a) the ceilometer data and the AERONET inversion products
> > b) the elevation dependence
> > c) the SZA dependence
>
> Missing here is d) near surface PM levels on 18 June are significantly (factor ~2 times?) higher than on 8 July. To reproduce this gradient in surface PM between both days it is necessary to redistribute 30% of the AOD from lower to higher altitudes. I further elaborate on synergies between b and d to inform this below.
>
> In absence of potential temperature profiles, information from b) is helpful to assess whether the near surface PM is expected to be highly localized near the surface, or is indeed representative also at altitudes above ~200m. I elaborate below.
>
> During MADCAT, the effective pathlength of photons at 350nm has been quantified as 7km in the lower angles (Ortega et al. 2015; doi:10.5194/amt-8-2371-2015). This distance corresponds to an altitude of 120m for EA1, 367m for EA3, and 735m for EA6 for the effective last scattering event. The sensitivity studies shown in Fig. A33 of Wagner et al. reveal that O4 dAMFs should be sensitive to assess aerosol gradients over these altitudes. In particular, if there is a sharp gradient at 500m, the O4 dAMFs for EA1 are systematically larger than at EA3 (consistent with the tail-shape of the box-AMFs expected for these EAs, and the above altitudes for the last scattering event). However, no such behavior is observed in the measurements. In fact, the measured O4 dAMFs slightly decrease from EA3 to EA1. And this shape in EA splits of the O4 dAMFs is very well reproduced based on the ceilometer shape information (consistent with 82% of the ceilometer signal actually

originating from the fine mode also at 1064nm). The EA split among measured O4 dAMFs is well reproduced at all EAs by RTM.

This is only consistent with 1) the absence of sharp gradients below 500m, and suggests the PM gradients measured near the surface are in fact a good proxy. The gradient in surface PM strongly suggests a gradient in the surface extinction of a factor of 2 is expected between both days (compare Fig.1 and Fig. 9). This is only achieved if 2) 28% of the lower AOD are redistributed to higher altitudes. Any lower number would overestimate surface extinction. Note that surface extinction is probably the only place where MAX-DOAS can constrain altitude resolved extinction well. Finally, 3) if 28% of AOD are redistributed the measured and predicted O4 dAMFs values agree quantitatively at all EAs.

The information from b and d combined thus provide strong experimental evidence in support of the hypothesis that uncertain aerosol vertical profiles are the primary cause for the correction factor on 8 July.

I consider this evidence as fully consistent also with the information provided in Appendix 5D, which shows that the fine mode aerosol is responsible for the major share of the signal detected by the ceilometer on 8 July, but not on 18 June.

Interestingly, the inferred profiles vary much more strongly on 8 July, and do not vary much on 18 June, possibly providing an important clue on what is driving the different behavior between both case study days (by affecting the initialization of RTM). In this context, I liked to point out that in Ortega et al. 2016 uncertainty due to wavelength scaling of the aerosol extinction profile is minimized (airborne HSRL_532nm was compared with O4 at 477 and 360nm in both Ortega et al. 2016 and Volkamer et al. 2015). Generally speaking, no correction factor was needed if information about aerosols aloft was well characterized in our previous work where HSRL was available. As you know, HSRL overcomes the fundamental limitation of characterizing sub-Rayleigh aerosol by measuring Rayleigh back-scatter directly, which greatly enhances the aerosol contrast in air where aerosol extinction becomes sub-Rayleigh. Sub-Rayleigh aerosol extinction becomes an issue with interpreting ceilometer data, which require to define a "zero" aerosol aloft to decouple aerosols. There is a fundamental limitation in that the ceilometer cannot measure sub-Rayleigh aerosol. In our aircraft campaigns comparing HSRL and AMAX-O4 inferred aerosol extinction, both sensors find aerosols typically become sub-Rayleigh at altitudes above 4-6km (compare e.g., Fig. 3 in Volkamer et al. 2015). This behavior we have observed over continents, and over oceans, and it is further generally also the altitude range where the ceilometer profiles during MADCAT are close or below the Rayleigh extinction (you could calculate the Rayleigh extinction line, and add it into Fig. 9). This is probably the reason why the extinction profile shape extracted from identical ceilometer data by different groups varies so much at altitude (Fig. 9).

In summary, I see no information in this paper that would not be compatible with the explanation presented in Ortega et al. 2016. And in fact, your paper makes an important contribution in that it helps establish that uncertain aerosol

profiles aloft have probably a larger uncertainty than has previously been recognized.

> In addition to these observations, we can assume that coarse mode aerosol is probably located at lower altitudes than the fine mode aerosols.
> All of these observations and assumptions have their uncertainties. Nevertheless, taking all information into account, I conclude that the scenario of a 10% redistribution is the most probable.

Taking also the information from d) into account, a larger redistribution is justified. See above.

> The results of the sensitivity studies for the different profiles and their compatibility with the above stated observations and assumptions will of course be provided in the paper.

I agree.

> Then not only the editor and the reviewers, but also the readers can reach their own conclusion on what they think is most probable.
>
> I hope you can agree to that procedure.

I would strongly advise against an approach that involves the Editor. Let alone you are the Editor in Chief of the Journal, and this could open all kinds of worms... I see no reason not to resolve this before submission. Its mostly language really, as I see it. And input from the co-authors could also be helpful.

> Please find my response to the individual points below.

I am adding some short responses below as well.

> Best regards,
>
> Thomas

> On 07.02.2019 03:39, Theodore Konstantinos Koenig wrote:

Hello Thomas,

I present my thinking at more length below, but skipping to the conclusions: the SZA dependence highlights the 30% redistribution is not ultimately fully consistent, and I further suspect that a fully consistent solution is not easily found. Rather, these sensitivity studies can be presented and framed to highlight that discrepancies between modeled and measured O4 can be explained by such changes which cannot be ruled out by available data. This points to the potential to leverage the angle specific O4 dAMFs and SZA dependent dAMFs in conjunction with certain assumptions to make adjustments and perhaps reach a fully consistent solution. That is beyond the scope of the paper. Still, I think the sensitivity studies can highlight that while such a exercise is challenging, at present poorly constrained, and perhaps impractical, it is not impossible.

see my general comments above

I outline my thinking below, my apologies for the length:

As I understand the analysis in the manuscript, an angstom exponent was derived for a given point in time, and then applied to the entire ceiliometer profile.

This is not exactly true. The angstrom exponent was determined for the selected period. Also it is not applied to the entire ceilometer profile. The altitude dependence of the size distribution (and thus the angstrom exponent) is implicitly accounted for by the re-distribution of 10% of the total extinction.

It is not possible in my opinion to recover information at 360nm accurately from measurements at 1064nm. We also have more direct evidence that

supports a larger re-distribution of AOD.

> The extinction for monodisperse large particles is relatively flat with wavelength, whereas for smaller particles the extinction changes more rapidly with wavelength. However, the coarse fine dichotomy is not the only concern, see for instance Schuster et al., 2006. While the precise size distribution of coarse mode aerosol does not change the angstrom exponent, the specifics of particle distributions in the fine mode can act as a strong lever on the angstrom exponent.

Of course this is true. However, the wavelength dependencies are intrinsically taken into account by the use of the phase functions for fine and coarse mode aerosols derived from the AERONET inversion. This information is not perfect, but describes best the aerosol properties during that day.

I agree that you have done what can be done, Thomas. But a case with high wavelength dependence (8 July) should result in a larger uncertainty due to wavelength scaling than a case with a lower wavelength dependence (18 June). I think nobody would argue that a measurement at 360nm would be more valuable to inform 360nm than a measurement at 1064nm -- but it seems to me that your argument in Appendix 5D can be misunderstood that way. We can agree to disagree here.

> The absolute contribution of fine mode aerosol provides some measure of the expected inaccuracy of adapting an extinction profile from a different wavelength. I don't think that the difference in the relative contribution of the aerosol modes to the different measurements is a relevant metric for this effect. Unfortunately I don't have a firm constraint to offer beyond the fact that it should be more important when more aerosol volume is in the fine mode.

Intuitively, I had the same expectations at the beginning. Nevertheless, by taking the optical depths and the phase functions of fine and coarse mode aerosols into account, it turns out that on 8 July even at the rather large wavelength of the ceilometer measurements the fine mode dominates the ceilometer signal (82%). I think this is the key point and tells us that on 8 July the ceilometer measurements at 1064 nm are a

> very good proxy for the aerosol extinction profile shape at 360 nm.

I agree - and had made a similar point in my email in suggesting to construct Table A28 from a perspective of the fine mode. Note that Mie resonances of fine mode particles happen at the wavelengths around the O4 observations. They do not happen at the wavelengths where the ceilometer strongly interacts. The ceilometer thus does not constrain the Q (extinction enhancement) of fine mode particles well, even though it is sensitive to fine aerosols. The exact wavelength and magnitude of Q depends on the refractive index and many other parameters (see earlier email), which strongly vary with wavelength. And all of this introduces uncertainty that goes well beyond the scope of this paper.

Its your paper, Thomas, but I strongly advise against putting too much faith into the calculations in Appendix 5D.

> In this context it is not surprising that June 18 has a correction factor closer to 1 than July 8, because overall the aerosol are larger and therefore a constant angstrom exponent with altitude is more likely to be closer to the truth. For July 8, while aerosol size distribution profiles in the atmosphere are complex they generally tend to get smaller and narrower rising through the troposphere. Both these effects increase the angstrom exponent and as such there is expected to be a general tendency that when transferring an extinction profile from longer to shorter wavelengths that it will be relatively enhanced at higher altitudes. Atmospheric layering of course also plays a role.

> Of course, I agree that in general the size distribution varies with altitude. This is what our whole discussion is about. But I think this is the case for both days: We should expect that the size of the aerosols in general decreases with altitude. The important difference is that on 8 July the relative contribution from the coarse mode to the ceilometer signal is much larger than on 18 June which complicates the quantitative interpretation.

See above. I agree its complicated.

> The sensitivity studies in Appendix 6 are therefore consistent with expected effects in the absence of better constraints highlighting layers especially. I cannot offer a corollary to your ~10% bound, but I don't

> believe the ~30% effect can be completely rejected either. If another sensitivity study can further illustrate the principle while ignoring any layers, a naive smooth altitude dependence of scaling might serve i.e multiply the extinction profile by [(1-x) + (2x/7 km-1) * altitude], where x between between 0.1 and 0.3 should serve to illustrate.

Initially, I also had this thought. Such a smooth altitude dependence is surely more realistic than a re-distribution between layers. However, I decided to use the more extreme re-distributions between layers for the sensitivity studies because of two reasons:
a) we have no information on the altitude dependence of the fine and coarse mode fractions. All assumed re-distributions are simply assumptions (of course with some plausibility)
b) from the extreme scenarios the overall magnitude of the effect can be estimated, and that is what matters.

I agree with all that is said here. But I do think the combination of b) and d) above holds new merit that should be considered. It supports redistribution out of the surface layer. Since the ceilometer is sensitive mostly to fine particles, and faces the fundamental limitation of loosing sensitivity for sub-Rayleigh aerosols, a redistribution into the higher aerosol layer is plausible.

Note that we did not optimize elevated layers using information from a) yet. There would be lots of room to optimize this distribution, and i.e. elevated layers, based on the SZA dependence in future work. I think this is worth pointing out in the section on recommendations in the revised manuscript.

> Please let me know if there is anything compelling which I am overlooking. Perhaps we will not reach agreement. In any case, I would appreciate your thoughts on this perspective and framing, it would be useful in determining some of the specific comments. Thank you for bringing this extensive exercise together and for your responsiveness these last days.

My current plan is to prepare an updated version of the manuscript in the next two days and send it to you. If your detailed feedback contains further fundamental points, it would be good to know these points before I prepare the updated version. I want to avoid too many iterations.

I liked to resonate Ted comments, and thank you for your responsiveness, and your patience.

Its a massive piece of work, with many interlocking pieces. Its at present also a very complicated paper to read. I am hoping that our discussions, albeit lengthy at times, are helpful, and can be used to simplify the paper. I look forward to seeing the revised version.

-Rainer

email from Thomas, 11.02.2019

Dear Rainer, dear all,

please find attached the updated version of the paper. The changes compared to the previous version are in sections 4.2.1, 5.2, and appendices A5 and A6.
Please let me know if you can agree to this version. Then I will send it around to the other co authors.

Concerning the last email from Rainer, I don't want to respond to each individual point, because the communication is already quite complicated. Below I give my feedback to the points which - in my view - are the most important ones:

a) what can we learn from the in situ measurements? In my opinion we can use them only in a qualitative way for the comparison between both days (as already discussed in the paper). But it is not possible to make a direct quantitative link between the in situ measurements and the ceilometer profiles, because different quantities are measured (backscatter signal versus aerosol mass concentration).

b) Rainer states that he agrees that we disagree. I also agree to that. Overall, the input from your group has led to large improvements of the paper, especially with respect to the shape of the aerosol extinction profile and its uncertainties, and to the recommendations for future comparison exercises. I hope that you find the discussion of the aerosol profiles and their uncertainties in the revised version acceptable for you.

c) Rainer suggests that the discussion between him/his group and me should not be shared with the editor. I must say that I strongly disagree. It is one important feature of AMT that important discussions have to be made available to the editor, to the reviewers, and also on the discussion web page.

Best regards,

Thomas

The attached pdf file is: O4_scaling_factor_10022019.pdf

email from Ivan, 11.0.2019

Dear Thomas,

Thanks for sending a revised version. It was hard to track down everything based on emails. I included a few comments using the annotation tools in Adobe Reader (see attachment). Below are some general comments:
- The appendix is quite long. Sometimes I had to look for key information in the appendix when, in my opinion, it should be included in the main text. Especially, regarding how the aerosol extinction profile at 360nm was derived. I think all assumptions should be included clearly in the main text instead of directing the reader to the appendix several times in a single paragraph.
- In the abstract you mention:
"*One important recommendation for future studies is that aerosol profile data should be measured at the same wavelengths as the MAX-DOAS measurements."*
Similarly in the conclusion:

"one important quality of the aerosol data sets is crucial to constrain the radiative transfer simulations. For example, it is
recommended that LIDAR instruments are operated at wavelengths close to those of the MAX-DOAS".
I fully agree. Note that in Ortega et al. (2016) this approach was already used. In that study, we used highly resolved independent extinction profiles measured at 355, 532, and 1064 nm from HSRL, i.e., no assumptions about construction of extinction profiles. However, detail information in that regard is missing in the manuscript. Ortega et al. (2016) and Volkamer et al are mentioned in the manuscript but it is not recognized the approach used and the key HSRL products used.
- As far as I can tell from the manuscript the only parameter that brings SF to unity is if aerosol extinction aloft is included, is that correct?. Maybe I am missing another factor that brings the SF to unity?. Leaving behind assumptions and whether this is true or not I would mention parameters that bring SF to unity in the abstract/conclusions and of course that more measurements are needed, as you already mentioned. I am mentioning this because still you mention in the abstract that *"Besides the inconsistent comparison results for both days, also no explanation for a O4 scaling factor could be derived in this study"* . It is hard to reconcile this though. It is clear that the SF is not explained by O4 MAX-DOAS measurements, but it has to be something in the state of the atmosphere causing the need of SF. In Ortega et al (2016) we concluded that independent highly-resolved profiles were needed and elevated aerosol layers were identified and if not accounted for the SF < 1 is needed. I am not saying this is the case always, but it has to be something in the state of the atmosphere causing this the forward model.
- Following up above, elevated aerosol layers are really more frequent that we thought, see Berg et al. (2015) and references therein.
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015JD023848
- I am not really sure but have you seen if using CALIPSO extinction profiles might help?. Again, I am not sure if there is an overpass or if they measure in boundary layer and even if they can be used we might have the same issues.
Thanks for all this important work and I apologize for not sending comments before.
Greeting to you & your group,
Ivan

The attached pdf file is O4_scaling_factor_10022019_io.pdf

Email from Thomas, 12.02.2019

Dear Ivan,

many thanks for your feedback! Please find my replies below.

Please let me know until 13 February if you agree to be co-author of the paper in the current form (including the changes described below). I have to send the updated version to all other co-authors to receive their feedback before I submit the revised version of the paper.

Rainer, Ted, please also let me know until 13 February if you agree to be co-author of the paper.

Many thanks,

Thomas

On 11.02.2019 20:51, Ivan Ortega wrote:

> Dear Thomas,
>
>
> Thanks for sending a revised version. It was hard to track down everything based on emails. I included a few comments using the annotation tools in Adobe Reader (see attachment). Below are some general comments:
>
> - The appendix is quite long. Sometimes I had to look for key information in the appendix when, in my opinion, it should be included in the main text. Especially, regarding how the aerosol extinction profile at 360nm was derived. I think all assumptions should be included clearly in the main text instead of directing the reader to the appendix several times in a single paragraph.

I would prefer to leave the structure as it is. I understand your concern, but there is so much information in the paper that a lot of details have to be put to the appendix. Nevertheless, in the main text it is clearly stated how the details can be found.

> - In the abstract you mention:
>
> "*One important recommendation for future studies is that aerosol profile data should be measured at the same wavelengths as the MAX-DOAS measurements.*"
>
> Similarly in the conclusion:
>
> *"one important quality of the aerosol data sets is crucial to constrain the radiative transfer simulations. For example, it is recommended that LIDAR instruments are operated at wavelengths close to those of the MAX-DOAS".*

I fully agree. Note that in Ortega et al. (2016) this approach was already used. In that study, we used highly resolved independent extinction profiles measured at 355, 532, and 1064 nm from HSRL, i.e., no assumptions about construction of extinction profiles. However, detail information in that regard is missing in the manuscript. Ortega et al. (2016) and Volkamer et al are mentioned in the manuscript but it is not recognized the approach used and the key HSRL products used.

In the parts of the text, where it is stated that it is important to use LIDAR measurements at the same wavelength, the reference to Ortega et al., 2016 was added.

- As far as I can tell from the manuscript the only parameter that brings SF to unity is if aerosol extinction aloft is included, is that correct?.

No, that is not correct, see section 5.2. There several potential reasons for the discrepancies are listed.

Maybe I am missing another factor that brings the SF to unity?. Leaving behind assumptions and whether this is true or not I would mention parameters that bring SF to unity in the abstract/conclusions and of course that more measurements are needed, as you already mentioned. I am mentioning this because still you mention in the abstract that *"Besides the inconsistent comparison results for both days, also no explanation for a O4 scaling factor could be derived in this study"* . It is hard to reconcile this though. It is clear that the SF is not explained by O4 MAX-DOAS measurements, but it has to be something in the state of the atmosphere causing the need of SF.

I disagree here. It is not clear that the reason has to be something in the atmosphere. Also high levels of instrument straylight or wrong O4 cross sections could explain the differences, see section 5.2.

In Ortega et al (2016) we concluded that independent highly-resolved profiles were needed and elevated aerosol layers were identified and if not accounted for the SF < 1 is needed. I am not saying this is the case always, but it has to be something in the state of the atmosphere causing this the forward model.

I think it is not clear that it has to be something in the atmosphere, see comment above.

- Following up above, elevated aerosol layers are really more frequent that we thought, see Berg et al. (2015) and references therein.
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015JD023848

This might be the case, and it is indeed an interesting finding. But I don't see the relevance for this study. Here two days were selected, and all relevant available information is considered.

- I am not really sure but have you seen if using CALIPSO extinction profiles might help?. Again, I am not sure if there is an overpass or if they measure in boundary layer and even if they can be used we might have the same issues.

Unfortunately, Mainz is not seen by CALIOP

Thanks for all this important work and I apologize for not sending comments before.

Greeting to you & your group,

Ivan

Please find below my replies to the individual comments in the pdf. There have been a few comments without text. Maybe my pdf reader has problems here. Please let me know if I missed something important.

Comment 1: This still reads as if only direct sun observations found no need of correction factor. I would change by: "However, many studies came to opposite conclusion, that there is no need for a scaling factor"

My intention was to mention that even direct sun light measurements came to that conclusions, because such measurements are not affected by AMF uncertainties. I thus still think it would make sense to keep the formulation as it is.

Comment 2: Ortega et al. (2016) already presented a study where MAX-DOAS O4 and aerosol extinction were measured at the same wavelength. This important description is missing in the current manuscript.

The reference to Ortega et al. 2016 is given later at several parts in the text. References should be avoided in the abstract. So I prefer not add a reference to Ortega et al., 2016 there.

Comment 3: After this paragraph. I also suggest to include a short description of the methodology followed by studies where SF is unity, i.e., Volkamer et al (20) and Ortega et al. (2016) used independent highly resolved extinction profiles. in Ortega et al. (2016) aerosol extinction was measured at 355, 532, and 1064 nm.

I don't see the need to add such descriptions here.

Comment 4: it is strange to see ? in this equation

The question mark indicates the question whether the expected equality is true.

Comment 5: A description of how this was concluded is missing. Is it based on the 1064 nm ceilometer signal?

The description is given in appendix A5. I see no need to add more information here.

Comment 6: It may be too late to re-arrange, but it would have been nice to read first how the atmospheric conditions were derived before reading about the need of SF.

I see no need to add this information here. What would be gained from it? All relevant detailed information is given later.

Comment 7: Including wavelength of the ceilometer and AERONET is important but missing here. Also, assumptions about extrapolating ceilometer to extinction profiles are 360nm is missing

Please note that the paragraph starts with 'In short, the ceilometer measurements....' This indicates that only the basic principle is described here. In the next sentences the link to appendix A5 is given, where all further details are provided.

Comment 8: Not sure why is set to zero?. Realistically, the aerosol extinction would not be zero.

The reason is stated in the remainder of the sentence: '...because of the further increasing scatter and the usually small extinctions.'

Comment 9: Maybe I am missing an explanation but I don't see the value of comparing extracted exticntion profiles from different groups if all of them are constructed the same way, i.e., scaled by AOD and shape of the ceilometer at 1064nm. I would think extracted exticntion profiles using different methods, based on current independent measurements would be better.

The value of this comparison is to investigate the effects of different procedures. This is important information because not only the fundamental assumptions matter, but also the details of the extraction.

Comment 10: How was this derived?. Instead of showing key information in the appendix, I suggest to include it here.

I prefer to leave the structure as it is, because the derived results matter in the main text. The details for the interested readers are given in the appendix.

Comment 11: Again, I think mentioning that highly resolved and independent extinction profiles were measured in Ortega et al (2016), without assumptions about wavelength dependency is missing

Here the point is the altitude range. I don't see why information on the wavelength is this important here.

Comment 12: Again, Ortega et al. (2016) already use this approach. HSRL measured extinction profiles at 355, 532, and 1064 nm.

The reference to Ortega et al., 2016 was added

Comment 12: Ortega et al. (2016) already use this approach.

The reference to Ortega et al., 2016 was added

Comment 13: Again, one key aspect of Ortega et al. (2016) is that they use HSRL extinction profiles at 355, 532, 1064 nm products.

This information is added.

email from Rainer, 14.02.2019

Dear Thomas,

find attached the comments from Ted and me combined into a single file. Some short replies to your latest summary is below.

On 2/10/2019 4:00 PM, Thomas Wagner wrote:

> Dear Rainer, dear all,
>
> please find attached the updated version of the paper. The changes compared to the previous version are in sections 4.2.1, 5.2, and appendices A5 and A6. Please let me know if you can agree to this version. Then I will send it around to the other co authors.

I am fine with your changes in Sections 4.2.1, 5.2, and made some additions to reflect what I found was missing. I also added comments in section 4.3.5 "Effect of the temperature dependence of the O4 cross section". Please take a look, and let us know you are on board with the suggested changes.

> Concerning the last email from Rainer, I don't want to respond to each individual point, because the communication is already quite complicated. Below I give my feedback to the points which - in my view - are the most important ones:
>
> a) what can we learn from the in situ measurements? In my opinion we can use them only in a qualitative way for the comparison between both days (as already discussed in the paper). But it is not possible to make a direct quantitative link between the in situ measurements and the ceilometer profiles, because different quantities are measured (backscatter signal versus aerosol mass concentration).

Surface PM scales as volume, and so does surface extinction. So I think my argument carries merit. It is true that in-situ / column comparisons are always complicated, but this is probably partially mitigated if temporal averages are compared in a relative sense (as I did). But ok to frame this as a qualitative argument (as done in Section 5.2)

> b) Rainer states that he agrees that we disagree. I also agree to that. Overall, the input from your group has led to large improvements of the paper,

especially with respect to the shape of the aerosol extinction profile and its uncertainties, and to the recommendations for future comparison exercises. I hope that you find the discussion of the aerosol profiles and their uncertainties in the revised version acceptable for you.

I am glad you feel that way. It was an interesting and somewhat open ended discussion. I see nothing in this paper that contradicts our own earlier work on the topic (incl. Thalman and Volkamer, 2010; Thalman and Volkamer, 2013; Spinei et al., 2015; Volkamer et al., 2015; Ortega et al. 2016). Several of these papers include data from collocated airborne multi-wavelength HSRL, near surface extinction all the way to the surface, and comparisons at multiple O4 wavelengths. That is not a trivial statement.
I am fine to be a co-author (with the attached changes).

c) Rainer suggests that the discussion between him/his group and me should not be shared with the editor. I must say that I strongly disagree. It is one important feature of AMT that important discussions have to be made available to the editor, to the reviewers, and also on the discussion web page.

This is not what I said. You are misrepresenting my email.
My point is that differences are best sorted among all co-authors first, and ideally are reflected in the paper.
Feel free to use for following text for the purposes of circulating to co-authors (I'd need to think a bit more if I was to write for a permanent archive such as AMTD):
The remaining disagreement ranks around the importance of lacking vertically resolved aerosol properties, and uncertainties specific to MADCAT. Boulder and MPI agree that Appendix A5 provides an interesting and useful semi-quantitative argument about the origin of ceilometer signal at 1064nm. MPI claims a quantitative and low uncertainty for inferred aerosol vertically resolved information at 360nm from AERONET measured column properties. Boulder notes that the uncertainty due to wavelength scaling needs to hold for both days, and the argument in Appendix A5 implies a four times smaller uncertainty on a day when AOD varies much more strongly with wavelength (8 July), than on a day when AOD varies weakly with wavelength (18 June). Boulder further notes that less than 3 times higher uncertainty holds potential to explain the correction factor quantitatively on 8 July for all EAs in form of elevated aerosol layers. The resulting aerosol distribution has not been further optimized for SZA effects. Boulder notes that 60-70% AOD to reside above 1km is fully consistent with previous observations during TCAP, where multi wavelength airborne HSRL measurements constrain aerosol extinction all the way to the surface, and elevated aerosol layers are key to providing closure on O4 (Ortega et al., 2016). Boulder further points out that the low uncertainty estimate provided in Appendix A5 does not resolve an inconsistency that still exists between the relative abundance of near surface PM (lower on 8 July), and the near surface aerosol extinction (roughly constant) between both days, and that the factor of 3 higher uncertainty holds potential to resolve this inconsistency. An attempt has been made to reflect this discussion in the revised Section 5.2.
Of course I would also be happy to post a public comment to this effect in the AMT discussion forum, if needed.
I have made an attempt to add suggestions in the file also on other points that I did not find in your above summary.
Best regards,
-Rainer

The attached pdf file is O4_scaling_factor_10022019_TKK_RMV-1.pdf

Individual replies to the comments from Rainer and Ted in the pdf (O4_scaling_factor_10022019_TKK_RMV-1.pdf).

Comment from Ted, page 1:
Name removed, because it is not expected that Ted and the others agree to be co-author of the current version. They will be asked again if they want to be co-author after feedback from all other co-authors has become available.
Comment from Rainer, page 1: see above
Comment from Ted, page 1: see above
Comment from Ted, page 2: This is 0.81 below.
Corrected (further corrected tp 0.82, because slightly wrong factor for the profile '0.75km' was used.
Comment from Rainer, page 2: do you mean "should be collected"?
The text was changed to 'should be collected and used'
Comment from Ted, page 3: For our internal discussions we have found the form of dSCD/VCD for dAMF most simple to consider. Could it perhaps be added as further equality?
In principle, this could be added. But I think, the equation in its current form is more consistent.
Comment from Ted, page 7: 'n' was added
Comment from Ted, page 8: 'to' deleted
Comment from Rainer, page 12: Add: These deviations are lower than during the case study days in Ortega et al. (2016), where deviations between observed and calculated O4 profiles in the U.S standard atmosphere were found to be 13-18%.
Why should this be added here? It is later discussed that of course the deviations can be stronger for different locations and seasons. There the reference to Ivan's paper was added.
Comment from Rainer, page 12: the suggested sentence was added: 'This assumption reflects a practical limitation of the ceilometer likely responsible for the larger variability in the profile shape aloft by different groups.'
Comment from Ted, page 12: The sentence was changed to 'This assumption reflects a practical limitation of the ceilometer likely responsible for the larger variability in the profile shape aloft by different groups.'
Comment from Ted, page 12: The suggested sentence was added: 'This effect is further examined in Appendix A6'
Comment from Ted, page 13: I do not agree with this conclusion. Because Appendix A5 leverages the AERONET size distributions and AOD which are column properties I do agree that the column properties of the fine mode aerosol are well represented. I do not think a statement can be made regarding the profile. Can we present the information without this specific statement?
I still think the statement is correct.
Comment from Rainer, page 13: I disagree with this statement. While backscatter signal at 1064nm originates largely from fine mode aerosoll, there is no profile shape information at 360nm that can be recovered from column properties.
I think this statement is correct. This is one important point where we disagree.
Comment from Ted, page 13: I would rather consider the 10% and 30% redistribution from lowest layer to highest layer cases quoted in the main text. Understanding that we disagree as to whether the latter is reasonable, would the following language be acceptable?
"... subdivided into 3 layers (0-1.7 km; 1.7-4.9 km; 4.9 - 7 km), and extinction was redistributed from the lowest layer to the highest layer. It was found that redistributing 30% of total AOD this way increased the O4 dAMFs by 25%. However, such redistribution cannot be specifically justified."

I use 25% here as I don't have a more accurate number.
I think the case with the 25% or 30% redistribution fits best to section 5.2
Comment from Rainer, page 13: , where extinction is less well constrained (i.e., assumptions of zero aerosol extinction give rise to significant variability above 5km on 8 July, compare Fig. 9).
I think it makes no sense to add this statement here. The uncertainty in the study of Ortega et al., 2016 is probably even larger, because the maximum altitude of the profiles was even lower.
Comment from Ted, page 14: appendix is now consistently written in lower case.
Comment from Ted, page 14: Might be worth referencing the primary viewing direction briefly, since the sensitivity to the phase function is in part a function of the prevailing solar relative azimuth angle.
I think it is not necessary to add this information here. The azimuth angle was provided in the general description of the measurements. Also, the sacttering angle also depends on the solar zenith angle.
Comment from Rainer page 17: corrected
Comment from Rainer, page 18: since simulated spectra have access to complete information, the larger difference in synthetic data cannot be a problem with the cross-sections.
The synthetic spectra used the cross sections for all temperatures. However, they show a slight inconsistency as function of temperature. Thus in the syntehtic spectra the temperature dependence is not as smooth as it (very probably) should be. The temperature range of this inconsistency (~210 – 265 K) corresponds to a large atmospheric altitude range (~5 – 10 km). The non smooth temperature dependence of the O_4 cross sections are likely to explain the observed inconsistent fit results, because the fit results largely improve if only one O_4 absorption band is used in the fit.
It must be a problem introduced by the noise, which I read somewhere is larger than noise in the measurements.
Actually, here results for spectra without noise were shown. This information was added to the text (sections 4.3.1 and 4.3.2).
Comment from Rainer, page 18: This is not an "inconsistency" in our data, but rather the wavelength dependent dsigma/dT is a "feature" of the spectra.
The word inconsistency suggests something is wrong with the spectra. While indeed the xs is a physical property. The inconsistency here is introduced by the DOAS fit, and driven by differences in dsigma/dT between different O4 bands that leads to bias if two bands with different dsigma/dT are fitted as part of a single fit window.
I think that's what you are trying to say here. But it is not what the text currently says.
I think the non smooth temperature dependene shown in Fig. A27 indicates a (slight) inconsistency. Of course the Thalman and Volkamer O4 cross sections are a very useful data sets, which helped a lot in improving the O4 spectral analyses compared to earlier O4 cross section measurements.
But this non-smooth temperature dependence is not what should be expected.
Comment from Rainer, page 18: 'Thalman and Volkamer' corrected
Comment from Rainer, page 18: The temperature dependence of the peak xs at 380nm is shown in Fig. S2 of T&V2013. It looks continuous there.
This is not true. Have a closer look at the values for 380 nm. There is exactly the same slight inconsistency found between 233K and 253K. The reference to Fig. S2 is added to the text.
dsigma/dT for each band behaves a little differently, i.e., larger at 380nm than at 360nm. For some bands, the peak xs is flat down to 253K, and a change in the band shape kicks in only at lower temperatures. When you renormalize at 360nm, you are transferring some of the 360nm behavior to 380nm, which distorts the picture at 380nm.

The normalisation was applied to make the ratio between both peeks more clearly visible.
The caption to Fig. A27 should clearly state that the Figure is constructed from a perspective of a least-squares DOAS fit which weights the peak_sigma more strongly than the band integral absorption.
Probably there is a misunderstanding here. Fig. A27 is simply showing values directly calculated from the original cross sections. No DOAS is applied.
Note that Fig. A27 was updated. In the original version the wrong temperature (223K instaed of 233K) was shown. This mistake is now corrected.
Comment from Rainer, page 18: I disagree with this statement. The change in the peak sigma is compensated by narrowing of the line, leaving the band integral independent of temperature. Compare Fig. 4 in T&V2013.
This does not change the fact that the ratio of the peaks shows a slight inconsistency.
Comment from Rainer, page 18: I am not sure I can agree with this statement. Also for a single band, the least-squares nature of a DOAS fit will weigh peak_xs more strongly than band integral. The situation has thus not changed fundamentally.
If a cross section with two separate peaks is fitted to a spectrum, in which the ratio of both peaks is different, the fit tries to find a compromise for both peaks (one will be over, the other underestimated). The resulting residual is large. If only one peak is fitted, this problem does not occur.
An improvement arises from the fact that dsigma/dT is lower at 360nm than 380nm, and better defined for a single band; furthermore, the temperature dependence in peak-sigma is partially compensated by the lack of a temperature dependence in the band integral (exactly speaking, dsigma/dT is not constant, but itself a function of temperature...).
I still think that the above eplanation is correct.
Comment from Rainer, page 18: , and wavelength dependent
I still belive that the statement is correct as it is.
Comment from Rainer, page 19: After reading it several times, I think the last four lines here really belong into the next paragraph.
I still think the text is correct here as it is.
The larger difference for synthetic data surprises me. One important difference being noise, which may be shielding band-shape differences, and mislead the DOAS fit into wavelength dependent differences in dsigma/dT as described in the previous paragraph.
As stated above, the synthetic spectra without noise were used here.
Comment from Rainer, page 19: This is not obvious to me.
It seems to me then that there is no difference expected between synthetic and measured data; in particular, the synthetic spectra are based on complete information. The differences are thus surprising. The differences could be due to either the other cross-sections, or noise, or a combination of the two.
Indeed, the differences are surprising, but can be understood as explained in the text..
Comment from Rainer, page 19: An alternative explanation for the different behavior of synthetic and measured spectra is noise.
The band shape effect is tough to unravel experimentally from noise at the UV wavelengths. We tried in the cold uFT... see Fig. 5 in Spinei et al. 2015.
As stated above, the synthetic spectra without noise were used here.
An important evidence is also Fig. S2 in the Supplement of T&V2013, where the peak-sigma is compared with the balloon profiles from Klaus Pfeilsticker. The agreement is remarkeable at all temperatures. However, noise in the ballon spectra did not reveal a band-shape change, which lead Klaus to attribute the change in peak sigma (T) to a change in the equilibrium constant of O4 (we now know its a band-shape change, and the integral absorption is independent of temperature).

As stated above, the slight inconsistency is also seen in Fig. S2 in the Supplement of T&V2013
(the Fig. S2 of the Supplement is copied below; the magenta ellipse was added to mark the (slight) inconsistency)

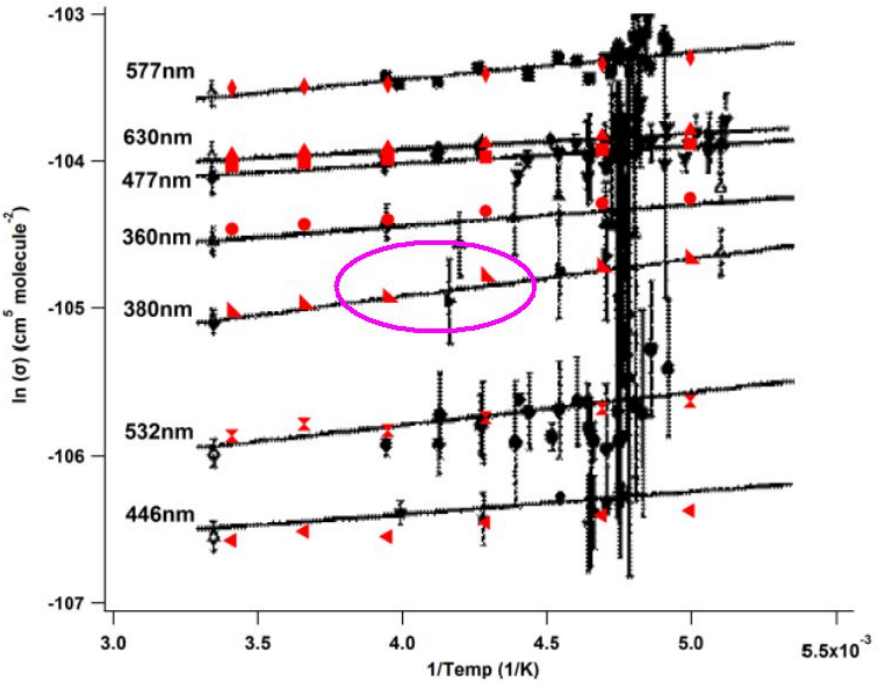


Figure S2: Comparison of Peak $\sigma_{O4\text{-}CIA}$ of this work to Figure 3 of Pfeilsticker et al. 2001. The red triangles represent our data, while the black circles are measurements from a balloon-borne DOAS instrument, combined with the room temperature cross-section of Greenblatt et al. (black diamonds).

Can noise be ruled out to explain the different behavior in the synthetic data? I think its worth mentioning here.
See above
Comment from Rainer, page 19: you mean to say "unity"?
No, 'zero' is correct
Comment from Rainer, page 19: wavelength dependent differences in the temperature dependence
I still think 'inconsistency' is correct.
Comment from Ted, page 20: I understand this to the random uncertainty only, but that may be worth saying more explicitly.
OK, this information is added.
Comment from Ted, page 21: corrected
Comment from Ted, page 22: There is not much discussion of July 8. I would add a sentence addressing Appendix 6 here. Perhaps:
"Significant redistribution of aerosol extinction to high altitudes on July 8 results in ratios of simulated and measured dAMFs not significantly different from unity."
I think this still fits under the subsection heading even if it is more explicitly dealt with in the next subsection.
I think this statement does not fit well here. This section is on the differences between both days. The uncertainty about the profile shape of the extinction at 360 nm is much larger on 18 June than on 8 July.
Comment from Rainer, page 22: possibly unrealistic

The text was changed to: ‘This section describes possible (but probably unrealistic) changes…’

Comment from Ted, page 22: I am not aware of a specific lack of agreement of contradiction.

From the AERONET inversion products a 10% redistribution was obtained, but not 27%.

Comment from Rainer, page 23: No attempts were made to optimize layers aloft in this respect. At the same time, the lower near surface extinction in this scenario is qualitatively consistent with the lower PM mass loadings measured near the surface on 8 July, which appear to be at odds with the rather constant near surface extinction between both days.

Only with the very large 27% redistribution the measured O4 dAMFs could be matched. But this scenario systematically understimates the zenith observations at large SZA.

It is only possible to ‘optimise’ the one aspect or the other.

Comment from Ted, page 23: I would cite here also that in situ aerosol measurements at the distributed site are consistent.

Good point! This information was added.

Comment from Ted, page 23: This would need to manifest differently between the two days and consistently so across the different instruments, correct? That would require some common environmental cause for increased straylight or else be vanishingly improbable.

Comment from Rainer, page 23: I agree. Consider removing these sentences.

In principle I agree to this argument. However, on 18 June the uncertainties of the aerosol extinction profile are much larger than on 8 July. So we don’t have to expect similar results on both days.

Comment from Ted, page 24: Several nested subclauses here, and I think an unnecessary comma. Perhaps rephrase to: "However, as long as the reason for this deviation is not understood, it is unclear how ..."

OK, corrected.

Comment from Ted, page 26: Replaced by more specific table footnotes below?

Not clear what is suggested. It seems to me that everything is OK here.

Comment from Rainer, page 27: V15 should be listed as separate from S15. Note that they deal with different case studies. V15 is focused on an evaluation of O4-inferred aerosol (case "with aerosol"), while S15 is focused on an evaluation of O4 in a Rayleigh atmosphere. Maybe add a footnote to this effect here.

Both references are now separately liested.

Comment from Ted, Page 90: corrected

Comment from Ted, page 92: ‘along’ inserted

Comment from Ted, page 92: changed as suggested

Comment from Rainer, page 93: Please show both case study days up to 8km. The scale seems to be cut off on 8 July.

These are examples of the MPIC extraction which set the extinction to zero above 6 km.

At the beginning of appendix A5 the following information was added: ‘Note that in this section the individual steps are described according to the MPIC procedure. The extracted profiles from other groups differ slightly compared to the results of the MPIC procedure, especially with respect to the altitude above which the extinction was set to zero (see Fig. 9).’

Comment from Rainer, page 96: a modest enhancement of the O4 dAMF is found in the elevated EAs (see appendix A6).

In both cases the enhancement is the same (+17%). The text (wrongly stating +15%) was corrected accordingly.

Comment from Ted, page 97: corrected

Comment from Ted, page 97: corrected: 15% => 17%

Comment from Ted, page 98: One or more numbers to compare with the 3% and 7% effects quoted above would be helpful. Similarly, regarding my suggestion to quote one such number in the main text.
Not clear what is suggested here.
Comment from Ted, page 98: corrected
Comment from Ted, page 98: As an aside, I think the remaining shortcomings in fact indicate that there is unleveraged information from the MAX-DOAS measurements, but that is beyond the scope of this work.
Not clear to me, what exactly is meant here.

Email from Thomas, 17.02.2019

Dear Rainer,

many thanks for your feedback!

It seems that we can not come to an agreement about the extraction of the profile shape. I understand your email that you will not agree to be co-author if your suggested changes are not implemented.

I have now sent the revised manuscript as well as the responses to the reviewer comments to the co-authors. I will also send them the protocol of our discussions. After I have received their feedback, I will send the manuscript again to you. Maybe you can then still agree to become co-author(s).

Best regards,

Thomas

Email from Thomas, 09.03.2019

Dear Rainer, Ivan, Ted,

attached I send you the updated version of the paper based on the feedback of the other co-authors. Please have a look at it and let me know until Sunday if you agree to be co-author.

Also attached is the protocol of our email exchanges, which will be uploaded to the discussion page.

Many thanks for your feedback!

Thomas

Revised paper

Is a scaling factor required to obtain closure between measured and modelled atmospheric O_4 absorptions? – An assessment of uncertainties of measurements and radiative transfer simulations ~~case study~~ for two days during the MAD-CAT campaign

Thomas Wagner[1], Steffen Beirle[1], Nuria Benavent[2], Tim Bösch[3], Ka~~i~~ Lok Chan[4], Sebastian Donner[1], Steffen Dörner[1], Caroline Fayt[5], Udo Frieß[6], David García-Nieto[2], Clio Gielen[5]*, David González-Bartolome[7], Laura Gomez[7], François Hendrick[5], Bas Henzing[8], Jun Li Jin[9], Johannes Lampel[6], Jianzhong Ma[10], Kornelia Mies[1], Mónica Navarro[7], Enno ~~Peters~~[4]Peters[3]**, Gaia Pinardi[5], Olga Puentedura[7], Janis Puķīte[1], Julia Remmers[1], Andreas Richter[3], Alfonso Saiz-Lopez[2], Reza Shaiganfar[1], Holger Sihler[1], Michel Van Roozendael[5], Yang Wang[1], Margarita Yela[7]

[1] Max Planck Institute for Chemistry, Mainz, Germany
[2] Department of Atmospheric Chemistry and Climate, Institute of Physical Chemistry Rocasolano (CSIC), Spain.
[3] University of Bremen, Germany
[4] Meteorological Institute, Ludwig-Maximilians-Universität München, Germany
[5] Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium
[6] University of Heidelberg, Germany
[7] Instituto Nacional de Tecnica Aeroespacial (INTA), Spain
[8] TNO, Netherlands Institute for Applied Scientific Research
[9] CMA Meteorological Observation Center, China
[10] Chinese Academy of Meteorological Science, China
* currently at the Institute of Astronomy, KU Leuven, Belgium
** Now at Institute for protection of maritime infrastructures, German Aerospace Center (DLR), Bremerhaven, Germany

Abstract

In this study the consistency between MAX-DOAS measurements and radiative transfer simulations of the atmospheric O_4 absorption is investigated on two mainly ~~clear~~cloud-free days during the MAD-CAT campaign in Mainz, Germany, in Summer 2013. In recent years several studies indicated that measurements and radiative transfer simulations of the atmospheric O_4 absorption can only be brought into agreement if a so-called scaling factor (<1) is applied to the measured O_4 absorption. However, many studies, ~~in particular~~including such based on direct sun light measurements, came to the opposite conclusion, that there is no need for a scaling factor. Up to now, there is no broad consensus for an explanation ~~for~~ of the observed discrepancies between measurements and simulations. Previous studies inferred the need for a scaling factor from the comparison of the aerosol optical depth derived from MAX-DOAS O_4 measurements with that derived from coincident sun photometer measurements. In this study a different approach is chosen: the measured O_4 absorption at 360 nm is directly compared to the O_4 absorption obtained from radiative transfer simulations. The atmospheric conditions used as input for the radiative transfer simulations were taken from independent

data sets, in particular from sun photometer and ceilometer measurements at the measurement site. ~~The comparisons are performed for two selected clear days with similar aerosol optical depth but very different aerosol properties.~~ This study has three main goals: First ~~For both days not only the O_4 absorptions are compared, but also~~ all relevant error sources of the spectral analysis, the radiative transfer simulations as well as the extraction of the input parameters used for the radiative transfer simulations are quantified. One important result obtained from the analysis of synthetic spectra is that the O_4 absorptions derived from the spectral analysis agree within 1% with the corresponding radiative transfer simulations at 360 nm. Based on the results from sensitivity studies, recommendations for optimised settings for the spectral analysis and radiative transfer simulations are given.~~The performed tests and sensitivity studies might be useful for the analysis and interpretation of O_4 MAX-DOAS measurements in future studies.~~ Second, the measured and simulated results are compared for two selected cloud free days with similar aerosol optical depth but very different aerosol properties.~~Different comparison results are found for both days~~: On 18 June, measurements and simulations agree within their (rather large) ~~errors~~ uncertainties (the ratio of simulated and measured O_4 absorptions is found to be 1.01±0.16). In contrast, on 8 July measurements and simulations significantly disagree: For the middle period of that day the ratio of simulated and measured O_4 absorptions is found to be 0.~~71~~ 82 ±0.~~12~~10, which differs significantly from unity. Thus for that day a scaling factor is needed to bring measurements and simulations into agreement. Third, recommendations for further intercomparison exercises are derived. ~~One possible reason for the comparison results on 18 June is the rather large aerosol extinction (and its large uncertainty) close to the surface, which has a large effect on the radiative transfer simulations.~~ One important recommendation for future studies is that aerosol profile data should be measured at the same wavelengths as the MAX-DOAS measurements. Also the altitude range without profile information close to the ground should be minimised and detailed information on the aerosol optical and/or microphysical properties should be collected and used.

~~Besides the inconsistent comparison~~The results for both days are inconsistent, ~~also~~ and no explanation for a O_4 scaling factor could be derived in this study. Thus similar, but more extended future studies should be performed, ~~which preferably include~~including more measurement days, and more instruments~~ and should be supported by more detailed independent aerosol measurements~~. Also additional wavelengths should be included. ~~The MAX-DOAS measurements collected during the recent CINDI-2 campaign are probably well suited for that purpose.~~

1 Introduction

Observations of the atmospheric absorption of the oxygen collision complex $(O_2)_2$ (in the following referred to as O_4, see Greenblatt et al. (1990)) are often used to derive information about atmospheric light paths from remote sensing measurements of scattered sun light (made e.g. from ground, satellite, balloon or airplane). Since atmospheric radiative transport is strongly influenced by scattering on aerosol and cloud particles, information on the presence and properties of clouds and aerosols can be derived from O_4 absorption measurements.

Early studies based on O_4 measurements focussed on the effect of clouds (e.g. Erle et al., 1995; Wagner et al., 1998; Winterrath et al., 1999; Acarreta et al., 2004; Sneep et al., 2008; Heue et al., 2014; Gielen et al., 2014; Wagner et al., 2014), which is usually stronger than that of aerosols. Later also aerosol properties were derived from O_4 measurements, in particular from Multi-AXis- (MAX-) DOAS measurements (e.g. Hönninger et al., 2004; Wagner et al., 2004; Wittrock et al., 2004; Friess et al., ~~2004~~2006; Irie et al., 2008; Clémer 2010; Friess et al., 2016 and references therein). For the retrieval of aerosol profiles usually forward model simulations for various assumed aerosol profiles are compared to measured O_4 slant column

densities (SCD, the integrated O_4 concentration along the atmospheric light path). The aerosol profile associated with the best fit between the forward model and measurement results is considered as the most probable atmospheric aerosol profile (for more details, see e.g. Frieß et al., 2006). Note that in some cases no unique solution might exist, if different atmospheric aerosol profiles lead to the same O_4 absorptions. MAX-DOAS aerosol retrievals are typically restricted to altitudes below about 4 km; see Friess et al. (2006).

About ten years ago, Wagner et al. (2009) suggested to apply a scaling factor (SF <1) to the O_4 SCDs derived from MAX-DOAS measurements at 360 nm in Milano in order to achieve agreement with forward model simulations. They found that on a day with low aerosol load the measured O_4 SCDs were larger than the model results, even if no aerosols were included in the model simulations. If, however, the measured O_4 SCDs were scaled by a SF of 0.81, good agreement with the forward model simulations (and nearby AERONET measurements) was achieved. Similar findings were then reported by Clémer et al. (2010), who suggested a SF of 0.8 for MAX-DOAS measurements in Beijing. Interestingly, they applied this SF to four different O_4 absorption bands (360, 477, 577, and 630 nm).

While with the application of a SF the consistency between forward model and measurements was substantially improved, both studies could not provide an explanation for the physical mechanism behind such a SF. In the following years several research groups applied a SF in their MAX-DOAS aerosol profile retrievals. However, a similarly large fraction of studies (including direct sun measurements and aircraft measurements, see Spinei et al. (2015)) did not find it necessary to apply a SF to bring measurements and forward model simulations into agreement. An overview on the application of a SF in various MAX-DOAS publications after 2010 is provided in Table 1. Up to now, there is no community consensus on whether or not a SF is needed for measured O_4 DSCDs. This is a rather unfortunate situation, because this ambiguity directly affects the aerosol results derived from MAX-DOAS measurements and thus the general confidence in the method.

So far, most of the studies deduced the need for a SF in a rather indirect way: aerosol extinction profiles derived from MAX-DOAS measurements using different SF are usually compared to independent data sets (mostly AOD from sun photometer observations) and the SF leading to the best agreement is selected. In many cases SF between 0.75 and 0.9 were derived.

In this study, we follow a different approach: similar to Ortega et al. (2016) we directly compare the measured O_4 SCDs with the corresponding SCDs derived ~~from~~ with a forward model (consisting of a radiative transfer model and assumptions of the state of the atmosphere). For this comparison, atmospheric conditions which are well characterised by independent measurements are chosen. Such a procedure allows in particular quantifying the influence of the ~~errors~~ uncertainties of the individual processing steps.

One peculiarity of this comparison is that the measured O_4 SCDs are first converted into their corresponding air mass factors (AMF), which are defined as the ratio of the SCD and the vertical column density (VCD, the vertically integrated concentration) (Solomon et al., 1987).

$$AMF = \frac{SCD}{VCD} \quad (1)$$

The 'measured' O_4 AMF is then compared to the corresponding AMF derived from radiative transfer simulations for the atmospheric conditions during the measurements:

$$AMF_{measured} \overset{?}{=} AMF_{simulated} \quad (2)$$

The conversion of the measured O_4 SCDs into AMFs is carried out to ensure a simple and direct comparison between measurements and forward model simulations. Here it should be noted that in addition to the AMFs also so-called differential AMFs (dAMFs) will be compared in this study. The dAMFs represent the difference between AMFs for measurements at non-zenith elevation angles α and at 90° for the same elevation sequence:

$$dAMF_{\alpha} = AMF_{\alpha} - AMF_{90°} \quad (3)$$

For the comparison between measured and simulated O_4 (d)AMFs, two mostly ~~clear~~cloud-free days (18 June and ~~0~~8 July 2013) during the Multi Axis DOAS Comparison campaign for Aerosols and Trace gases (MAD-CAT) campaign are chosen (http://joseba.mpch-mainz.mpg.de/mad_cat.htm). As discussed in more detail in section 4.2.2, based on the ceilometer and sun photometer measurements, three periods on each of ~~both~~ the two days are selected, during which the variation of the aerosol profiles was relatively small (see Table 2). In addition to the aerosol profiles, also other atmospheric properties are averaged during these periods before they are used as input for the radiative transfer simulations.

The comparison is carried out for the O_4 absorption band at 360 nm, which is the strongest O_4 absorption band in the UV. In principle also other O_4 absorption bands (e.g. in the visible spectral range) could be chosen, but these bands are not covered by the wavelength range of the MPIC instrument. Thus they are not part of this study.

The comparison between measurements and simulations is performed in three different steps: First, for two selected periods in the middle of both days, the ratios between measured and simulated O_4 (d)AMFs are calculated for standard settings of the spectral retrieval and radiative transfer simulations (for details see below). In a second step the uncertainties of the measurements and simulations are investigated. In the final step, it is investigated whether the ratio of measured and simulated O_4 (d)AMFs agree with unity taking into account these uncertainties.

Deviations between forward model and measurements can have different reasons. ~~:~~ In the following an overview on these error sources and the way they are investigated in this study are given:

a) Calculation of O_4 profiles and O_4 VCDs (eq. 1):

Profiles and VCDs of O_4 are derived from pressure and temperature profiles. The ~~errors~~ uncertainties of the pressure and temperature profiles are quantified by sensitivity studies and by the comparison of the extraction results derived from different groups/persons (see Table 3).

b) Calculation of O_4 (d)AMFs from radiative transfer simulations:

Besides differences between the different radiative transfer codes, the dominating ~~error~~ sources of uncertainty are ~~the uncertainties of~~those related to the input parameters. They are investigated by sensitivity studies and by the comparison of extracted input data by different groups/persons. Also the effects of operating different radiative transfer models by different groups are investigated.

c) Analysis of the O_4 (d)AMFs from MAX-DOAS measurements:

Uncertainties of the spectral analysis results are caused by errors and imperfections of the measurements/instruments, by the dependence of the analysis results on the specific fit settings, and the uncertainties of the O_4 cross sections including their temperature dependence. They are investigated by systematic variation of the DOAS fit settings (for measured and synthetic spectra), and by comparison of analysis results obtained from different groups and/or instruments.

The paper is organised as follows: in section 2, information on the selected days during the MAD-CAT campaign, on the MAX-DOAS measurements, and on the data sets from independent measurements is provided. Section 3 presents initial comparison results for the

selected days using standard settings. In section 4 the uncertainties associated with each of the various processing steps of the spectral analysis and the forward model simulations are quantified by comparing them to the results for the standard settings. Section 5 presents a summary and conclusions.

2 MAD-CAT campaign, MAX-DOAS instruments and other data sets used in this study

The Multi Axis DOAS Comparison campaign for Aerosols and Trace gases (MAD-CAT) (http://joseba.mpch-mainz.mpg.de/mad_cat.htm) took place in June and July 2013 on the roof of the Max-Planck-Institute for Chemistry in Mainz, Germany. The main aim of the campaign was to compare MAX-DOAS retrieval results of several atmospheric trace gases like NO_2, HCHO, HONO, CHOCHO as well as aerosols. The measurement location was at 150m above sea level at the western edge of the city of Mainz.

2.1 MAX-DOAS instruments

During the MAD-CAT campaign, 11 MAX-DOAS instruments were operated by different groups; an overview can be found at the website http://joseba.mpch-mainz.mpg.de/equipment.htm. The main viewing direction of the MAX-DOAS instruments was towards north-west (51° with respect to North). Measurements at this viewing direction were the main focus of this study, but a few comparisons using the 'standard settings' (see section 3) were also carried out for three other azimuth angles (141°, 231°, 321°, see Fig. A2 I in appendix A1). Each elevation sequence contains the following elevation angles: 1, 2, 3, 4, 5, 6, 8, 10, 15, 30 and 90°. In this study, in addition to the MPIC instrument, also spectra from 3 other MAX-DOAS instruments were analysed. The instrumental details are given in Table 4. The spectra of the MPIC instrument are available at the website http://joseba.mpch-mainz.mpg.de/e_doc_zip.htm.

2.2 Additional data sets

In order to constrain the radiative transfer simulations, independent measurements and data sets were used. In particular, information on atmospheric pressure, temperature and relative humidity, as well as aerosol properties is used. In addition to local in situ measurements from air quality monitoring stations and remote sensing measurements by a ceilometer and a sun photometer, also ECMWF reanalysis data were used. An overview on these data sets is given in Table 5. The data sets used in this study are available at the websites http://joseba.mpch-mainz.mpg.de/a_doc_zip.htm and http://joseba.mpch-mainz.mpg.de/c_doc_zip.htm.

2.3 RTM simulations

Several radiative transfer models are used to calculate O_4 (d)AMFs for the selected days. As input, vertical profiles of temperature, pressure, relative humidity and aerosol extinction extracted from the independent data sets (see section 2.2 and 4) were used. The vertical resolution is high in the lowest layers and decreases with increasing altitude (see Table A1 in appendix A1). The upper boundary of the vertical grid is set to 1000 km. The lower boundary of the model grid represents the surface elevation of the instrument (150 m above sea level). For the 'standard run', a surface albedo of 5% is assumed and the aerosol optical properties are described by a Henyey-Greenstein phase function with an asymmetry parameter of 0.68 and a single scattering albedo of 0.95. Both values represent typical urban aerosols (see e.g. Dubovik et al., 2002). Ozone absorption was not considered, because it is very small at 360

nm. The MAD-CAT campaign took place around summer solstice. Thus the same dependence of the solar zenith angle (SZA) and relative azimuth angle (RAZI) on time is used for both days (see Table A2 in the appendix A1). The input data used for the radiative transfer simulations are available at the website http://joseba.mpch-mainz.mpg.de/d_doc_zip.htm. In the following sub-sections the different radiative transfer models used in this study are described.

2.3.1 MCARTIM

The full spherical Monte Carlo radiative transfer model MCARTIM (Deutschmann et al., 2011) explicitly simulates individual photon trajectories including the photon interactions with molecules, aerosol particles and the surface. In this study two versions of MCARTIM are used: version 1 and version 3. Version 1 is a 1-D scalar model. Version 3 can also be run in 3-D and vector modes. In version 1 Rotational Raman scattering (RRS) is partly taken into account: the RRS cross section and phase function are explicitly considered for the determination of the photon paths, but the wavelength redistribution during the RRS events is not considered. In version 3 RRS can be fully taken into account. If operated in the same mode (1-D scalar) both models show excellent agreement.

2.3.2 LIDORT

In this study the LIDORT version 3.3 was used. The Linearized Discrete Ordinate Radiative Transfer (LIDORT) forward model (Spurr et al., 2001; Spurr et al., 2008) is based on the discrete ordinate method to solve the radiative transfer equation (e.g.: Chandrasekhar, 1960; Chandrasekhar, 1989; Stamnes et al., 1988). This model considers a pseudo-spherical multi-layered atmosphere including several anisotropic scatters. The formulation implemented corrects for the atmosphere curvature in the solar and single scattered beam, however the multiple scattering term is treated in the plane-parallel approximation. The properties of each of the atmospheric layers are considered homogenous in the corresponding layer. Using finite differences for the altitude derivatives, this linearized code converts the problem into a linear algebraic system. Through first order perturbation theory, it is able to provide radiance field and radiance derivatives with respect to atmospheric and surface variables (Jacobians) in a single call. LIDORT was used in several studies to derive vertical profiles of aerosols and trace gases from MAX-DOAS (e.g. Clémer et al., 2010; Hendrick et al., 2014; Franco et al., 2015).

2.3.3 SCIATRAN

The RTM SCIATRAN (Rozanov et al. 2014) was used in its full-spherical mode including multiple scattering but without polarization. In the operation mode used here, SCIATRAN solves the transfer equations using the discrete ordinate method. In this study, SCIATRAN was used by two groups: The IUP Bremen group used v3.8.3 ~~for the~~ for the O_4 dAMFs simulations (without Raman scattering). The MPIC group used v3.6.11 for the calculation of synthetic spectra (see Section 2.4) and for the O_4 dAMFs simulations (including Raman scattering).

2.4 Synthetic spectra

In addition to AMFs and dAMFs, also synthetic spectra were simulated. They are analysed in the same way as the measured spectra, which allows the investigation of two important aspects:

a) The derived O_4 dAMFs from the synthetic spectra can be compared to the O_4 dAMFs obtained directly from the radiative simulations at one wavelength (here: 360 nm) using the same settings. In this way the consistency of the spectral analysis results and the radiative transfer simulations is tested.

b) Sensitivity tests can be performed varying several fit parameters, e.g. the spectral range or the DOAS polynomial, and their effect on the derived O_4 dAMFs can be assessed.

Synthetic spectra are simulated using SCIATRAN taking into account rotational Raman scattering. The basic simulation settings are the same as for the RTM simulations of the O_4 (d)AMFs described above. In order to minimise the computational effort, for the profiles of temperature, pressure, relative humidity and aerosol extinction the input data for only two periods (18 June: 11:00 – 14:00, ~~0~~8 July: 7:00 – 11:00, see Table 2) are used for the whole day. Thus 'perfect' agreement with the measurements can only be expected for the two selected periods. Aerosol optical properties (phase function and single scattering albedo) are taken from AERONET measurements of the two selected days. Although the wavelength dependencies of both quantities (and also for the aerosol extinction) are considered, it should be noted that the associated uncertainties are probably rather large, since the optical properties in the UV had to be extrapolated from measurements in the visible spectral range. ~~Moreover, the phase functions were not available as fully consolidated AERONET level 2.0 data, but only as level 1.5 data.~~

Spectra were simulated at a spectral resolution of 0.01 nm and convolved with a Gaussian slit function of 0.6 nm full width at half maximum (FWHM), which is similar to those of the measurements. For the generation of the spectra a high resolution solar spectrum (Chance and Kurucz, 2010) and the trace gas absorptions of O_3, NO_2, HCHO, and O_4 are considered (see Table A3 in appendix A1). The assumed tropospheric profiles of NO_2 and HCHO are similar to those retrieved from the MAX-DOAS observations during the selected periods. Time series of the tropospheric VCDs of NO_2 and HCHO for the two selected days are shown in Fig. A1 in appendix 1.

Two sets of synthetic spectra were simulated, one taking into account the temperature dependence of the O_4 cross section and the other not. For the case without considering the temperature dependence, the O_4 cross section for 293 K is used. In addition to spectra without noise, also spectra with noise (sigma of the noise is assumed as $7.5 \cdot 10^{-4}$ times the intensity) were simulated. The synthetic spectra are available at the website http://joseba.mpch-mainz.mpg.de/f_doc_zip.htm.

3 Strategies used in this studies and comparison results for 'standard settings'

3.1 Selection of days

For the comparison of measured and simulated O_4 dAMFs, two mostly ~~clear~~cloud-free days during the MAD-CAT campaign (18 June and 8 July 2013) were selected. On both days the AOD measured by the AERONET sun photometer at 360 nm ~~is~~ was between 0.25 and 0.4 (see Fig. 1). In spite of the similar AOD, very different aerosol properties at the surface ~~are~~ were found on the two days: on 18 June much higher concentrations of large aerosol particles ($PM_{2.5}$ and PM_{10}) are found. These differences are also represented by the large differences of the Ångström parameter for long wavelengths (440 – 870 nm) on both days. Also the aerosol height profiles are different: On 8 July rather homogenous profiles with a layer height of

about 2 km occur. On 18 June the aerosol profiles reach to higher altitudes, but the highest extinction is found close to the surface. Also the temporal variability of the aerosol properties, especially the near-surface concentrations, is much larger on 18 June.

3.2 Different levels of comparisons

The comparison between the forward model and MAX-DOAS measurements is performed in different depth for different subsets of the measurements:

a) A quantitative comparison of O_4 AMFs and O_4 dAMFs is performed for 3° elevation angle at the standard viewing direction (51° with respect to North) for the middle periods of both selected days. During these periods the uncertainties of the measurement and the radiative transfer simulations are smallest because around noon the measured intensities are high and the variation of the SZA is small. During the selected periods, also the variation of the ceilometer profiles is relatively small. These comparisons thus constitute the core of the comparison exercise and all sensitivity studies are performed for these two periods. The elevation angle of 3° is selected because for such a low elevation angle the atmospheric light paths and thus the O_4 absorption are rather large. Moreover, as can be seen in Fig. 2, the O_4 (d)AMFs for 3° are very similar to those for 1° and 6°, especially on 8 July 2013. Sensitivity studies showed that a wrong elevation angle calibration (±0.5°) led to only small changes (<1%) of the O_4 (d)AMFs. Changes of the field of view between 0.2 and 1.1° led to even smaller differences. ~~This~~ These findings indicate~~s~~ that possible uncertainties of the calibration of the elevation angles of the instruments can be neglected. Here it is interesting to note that on 18 June even slightly lower O_4 (d)AMFs are found for the low elevation angles. This is in agreement with the finding of high aerosol extinction in a shallow layer above the surface (see Fig. 1). The azimuth angle of 51° is chosen, because it was the standard viewing direction during the MAD-CAT campaign and measurements for this direction are available from different instruments.

b) The quantitative comparison for 3° elevation and azimuth of 51° is also extended to the periods prior and after the middle periods of the selected days. However, to minimise the computational efforts, some sensitivity studies are not carried out for the first and last periods.

c) The comparison is extended to more elevation angles (1°, 3°, 6°, 10°, 15°, 30°, 90°) and azimuth angles (51°, 141°, 231°, 321°). For this comparison only the standard settings for the DOAS analysis and the radiative transfer simulations are applied (see Tables 6 and 7). The comparison results for the MPIC MAX-DOAS measurements are shown in appendix A2. The purpose of this comparison is to check whether for other viewing angles similar results are found as for 3° elevation at 51° azimuth direction.

3.3 Quantitative comparison for 3° elevation in standard azimuth direction

Fig. 3 presents a comparison of the measured and simulated O_4 (d)AMFs for 3° elevation and 51° azimuth on both days. For the spectral analysis and the radiative transfer simulations the respective 'standard settings' (see Tables 6 and 7) were used. On 8 July the simulated O_4 (d)AMFs systematically underestimate the measured O_4 (d)AMFs by up to 40%. Similar results are also obtained for other elevation and azimuth angles (see appendix ~~A1~~A2), the differences becoming smaller towards higher elevation angles. In contrast, no systematic underestimation is observed for most of 18 June. For some periods of that day the simulated O_4 (d)AMFs are even larger than the measured O_4 (d)AMFs. However, here it should be noted that the aerosol extinction profile of the 'standard settings' (using linear extrapolation below 180 m where no ceilomter data are available) probably underestimates the aerosol extinction close to the surface. If instead a modified aerosol profile with strongly increased aerosol extinction below 180 m and the maximum AOD during that period is used (see Fig.

A31 in appendix A5) the corresponding (d)AMFs fall below the measured O_4 (d)AMFs (green curves in Fig. A4 in appendix A2). More details on the extraction of the aerosol extinction profiles are given in section 4.2.2 and appendix A5).
The average ratio of simulated to measured (d)AMFs (for the standard settings) during the middle periods on both days are given in Table 8. For 18 June they are close to unity, for 8 July they are much lower (0.83 for the AMF, and 0.69 for the dAMF).

4 Estimation of the uncertainties of the different processing steps

There are 3 major processing steps, for which the uncertainties are quantified in this section:
a) The determination of the O_4 height profiles and corresponding O_4 vertical column densities.
b) The simulation of O_4 (d)AMFs by the forward model
c) The analysis of O_4 (d)AMFs from the MAX-DOAS measurements.

4.1 Determination of the vertical O_4 profile and the O_4 VCD

The O_4 VCD is required for conversion of measured (d)SCDs into (d)AMFs (eq. 1). O_4 profiles are also needed for the calculation of O_4 (d)AMFs. The accuracy of the calculated O_4 height profile and the O_4 VCD depends in particular on two aspects:
a) is profile information on temperature, pressure and (relative) humidity available?
b) what is the accuracy of these data sets?
Additional uncertainties are related to the details of the calculation of the O_4 concentration and O_4 VCDs from these profiles. Both ~~error~~ sources of uncertainties are investigated in the following sub sections.

4.1.1 Extraction of vertical profiles of temperature and pressure

The procedure of extracting temperature and pressure profiles depends on the availability of measured profile data or surface measurements. If profile data are available (e.g. from sondes or models) they could be directly used. If only surface measurements are available, vertical profiles of temperature and pressure could be calculated making assumptions on the lapse rate (here we assume a value of -0.65 K / 100 m). If no measurements or model data are available, profiles from the US standard atmosphere might be used (United States Committee on Extension to the Standard Atmosphere, 1976). In appendix A3 the different procedures for the extraction of pressure and temperature profiles are described in detail for the two days of the MAD-CAT campaign. For these days the optimum choice was to combine the model data and the surface measurements. In that way, the diurnal variation in the boundary layer could be considered. In Fig. 4 temperature and pressure profiles extracted from the combination of in situ measurements and ECMWF data are shown. These profiles probably best match the true atmospheric profiles.
~~For the two selected days during the MADCAT campaign two data sets of temperature and pressure are available: surface measurements close to the measurement site and vertical profiles from ECMWF ERA Interim re-analysis data (see Table 5). Both data sets are used to derive the O_4 concentration profiles for the three selected periods on both days. The general procedure is that first the temperature profiles are determined. In a second step, the pressure profiles are derived from the temperature profiles and the measured surface pressure. For the temperature profile extraction, three height layers are treated differently:~~
~~- below 1 km~~
~~Between the surface (~150 m above sea level) and 1 km, the temperature is linearly interpolated between the average of the in situ measurements of the respective period and the~~

~~ECMWF data at 1 km (see next paragraph). This procedure is used to account for the diurnal variation of the temperature close to the surface. Here it is important to note that for this surface near layer the highest accuracy is required, because a) the maximum O_4 concentration is located near the surface, and b) the MAX-DOAS measurements are most sensitive close to the surface.~~

~~- 1 km to 20 km~~

~~In this altitude range, the diurnal variation of the temperature becomes very small. Thus the average of the four ECMWF profiles of each day is used (for simplicity, a 6th order polynomial is fitted to the ECMWF data).~~

~~- Above 20 km~~

~~In this altitude range the accuracy of the temperature profile is not critical and thus the ECMWF temperature profile for 00:00 UTC of the respective day is used for simplicity.~~

~~The temperature profiles for 8 July 2013 extracted in this way are shown in Fig. 4 (left). Close to the surface the temperature variation during the day is about 10 K.~~

~~In the next step, the pressure profiles are determined from the surface pressure (obtained from the in situ measurements) and the extracted temperature profiles according to the ideal gas law. In principle the effect of atmospheric humidity could also be taken into account, but the effect is very small for surface near layers and is thus ignored here. The derived pressure profiles for 8 July 2013 are shown in Fig. 4 (right). Excellent agreement with the corresponding ECMWF pressure profiles is found.~~

~~Here it should be noted that in principle also the ECMWF pressure profiles could be used. However, we chose to determine the pressure profiles from the surface pressure and the extracted temperature profiles, because this procedure can also be applied if no ECMWF data (or other information on temperature and pressure profiles) is available.~~

~~If no profile data (e.g. from ECMWF) are available, temperature and pressure profiles can also be extrapolated from surface measurements e.g. by assuming a constant lapse rate of -0.65 K / 100 m for the altitude range between the surface and 12 km, and a constant temperature above 12 km (as stated above, uncertainties at this altitude range have only a negligible effect on the O_4 VCD). If no measurements or model data are available at all, a fixed temperature and pressure profile can be used, e.g. the US standard atmosphere (United States Committee on Extension to the Standard Atmosphere, 1976).~~

A comparison of ~~the different~~ temperature profiles extracted by different methods for two selected periods on both days is shown in Fig. 5. For 8 July (right), rather good agreement is found, but for 18 June (left) the agreement is worse (differences up to 20 K). Of course, the differences between the true and the US standard atmosphere profiles can become even larger, depending on location and season. So the use of a fixed temperature and pressure profile should always be the last choice. In contrast, the simple extrapolation from surface values can be very useful if no profile data are available, because the uncertainties of this method are usually smallest at low altitudes, where the bulk of O_4 is located.

4.1.2 Calculation of O_4 concentration profiles and O_4 VCDs

From the temperature and pressure profiles the oxygen (O_2) concentration is calculated. Here also the effect of the atmospheric humidity profiles should be taken into account (see ~~below~~appendix A3), because it can have a considerable effect on the near-surface-~~near~~ layers (at least for temperatures of about > 20°C). Finally, the square of the oxygen concentration is calculated and used as proxy for the O_4 concentration consistently with assumptions made in the determination of the absorption cross-sections (see Greenblatt et al., 1990). The uncertainties of the derived O_4 concentration (and the corresponding O_4 VCD) caused by the uncertainty of the input profiles is estimated by varying the input parameters (for details see appendix A3). ~~The following uncertainties are derived:~~

~~-The variation of the temperature (whole profile) by about 2K leads to variations of the O_4 concentration (or O_4 VCD) by about 0.8%.~~
~~-The variation of the surface pressure by about 3 hPa leads to variations of the O_4 concentration (or O_4 VCD) by about 0.7%.~~
~~-The effect of uncertainties of the relative humidity depends strongly on temperature: For surface temperatures of 0°C, 10°C, 20°C, 30°C, and 35°C a variation of the relative humidity of 30% leads to variations of the O_4 concentration (or O_4 VCDs) of about 0.15%, 0.3%, 0.6%, 1.2%, and 1.6%, respectively. If the effect of atmospheric humidity is completely ignored (dry air is assumed), the resulting O_4 concentrations (or O_4 VCDs) are systematically overestimated by about 0.3%, 0.7%, 1.3%, 2.5%, and 4% for surface temperatures of 0°C, 10°C, 20°C, 30°C, and 35°C, respectively (assuming a relative humidity of 70%). In this study we used the relative humidity measured by the in situ sensors. We took these values not only for the surface layers, but also for the whole troposphere. Here it should be noted that the related uncertainties of the absolute humidity decrease quickly with altitude because the absolute humidity itself decrease quickly with altitude. Since both selected days were warm or even hot summer days, we estimate the uncertainty of the O_4 concentration and O_4 VCDs due to uncertainties of the relative humidity to 1% and 0.4% on 18 June and 8 July, respectively.~~
For both selected days during the MAD-CAT campaign ~~Assuming that the uncertainties of the three input parameters are independent,~~ the total uncertainty ~~related to the~~ is ~~se factors is~~ estimated to be about 1.5% assuming that the uncertainties of the individual input parameters are independent,.
Further uncertainties arise from the procedure of the vertical integration of the O_4 concentration profiles. We tested the effect of using different vertical grids and altitude ranges. It is found that the vertical grid should not be coarser than 100 m (for which a deviation of the O_4 VCD of 0.3% compared to a much finer grid is found). If e.g. a vertical grid with 500 m layers is used, the deviation increases to about 1.3%. The integration should be performed over an altitude range up to 30 km. If lower maximum altitudes are used, the O_4 VCD will be substantially underestimated: deviations of 0.1 %, 0.5 %, and 11% are found if the integration is performed only up to 25 km, 20 km, and 10 km, respectively. Here it should be noted that the exact consideration of the altitude of the measurement site is also very important: A deviation of 50 m already leads to a change of the O_4 VCD by 1%. For the MAD-CAT measurements the altitude of the instruments is 150m ±20m.
Finally, the effects of individual extraction and integration procedures are investigated by comparing the results from different groups (see Fig. 6, and Fig. A5 in appendix A3). Except for some extreme cases, the extracted temperatures typically differ by less than 3 K below 10 km. However, the deviations are typically larger for the profiles extrapolated from the surface values and in particular for the US standard atmosphere (up to > 10 K below 10 km). ~~Also t~~The variations of the extracted pressure profiles are in general rather small (< 1% below 10 km, except one obvious outlier). ~~Also here~~However, the deviations of the profiles extrapolated from the surface values and especially the US standard atmosphere are much larger (up to > 5 % below 10 km). The resulting deviations of the O_4 concentration from the different extractions are typically <3% below 10 km (and up to > 20 % ~~below~~ above 10 km for the US standard atmosphere).
In Fig. 7 the O_4 VCDs calculated for the O_4 profiles extracted from the different groups and for the profiles extrapolated from the surface values and the US standard atmosphere are shown. The VCDs for the profiles extracted by the different groups agree within 2.5%. The deviations for the profiles extrapolated from the surface values are only slightly larger (typically within 3%), but show a large variability throughout the day, which is caused by the systematic increase of the surface temperature during the day (with temperature inversions in the morning on the two selected days). The deviations of the US standard atmosphere are up

to 5% (but can of course be larger for other seasons and locations, see also Ortega et al. (2016).
Ultimately, the accuracy with which O_4 concentrations can be calculated is limited by the assumption that O_4 (O_2-O_2) is pure collision induced absorption. If the oxygen concentration profile is well known, the uncertainty due to bound O_4 is smaller than 0.14% in Earth's atmosphere (Thalman and Volkamer, 2013).
Together with the uncertainties related to the input data sets, the total uncertainty of the O_4 VCDs determined for both selected days is estimated as 3%.

4.2 Uncertainties of the O_4 (d)AMFs derived from radiative transfer simulations

The most important ~~errors~~ uncertainties of the simulated O_4 (d)AMFs are related to the uncertainties of the input parameters used for the simulations, in particular the aerosol properties. Further uncertainties are caused by imperfections of the radiative transfer models. These ~~error~~ sources of uncertainty are discussed and quantified in the following sub sections.

4.2.1 Uncertainties of the O_4 (d)AMFs caused by uncertainties of the input parameters

In this section the effect of the uncertainties of various input parameters on the O_4 (d)AMFs is investigated. The general procedure is that the input parameters are varied individually and the corresponding changes of the O_4 (d)AMFs compared to the standard settings are quantified.
First, the effect of the O_4 profile shape is investigated. In contrast to the effect of the (absolute) profile shape on the O_4 VCD (section 4.1), here the effect of the relative profile shape on the O_4 AMF is investigated. The O_4 (d)AMFs simulated for the O_4 profiles extracted by the different groups (and for those derived from the US standard atmosphere and the profiles extrapolated from the surface values, see section 4.1) are compared to those for the MPIC O_4 profiles (using the standard settings). The corresponding ratios are shown in Fig. A6 and Table A4 in appendix A4. For the O_4 profiles extracted by the different groups, and for O_4 profiles extrapolated from the surface values, small variations are found (typically < 2%). For the ~~O_4~~ US standard atmosphere larger deviations (up to 7%) are derived.
Next the effect of the aerosol extinction profile is investigated. In this study, aerosol extinction profiles are derived from the combined ceilometer and sun photometer measurements (see Table 5). In short, the ceilometer measurements of the attenuated backscatter are scaled by the simultaneously measured aerosol optical depth (AOD) from the sun photometer to obtain the aerosol extinction profile. Also the self-attenuation of the aerosol is taken into account. The different steps are illustrated in Fig. 8 and described in detail in appendix A5. In the extraction procedure, several assumptions have to be made: First, the ceilometer profiles have to be extrapolated for altitudes below 180 m, for which the ceilometer is not sensitive. Furthermore, they have to be averaged over several hours and are in addition vertically smoothed (above 2 km) to minimise the rather large scatter. Finally, above 5 to 6 km (depending on the ceilometer profiles) the extinction is set to zero because of the further increasing scatter and the usually small extinctions. This assumption reflects a practical limitation of the ceilometer likely responsible for the larger variability in the profile shape aloft by different groups. Another assumption is that the Angström exponent and the LIDAR ratio ~~is~~ are independent of altitude, which is typically not strictly fulfilled (the LIDAR ratio describes the ratio between the extinction and backscatter probabilities of the molecules and aerosol particles).
~~Some of t~~These uncertainties are quantified by sensitivity studies, in particular the effect of the extrapolation below 180 m and the altitude above which the aerosol extinction is set to zero. Other uncertainties, like the effect of the assumption of a constant LIDAR ratio are more

difficult to quantify without further information (see below). ~~While a constant LIDAR ratio is probably a good assumption for 8 July, for 18 June the surface measurements indicate that the aerosol properties strongly change with time. Thus the LIDAR ratio might also vary stronger with altitude on that day.~~ The effect of temporal averaging and smoothing is probably negligible for 8 July, because similar height profiles are found for all three periods of that day, but on 18 June the effect might be more important.

Fig. 9 shows a comparison of the aerosol extinction profiles extracted by the different groups for the three periods on both days. Especially on 8 July systematic differences are found. They are caused by the different altitudes, above which the aerosol extinction is set to zero. In combination with the scaling of the profiles with the AOD obtained from the sun photometer, this also influences the extinction values close to the surface. Deviations up to 18% are found for the first period of 8 July. These deviations also have an effect on the corresponding O_4 (d)AMFs, where higher values are obtained for the profiles (INTA and IUPB 300m) which were extracted for a larger altitude range (Fig. A7 and Table A5 in the appendix A4). Here it is interesting to note that these differences are not related to the direct effect of the aerosol extinction at high altitude, but to the corresponding (via the scaling with the AOD) decrease of the aerosol extinction close to the surface. Larger deviations (up to 4%) are found for 8 July, while the deviations on 18 June are within 3%. This effect is further examined in appendix A6.

In Fig. A8 and Table A6 in appendix A4, the effect of the different extrapolations of the aerosol extinction profile below 180 m on the O_4 (d)AMFs is quantified. Similar deviations (up to 5 %) are found for both days.

Finally, we investigated the effect of changing aerosol optical properties with altitude (changing LIDAR ratio). Such effects are in particular important if the wavelength of the ceilometer measurements (1064 nm) differs largely from that of the MAX-DOAS observations (360 nm). Based on the partitioning into fine and coarse mode aerosols (derived from the sun photometer observations) and the corresponding phase functions and optical depths, the sensitivity of the ceilometer to fine mode aerosols were estimated (for details see appendix A5). While for 18 June the contribution of the fine mode to the ceilometer signal is about 32% on 8 July it is much larger (about 82 %). Thus it can be concluded that the aerosol extinction profile derived from the ceilometer is largely representative for the fine mode aerosols on that day. To investigate the effect of the remaining uncertainties, the shape of the aerosol extinction profile was further modified (for details see appendix A5) taking into account that the coarse aerosols are typically located at low altitudes. The corresponding repartitioning of the aerosol extinction profile led to a decrease of the aerosol extinction close to the surface which is balanced by an increase at higher altitudes (see Fig. A34). The O_4 dAMFs calculated for the modified profile are by about 17 % larger than those for the standard settings (for details see appendix A5).

The effect of elevated aerosol layers (see Ortega et al., 2016) was further investigated by systematic sensitivity studies (appendix A6). On both selected days enhanced aerosol extinction was found at elevated layers (Fig. 9). Compared to those reported by Ortega et al. (2016) the profiles extracted in this study reach even up to higher altitudes. For the investigation of the effect of changes of the aerosol extinction at different altitudes, the aerosol extinction profile on 8 July was subdivided into 3 layers (0-1.7 km; 1.7 – 4.9 km; 4.9 – 7 km), and the extinction in the individual layers was increased by +40 %. It was found that even a strong increase of the aerosol extinction at high altitudes by 40% leads only to an increase of the O_4 dAMFs by 7 %.

Also the effect of horizontal gradients should be briefly discussed. For the selected periods of both days, the wind direction and wind speed were rather constant. On 18 June the wind direction was between 80° and 150° with respect to North, and the wind speed was about 2 m/s. On 8 July the wind direction was between 70° and 90° (the wind came from almost the

same direction at which the instruments were looking), and the wind speed was about 3 m/s. During the 4 hours of the selected period on 8 July, the air masses moved over a distance of about 40 km. During the 3 hours of the selected period on 18 June, the air masses moved over a distance of about 20 km. These distances are larger than the distances for which the MAX-DOAS observations are sensitive (about 5 – 15 km). Since also the AOD and the aerosol extinction profiles were rather constant during both selected periods, we conclude that for the measurements considered here horizontal gradients can be neglected. It should also be noted that the discrepancies between measurements and simulations were simultaneously observed at all 4 azimuth directions.

In Fig. A9 and Table A7 in appendix A4, the effect of different single scattering albedos (between 0.9 and 1) on the O_4 (d)AMFs is quantified. The effect on the O_4 (d)AMFs is up to 4 % on 18 June and up to 2 % on 8 July 2013.
The impact of the aerosol phase function is investigated in two ways: First, simulation results are compared for Henyey Greenstein phase functions with different asymmetry parameters. The corresponding results are shown in Fig. A10 and Table A8 in appendix A4. The differences of the O_4 (d)AMFs for the different aerosol phase functions are rather strong: up to 3% for the O_4 AMFs and up to 8% for the O_4 dAMFs (larger uncertainties for the dAMFs are found because of the strong influence of the phase function on the 90° observations). Here it should be noted that the actual deviations from the true phase function might be even larger. In order to better estimate these uncertainties, also simulations for phase functions derived from the sun photometer measurements based on Mie theory (in the following referred to as Mie phase functions) were performed. A comparison of these Mie phase functions with the Henyey Greenstein phase functions is shown in Fig. 10. Large differences, especially in forward direction are obvious. The O_4 (d)AMFs for the Mie phase functions are compared to the standard simulations (using the HG phase function for an asymmetry parameter of 0.68) in Fig. A11 and Table A9 in aAppendix A4. Again rather large deviations are found, which are larger on 18 June (up to 9 %) than on 8 July (up to 5%).
In Fig. A12 and Table A10 in aAppendix A4, the effect of different surface albedos on the O_4 (d)AMFs is quantified. For the considered variations (0.03 to 0.1) the changes of the O_4 (d)AMFs are within 2 %.

4.2.2 Uncertainties of the O_4 (d)AMFs caused by imperfections of the radiative transfer models

The radiative transfer models used in this study are well established and showed very good agreement in several intercomparison studies (e.g. Hendrick et al., 2006; Wagner et al., 2007; Lorente et al., 2017). Nevertheless, they are based on different methods and use different approximations (e.g. with respect to the Earth's sphericity). Thus we compared the simulated O_4 (d)AMFs for both days in order to estimate the uncertainties associated to these differences. In Fig. A13 and Table A11 (appendix A4), the comparison results are shown. They agree within a few percent with slightly larger differences for 18 June (up to 6 %) than for 8 July (up to 3 %).
So far, all radiative transfer simulations were carried out without considering polarisation. Thus in Fig. A14 and Table A12 in appendix A4, the results with and without considering polarisation are compared. The corresponding differences are very small (<1%).

4.2.3 Summary of uncertainties of the O_4 AMF from radiative transfer simulations

Table 9 presents and overview on the different sources of uncertainties of the simulated O_4 (d)AMFs derived from the comparison of the results from different groups and the sensitivity

studies. The uncertainties are expressed as relative deviations from the results for the standard settings (see Table 6) derived by MPIC using MCARTIM.

In general, larger uncertainties are found for the O_4 dAMFs compared to the O_4 AMFs. This is expected because the uncertainties of the O_4 dAMFs contain the uncertainties of two simulations (at 90° elevation and at low elevation). Another general finding is that the uncertainties on 18 June are larger than on 8 July. This finding is mainly related to the larger uncertainties due to the aerosol phase function, which has an especially strong forward peak on 18 June. Also the ~~error contributions~~uncertainties from the O_4 profile extraction, the choice of the radiative transfer model and the extrapolation of the aerosol extinction below 180 m are larger on 18 June than on 8 July. These higher uncertainties are probably mainly related to the high aerosol extinction close to the surface on 18 June (see section 5.1, and appendices A2 and A5).

For the total uncertainties two values are given in Table 9: The 'average deviation' is the sum of all systematic deviations of the individual uncertainties (the corresponding mean of the maximum and minimum values). The second quantity (the 'range of uncertainties) is calculated from half the individual uncertainty ranges by assuming that they are independent.

Finally, it should be noted that for some ~~error sources~~uncertainties (e.g. the effects of the surface albedo or the single scattering albedo) the given numbers probably overestimate the true uncertainties, while for others, e.g. the uncertainties related to the aerosol extinction profiles or the phase functions they possibly underestimate the true uncertainties (although reasonable assumptions were made). The two latter ~~error sources~~uncertainties are especially large for 18 June. The differences between both days are discussed in more detail in section 5.

4.3 Uncertainties of the spectral analysis

The uncertainties of the spectral analysis are caused by different effects:

-the specific settings of the spectral analysis like the fit window or the degree of the polynomial. Of particular interest is the effect of choosing different O_4 cross sections as well as ~~its~~ their temperature dependence.

-the properties (and imperfections) of the MAX-DOAS instruments

-the effect of different analysis software and implementations

-the effect of the wavelength dependence of the AMF across the fit window.

These ~~error sources~~uncertainties are discussed and quantified in the following sub sections.

4.3.1 Comparison of O_4 (d)AMFs derived from the synthetic spectra with O_4 (d)AMFs directly obtained from the radiative transfer simulations

Synthetic spectra for both selected days were simulated using the radiative transfer model SCIATRAN (for details see section 2.4 and Table A3 in appendix A1). While spectra for the whole day are simulated (for the viewing geometry see Table A2 in appendix A1) it should be noted that the aerosol properties during the middle periods are used also for the whole day (to minimise the computational efforts). The spectra are analysed using the standard settings and the derived O_4 (d)SCDs are converted to O_4 (d)AMFs using eq. 1. In addition to the spectra, also O_4 (d)AMFs at 360 nm are simulated directly by the RT models using exactly the same settings. These O_4 (d)AMFs are used to test whether the spectral retrieval results are indeed representative for the simulated O_4 (d)AMFs at 360 nm.

Spectra are simulated with and without considering the temperature dependence of the O_4 cross section. Also one version of synthetic spectra with added random noise is processed.

First, the synthetic spectra are analysed using the standard settings (see Table 7). Examples of the O_4 fits for synthetic (and measured) spectra are shown in Fig. 11. Here it is interesting to

note that the ratios of the results for the measured and the simulated spectra are between 0.68 and 0.74, similar to ratio for the dAMFs on 8 July shown in Table 8.

In Fig. 12 the ratios of the O_4 (d)AMFs derived from the synthetic spectra versus those directly obtained from the radiative transfer simulations at 360 nm are shown. In the upper part (a) the results for synthetic spectra considering the temperature dependence of the O_4 cross section are presented (without noise). Systematically enhanced ratios are found in the morning and evening, while for most of the day the ratios are close to unity. The higher values in the morning and evening are probably partly caused by the increased light paths through higher atmospheric layers (with lower temperatures) when the solar zenith angle is high. Interestingly, if the temperature dependence of the O_4 cross section is not taken into account (Fig. 12 b), still slightly enhanced ratios during the morning and evening are found, which can not be explained anymore by the temperature dependence of the O_4 cross section. Thus we speculate ~~whether~~ that part of the enhanced values at high SZA are ~~probable~~ probably caused by the wavelength dependence of the O_4 AMFs. Nevertheless, for most of the day the ratio is very close to unity indicating that for SZA < 75° the O_4 (d)AMFs~~)~~ obtained from the spectral analysis are almost identical to the O_4 (dAMFs) directly obtained from the radiative transfer simulations (at 360 nm).

In Fig. 12 c results for spectra with added random noise (without consideration of the temperature dependence of the O_4 cross section) are shown. On average similar results as for the spectra without noise (Fig. 12 b) are found but the results now show a large scatter. From these results and also the spectral analyses (Fig. 11) we conclude that the noise added to the synthetic spectra overestimates that of the real measurements. For the sensitivity studies discussed in section 4.3.2 only synthetic spectra without noise were used.

In Table A13 in appendix A4 the average ratios for the middle periods on both selected days are shown. They deviate from unity by up to 2% indicating that the wavelength dependence of the O_4 (d)AMF is negligible for the considered cases for SZA < 75°.

4.3.2 Sensitivity studies for different fit parameters

In this section the effect of the choice of several fit parameters on the derived O_4 (d)AMFs is investigated using both measured and synthetic spectra. It should be noted that in the following only synthetic spectra without noise were used, because for the sensitivity studies we are interested in the systematic effects. Only one fit parameter is varied for each individual test, and the results are compared to those for the standard fit parameters (see Table 7).

First the fit window is varied. Besides the standard fit window (352 to 387 nm), which contains two O_4 bands, also two fit windows towards shorter wavelengths are tested: 335 – 374 nm (including two O_4 bands) and 345 – 374 nm (including one O_4 band at 360 nm). The ratios of the derived O_4 (d)AMFs versus those for the standard analysis are shown in Fig. A15 and Table A14 in appendix A2. On 18 June rather large deviations of the O_4 (d)AMFs are found for both measured (-12%) and synthetic spectra (-5%) for the spectral range 335 to 374 nm. On 8 July the corresponding differences are smaller (-6% and -2% for measured and synthetic spectra, respectively). For the spectral range 345 – 374 nm, smaller differences of only up to 1% are found for both days. The reason for the larger deviations on 18 June for the spectral range 335 – 374 nm is not clear. One possible reason could be the differences of the Ångström parameters (see Fig. 1) and phase functions (see Fig 10).

In Fig. A16 and Table A15 the results for different degrees of the polynomial used in the spectral analysis are shown. For the measured spectra systematically higher O_4 (d)AMFs (up to 6%) than for the standard analysis are found when using lower polynomial degrees. For the synthetic spectra the effect is smaller (<3%).

In Fig. A17 and Table A16 the results for different intensity offsets are shown. Again, for the measured spectra systematically higher O_4 (d)AMFs (up to 16%) than for the standard

analysis are found when reducing the order of the intensity offset, while for the synthetic spectra the effect is smaller (<3%). Higher order intensity offsets might compensate for wavelength dependent offsets (e.g. spectral straylight), which can be important for real measurements, while the synthetic spectra do not contain such contributions.
In Fig. A18 and Table A17 the results for spectral analyses with only one Ring spectrum are shown. In contrast to the standard analysis, which includes two Ring spectra (one for clear and one for cloudy sky, see Wagner et al., 2009), only the Ring spectrum for clear sky is used. For both selected days, only small deviations (within 2%) compared to the standard analysis are found.

4.3.3 Sensitivity studies using different trace gas absorption cross sections

In this section the impact of different trace gas absorption cross sections on the derived O_4 (d)AMFs is investigated.
In Fig. A19 and Table A18 the results for using two NO_2 cross sections (294 and 220 K) compared to the standard analysis (using only a NO_2 cross section for 294 K) are shown. The results are almost the same as for the standard analysis.
In Fig. A20 and Table A19 the results for using an additional wavelength-dependent NO_2 cross section compared to the standard analysis (using only one NO_2 cross section) are shown. The second NO_2 cross section is calculated by multiplying the original cross section with wavelength (Pukite et al., 2010). Again, only small deviations of the results from the standard analysis (1% for the measured spectra, and 2% for the synthetic spectra are found.
In Fig. A21 and Table A20 results for using and additional wavelength-dependent O_4 cross sections compared to the standard analysis (using only one O_4 cross section) are shown. The second O_4 cross section is calculated like for NO_2, but also an orthogonalisation with respect to the original O_4 cross section (at 360 nm) is performed. The derived O_4 (d)AMFs are almost identical to those from the standard analysis (within 1%).
For the spectral retrieval of HONO in a similar spectral range, a significant impact of water vapour absorption around 363 nm was found in Wang et al. (2017c) and Lampel et al. (2017). In Fig. A22 and Table A21 the O_4 results for including a H_2O cross section (Polyansky et al., 2018) compared to the standard analysis (using no H_2O cross section) are shown. The results are almost identical to those from the standard analysis (within 1%).
In Fig. A23 and Table A22 the results for including a HCHO cross section ~~(Polyansky et al., 2018)~~ compared to the standard analysis (using no HCHO cross section) are shown. Especially for 18 June a large systematic effect is found: the O_4 dAMFs are by 4 % or 6 % smaller than for the standard analysis for measured and synthetic spectra, respectively. On 8 July the underestimation is smaller (2% and 3% for measured and synthetic spectra, respectively).

4.3.4 Effect of using different O_4 cross sections

In Fig. A24 and Table A23 the results for different O_4 cross sections are compared to the standard analysis (using the Thalman O_4 cross section). The results for both days are almost identical. For the real measurements, the derived O_4 dAMFs using the Hermans and Greenblatt cross sections are by 3% smaller or 8 % larger than those for the standard analysis, respectively. However, if the Greenblatt O_4 cross section is allowed to shift during the spectral analysis, the overestimation can be largely reduced to only +3 %. This confirms findings from earlier studies (e.g. Pinardi et al., 2013) that the wavelength calibration of the original data sets is not very accurate.
For the synthetic spectra slightly different results than for the real measurements are found for the Hermans O_4 cross section. The reason for these differences is not clear. However, here it

should be noted that the temperature dependent O_4 absorption in the synthetic spectra does probably not exactly represent the true atmospheric O_4 absorption.

4.3.5 Effect of the temperature dependence of the O_4 cross section

The new set of O_4 cross sections provided by Thalman and Volkamer (2013) allows to investigate the temperature dependence of the atmospheric O_4 absorptions in detail. They provide O_4 cross sections measured at five temperatures (203, 233, 253, 273, 293 K) covering the range of temperatures relevant for atmospheric applications. Using these cross sections, the effect of the temperature dependence of the O_4 absorptions is investigated in two ways:
a) In a first test, synthetic spectra are simulated for different surface temperatures assuming a fixed lapse rate. These spectra are then analysed using the O_4 cross section for 293K (which is usually used for the spectral analysis of O_4). From this study the magnitude of the effect of the temperature dependence of the O_4 cross section on MAX-DOAS measurements can be quantified.
b) In a second test, measured and synthetic spectra for both selected days are analysed with O_4 cross sections for different temperatures. From this study it can be seen to which degree the temperature dependence of the O_4 cross section can be already corrected during the spectral analysis (if two O_4 cross sections are used simultaneously).
For the first study, MAX-DOAS spectra are simulated in a simplified way:
-Atmospheric temperature profiles are constructed for surface temperatures between 220 K and 310 K in steps of 10 K assuming a fixed laps rate of –0.656 K / 100 m.
-For each altitude layer (vertical extension: 20 m below 500m, 100 m between 500 m and 2 km, 200 m between 2 km and 12 km, 1 km above) the O_4 concentrations (calculated from the US standard atmosphere) are multiplied with the corresponding differential box-AMFs calculated for typical atmospheric conditions and viewing geometries (see Fig. A25 in appendix A4).
-High resolution absorption spectra are calculated by applying the Beer-Lambert-law for each height layer using the O_4 cross section of the respective temperature (interpolated between the two adjacent temperatures of the Thalman and Volkamer data set).
-The derived high resolution spectra are convolved with the instrument slit function (FWHM of 0.6 nm).
-The logarithm of the ratio of the spectra for the low elevation and zenith is calculated and analysed using the O_4 cross section for 293 K.
-The derived O_4 dAMFs are divided by the corresponding dAMFs directly obtained from the radiative transfer simulations.
These calculated ratios as function of the surface temperature are shown in Fig. 13. A strong and systematic dependence on the surface temperature is found (15 % for a change of the surface temperature between 240 and 310 K). However, except for measurements at polar regions, the deviations are usually small. Since for both selected days the temperatures were rather high (indicated by the two coloured horizontal bars in the figure), the effect of the temperature dependence of the O_4 absorption for the middle periods of both days is very small (-1 to -2% for 18 June, and 0 to +1% on 8 July). It should be noted that the results shown in Fig. 13 are obtained for generalised settings of the radiative transfer simulations. Thus it is recommended that future studies should investigate the effect of the temperature dependence in more detail and using the exact viewing geometry for individual observations. However, since the temperatures on both selected days were rather high, for this study the simplifications of the radiative transfer simulations have no strong influence on the derived results.
In the second test the measured and synthetic spectra are analysed using O_4 cross sections for different temperatures. The corresponding results are shown in Fig. A26 and Table A24.

If only the O_4 cross section at low temperature (203 K) is used, the derived O_4 AMFs and dAMFs are by about 16% and 30% smaller than for the standard analysis (using the O_4 cross section for 293 K). These results are consistently obtained for the measured and synthetic spectra. If, however, two O_4 cross sections (for 203 and 293 K) are simultaneously included in the analysis, different results are obtained for the measured and synthetic spectra: for the measured spectra the derived O_4 (d)AMFs agree within 4% with those from the standard analysis. In contrast, for the synthetic spectra, the derived O_4 (d)AMFs are systematically smaller (by about 6 to 18 %). This finding was not expected, because exactly the same cross sections were used for both the simulation and the analysis of the synthetic spectra. Detailed investigations (see appendix A4) led to the conclusion that there is a slight inconsistency in the temperature dependence of the O_4 cross sections from Thalman and Volkamer~~et al.~~ (2013): The ratio of the peak values of the cross section at 360 and 380 nm changes in a non-continuous way between 253 and 23~~2~~3 K (see Fig. A27 in appendix A4), see also~~.~~ Fig. S2 (values for 380nm) in the supplementary material of Thalman and Volkamer (2013). The reason for this inconsistency is currently not known. If these two O_4 bands are included in the spectral analysis (as for the standard settings), the convergence of the spectral analysis strongly depends on the ability to fit both O_4 bands well. Thus the fit results for both O_4 cross sections are mainly determined by the relative strengths of both O_4 bands (see Fig. A27 in appendix A4). If instead a smaller wavelength range~~s~~ is used containing only one absorption band (345 – 374 nm), the derived O_4 (d)AMFs are in rather good agreement with the results of the analysis (using only the O_4 cross section for 293 K), see Table A25 in appendix A4. In that case, the convergence of the fit mainly depends on the temperature dependence of the line width. It should be noted that the non-continuous temperature dependence of the O_4 absorption cross section only affects the analysis of the synthetic spectra, because for the simulation of the spectra all O_4 cross sections for temperatures between 23~~2~~3 and 293 K were used. For the measured spectra, no problems are found, because in the spectral analysis only the O_4 cross sections for 23~~2~~3 and 293 K were used.

In Fig. A28 in appendix A4 the ratios of both fit coefficients (for 203 and 293 K) as well as the derived effective temperatures for the analyses of measured and synthetic spectra are shown. For the measured spectra the ratios are close to zero and the derived temperatures are close to 300K ~~for~~ most of the time (except in early morning and evening), because the effective atmospheric temperature for both days is close to the temperature of the high temperature O_4 cross section (293 K) (see Fig. 13). Similar results (at least around noon) are also obtained for the synthetic spectra if the narrow spectral range (345 – 374 nm) is used. For the standard fit range (including two O_4 bands), however, the ratios are much higher again indicating the effect of the inconsistency of the temperature dependence of the O_4 cross sections (see Fig. A27 in appendix A4).

4.3.6 Results from different instruments and analyses by different groups

In this section the effects of using measurements from different instruments and having these spectra analysed by different groups are investigated. For that purpose three different procedures are followed: First, MPIC spectra are analysed by other groups; second, the spectra from other instruments are analysed by MPIC~~non-MPIC instruments are analysed by the respective group~~; third, the spectra from non-MPIC instruments~~other instruments~~ are analysed by the respective group~~by MPIC~~.

In Fig. 14a and Table A25 (in appendix A4) the comparison results of the analysis of MPIC spectra by other groups versus the analysis of MPIC spectra by MPIC are shown. Especially for 18 June rather large differences (between –6% / +5%) to the MPIC standard analysis are found. Interestingly the largest differences are found in the morning when the aerosol

extinction close to the surface was strongest. On 8 July smaller differences (between –6% and –1%) are found.

In Fig. 14b and Table A25 (in appendix A4) the comparison results of the analysis of spectra from other instruments by MPIC versus the analysis of MPIC spectra by MPIC are shown. For this comparison all analyses are performed in the spectral range 335 – 374 nm, because the standard spectral range (352 – 387 nm) is not covered by all instruments. Again, the largest differences are found for 18 June (up to ±11%). For 8 July the differences reach up to ±6%, but for this day only a few measurements in the morning are available.

In Fig. 14c and Table A25 (in appendix A4) the comparison results of the analysis of spectra from other instruments by the respective group versus the MPIC analysis by MPIC (standard analysis) is shown. From this exercise the combined effects of different instrumental properties and retrievals can be estimated. Interestingly, the observed differences are only slightly larger than those for the analysis of the spectra from the different instruments by MPIC (Fig. 14b). This indicates that the largest ~~errors~~ uncertainties are related to the differences of the different instruments and not to the settings and implementations of the different retrievals. For the middle period of 18 June the uncertainties are within 12%. This range is also assumed for 8 July. Here it is interesting to note that the derived uncertainties of the spectral analysis are probably not representative for most recent measurement campaigns. For example, during the CINDI-2 campaign (http://www.tropomi.eu/data-products/cindi-2) the deviations of the O_4 spectral analysis results were much smaller than for the selected days during the MAD-CAT campaign (Kreher et al., 2019).

4.3.7 Summary of uncertainties of the O_4 AMF from the spectral analysis

Table 10 presents an overview on the different sources of uncertainties of the measured O_4 (d)AMFs obtained in the previous sub-sections. The uncertainties are expressed as relative deviations from the results for the standard settings (see Table 7) derived by MPIC from spectra of the MPIC instrument.

Like for the simulation results, in general, larger uncertainties are found for the O_4 dAMFs compared to the O_4 AMFs. This is expected because the uncertainties of the O_4 dAMFs contain the uncertainties of two analyses (at 90° elevation and at low elevation). Also, the uncertainties on 18 June are again larger than on 8 July. This finding was not expected, but is possibly related to the higher trace gas abundances (see Fig. 1 and Table A3 in appendix A1) and the higher aerosol extinction close to the surface on 18 June.

Another interesting finding is that the uncertainties of the spectral analysis of O_4 are dominated by the effect of instrumental properties up to ±12% in the morning of 18 June. Further important uncertainties are associated with the choice of the wavelength range, the degree of the polynomial and the intensity offset. In contrast, the exact choices of the trace gas cross sections (including their wavelength- and temperature dependencies) play only a minor role (up to a few percent). Excellent agreement (within ±1%) is in particular found for the O_4 analysis of the synthetic spectra using the standard settings and the directly simulated O_4 (d)AMFs at 360 nm. This indicates that the O_4 (d)AMFs retrieved in the wavelength range 352 – 387 nm are indeed representative for radiative transfer simulations at 360 nm.

As for the uncertainties of the simulated O_4 (d)AMFs, the uncertainties of the spectral analysis are also split into a systematic and a random term: the systematic deviations of the O_4 dAMFs from those of the standard settings are about +1% and –1.5% for 18 June and 8 July, respectively. The range of uncertainty is calculated from the uncertainty ranges of the different ~~error sources~~contributions by assuming that they are all independent. The random uncertainty ranges for 18 June and 8 July are calculated as ±12.5% and ±10.8%, respectively.

4.4 Recommendations derived from the sensitivity studies

In this section a short summary of the most important findings from the sensitivity studies is given.

Temperature and pressure profiles

Temperature and pressure profiles from sondes or model data should be used if available. Alternatively, temperature and pressure profiles extrapolated from surface measurements could be used. Typical uncertainties of the O_4 VCD derived from such profiles are still < 2%. For high temperatures (>20°C) the atmospheric humidity should be considered. If no measurements are available, prescribed profiles, e.g. from the US standard atmosphere or climatologies of temperature and pressure profiles can be used. However, depending on location and season the uncertainties of the resulting O_4 VCD can be rather large (see also Ortega et al., 2016).

Integration of the O_4 VCD

The integration should be performed on a vertical grid with at least 100 m resolution up to an altitude of 30 km. The surface altitude should be taken into account with an accuracy of at least 20 m.

Measurements and spectral analysis

Instruments should have a small FOV (≤1°), an accurate elevation calibration (better than 0.5°), and a small and preferably well characterised stray light level. For the data analysis the standard settings as provided in Table 7 should be used. From the analysis of synthetic spectra it was found that the results for these settings are consistent with simulated O_4 (d)AMFs within 1 %.

Information on aerosols

Aerosol profiles should be obtained from LIDARs or ceilometers using similar wavelengths as the MAX-DOAS measurements if available (see e.g. Ortega et al., 2016). Preferred LIDAR types are HSRL or Raman LIDARs, which directly provide profiles of aerosol extinction and thus need no assumptions on the LIDAR ratio. They should also have high signal to noise ratios and shallow blind region at the surface in order to cover a large altitude range. Information on aerosol optical properties and size distributions from sun photometers or in situ measurements should be used.

RTM simulations

Radiative transfer models should use Mie phase functions and aerosol single scattering albedo e.g. derived from sun photometer observations. The consideration of polarisation and rotational Raman scattering is not necessary.

In summary, if the optimised settings described above are used, the uncertainties of the radiative transfer simulations and spectral analysis can be largely reduced: the uncertainties of the O_4 dAMFs related to radiative transfer simulations can be reduced from about ±8 % as in this study to about ±4 %; those related to the spectral analysis can be reduced from about ±10 % to about ±6 %.

4.4.1 Preferred scenarios for future studies

In addition to the recommendations given above, future campaigns should aim to cover different meteorological conditions (e.g. low temperatures), viewing geometries (e.g. low

SZA), surface albedos (e.g. snow and ice) and wavelengths (e.g. 477, 577, and 630 nm). Also different aerosol scenarios including those with low aerosol optical depths should be covered. MAX-DOAS measurements should be performed by at least 2, preferably more instruments. In order to minimise the effects of instrumental properties, the instruments should be well calibrated and should have low straylight levels. Measurements during the CINDI-2 campaign are probably well suited for a similar study.

5 Comparison of measurements and simulations

The comparison results for both days are different: On 18 June (except in the evening) measurements and simulations agree within uncertainties (the ratio of simulated and measured O_4 dAMFs for the middle period of that day is 1.01±0.16). In contrast, on 8 July measurements and simulations significantly disagree: Taking into account the uncertainties of the VCD calculation (3%), the radiative transfer simulations (+16±6.4%) and the spectral analysis (-1.5±10.8%) for the middle period of that day results in a ratio of simulated and measured O_4 dAMFs of 0.82 ±0.10, which differs significantly from unity.

5.1 Important differences between both days

On both selected days similar aerosol AOD were measured. Also the diurnal variation of the SZA was similar because of the proximity to summer solstice. However, also many differences are found for the two days, which are discussed below.

a) temperature, pressure, wind:
On 18 June surface pressure was lower by about 13 hPa and surface temperature was higher by about 7K than on 8 July, respectively. These differences were explicitly taken into account in the calculation of the O_4 profiles / VCDs, the radiative transfer simulations and the interpretation of the spectral analyses. Thus they can very probably not explain the different comparison results on the two days.
On both days, wind was mainly blowing from East-North-East, but on 18 June it was blowing from West before about 08:00 and after 20:00 UTC. Wind speeds were lower on 18 June (between 1 and 2 m/s) than on 8 July (between 1 and 3 m/s).

b) aerosol properties:
The in situ aerosol measurements show very different abundances and properties of aerosols close to the ground for the selected days. On 18 June much higher concentrations of larger aerosol particles are found, which cannot be measured by the ceilometer due to the blindness for the lowest 180m. Thus it can be concluded that the enhanced aerosol concentration on 18 June is confined to a shallow layer at the surface. In general the aerosol concentrations close to the surface are more variable on 18 June than on 8 July. The high aerosol concentrations close to the surface probably also affect the LIDAR ratio, which is thus probably more variable on 18 June. Similarly, also the phase function derived from the sun photometer (for the integrated aerosol profile) is probably less representative for the low elevation angles on 18 June because different aerosol size distributions probably existed at different altitudes. Finally, the Ångström parameter derived from AERONET observations is different for both days, especially for large wavelengths, which is in qualitative agreement with the higher in situ aerosol concentrations of large particles on 18 June. Also a larger forward peak of the derived aerosol phase function is found for 18 June. Both effects probably cause larger uncertainties on 18 June.

c) spectral analysis

Larger uncertainties of the spectral analysis are found for 18 June compared to 8 July. This finding was surprising, but was also partly reproduced by the analysis of the synthetic spectra. One possible explanation is the smaller wavelength dependence of aerosol scattering at low altitudes on 18 June, which mainly affects measurements at low elevation angles. When analysed versus a zenith reference, for which the broad band wavelength dependency is much stronger (because of the larger contribution from Rayleigh scattering), larger deviations can be expected (e.g. because of differences of instrumental straylight, or the different detector saturation levels). On 18 June also higher (about doubled) NO_2 and HCHO concentrations are present compared to 8 July possibly leading to increased spectral interferences with the O_4 absorption, but this effect is expected to be small.

5.2 Which conditions would be needed to bring measurements and simulations on 8 July into agreement

This section tentatively describes possible (although generally unrealistic) changes of the atmospheric scenario, the instrument properties or the input parameters, which could bring measurements and simulations on 08 July into agreement. If e.g. the whole aerosol extinction profile was scaled by 0.65, the corresponding O_4 dAMFs would almost perfectly match the measured ones.

Similarly good agreement could also be achieved if about 27% of the total AOD would be shifted from low layers (below 1.68 km) to high layers (above 4.9 km, see appendix A6). However, in this scenario, about 73% of the total aerosol extinction would be above 1.68 km. Such a scenario would not be in agreement with the AERONET inversion products and would also lead to an underestimation of the diurnal variation of the O_4 AMFs measured in zenith direction.

Also horizontal gradients of the aerosol extinction could in principle explain the discrepancy. While we are not able to quantify them, they surely would have to be of the order of several ten percent per 10 km. Such persistent horizontal gradients are not supported by the almost constant AOD during the day (and also by the consistent aerosol in situ observations at the different sites). Also the finding that mismatch between measurements and simulations is found for all azimuth angles indicates that horizontal gradients can not explain the observed discrepancies.

Another possibility would be aerosol phase functions with very high asymmetry parameters (>> 0.75). Also systematic errors of the O_4 cross section could explain the observed discrepancies. Finally, an overcorrection of spectrograph straylight (or any other intensity offset) could explain the discrepancies. However, a rather high overcorrection (by about 20%) would be needed, which is probably unrealistic.

~~5~~ 6 ~~Discussion and c~~Conclusions

We compared MAX-DOAS observations of the atmospheric O_4 absorption with corresponding radiative transfer simulations for two mainly cloud-free days during the MAD-CAT campaign. A large part of this study is dedicated to the extraction of input information for the radiative transfer simulations and the quantification of the associated uncertainties of the radiative transfer simulations and spectral retrievals. An important result from the sensitivity studies is that the O_4 results derived from the analysis of synthetic spectra using the

standard settings are consistent with the simulated O_4 air mass factors within 1%. Also recommendations for the settings of the radiative transfer simulations, in particular on the extraction of aerosol and O_4 profiles are given. Another important result is that the extent and quality of the aerosol data sets is crucial to constrain the radiative transfer simulations. For example, it is recommended that LIDAR instruments are operated at wavelengths close to those of the MAX-DOAS measurements (see Ortega et al., 2016) and have a small sensitivity gap close to the surface. Further aerosol properties (e.g. size distributions, phase functions) should be available from sun photometer and/or in situ measurements. If such aerosol data are available the corresponding uncertainties of the radiative transfer simulations could be largely reduced to about ±5%. Similar uncertainties can also be expected for optimum instrument operations and data analyses.

The comparison results for both days are different: On 18 June (except in the evening) measurements and simulations agree within ~~errors~~ uncertainties (the a ratio of simulated and measured O_4 dAMFs for the middle period of that day is 1.01±0.16). In contrast, on 8 July measurements and simulations significantly disagree: Taking into account the ~~errors~~ uncertainties of the VCD calculation (3%), the radiative transfer simulations (+16±6.4~~1~~%) and the spectral analysis (-1.5±10.8%) for the middle period of that day results in a ratio of simulated and measured O_4 dAMFs of 0.~~71~~ 81 ±0.~~12~~10, which differs significantly from unity. So far no plausible explanation for the observed discrepancies on 8 July was found.

~~On 18 June larger uncertainties both for the measurements and radiative transfer simulations exist, mainly related to the high aerosol concentration close to the surface. A summary of the most important differences between both days is given in section 5.1.~~

~~A large part of this study was dedicated to the extraction of input information for the radiative transfer simulations and to the quantification of the errors of the radiative transfer simulations and spectral retrievals. In particular, the analysis of synthetic spectra indicated that the O_4 results derived from the spectral analysis using the standard settings are consistent with the simulated O_4 air mass factors within 1%.~~

~~Based on this study, also recommendations for similar future studies are derived (see section 5.2). In general, the largest errors sources arise from spectral analyses (partly related to imperfections of the MAX-DOAS instruments) and the uncertainties of the aerosol phase functions and extinction profiles. Even if the aerosol extinction profiles could be better constraint, e.g. using results from Raman LIDARs or high spectral resolution LIDARs (HSRL), the uncertainties of the aerosol phase function will remain a critical error source. Future measurements should in particular try to minimize these error sources. Here it should be noted that the general larger errors obtained for 18 June are probably not representative for typical measurement conditions. For example, during the CINDI-2 campaign (http://www.tropomi.eu/data-products/cindi-2) the deviations of the O_4 spectral analysis results were much smaller than those for 18 June.~~

~~The main conclusion from this study is that on one of the two selected days during the MADCAT campaign (08 July) a scaling factor (of about 0.71±0.12) is needed to bring measurements and forward model into agreement.~~ However, as long as the reason for this deviation is not understood, it is unclear how~~As long as the reason for this deviation is not understood, it is, however, unclear, how~~ representative these findings are for other measurements (e.g. from other platforms, at other locations/seasons, for other aerosol loads, and other wavelengths). Thus further studies spanning a larger variety of measurement conditions and also including other wavelengths are recommended. The MAX-DOAS measurements collected during the recent CINDI-2 campaign are probably well suited for that purpose.

~~**5.1 Important differences between both days**~~

On both selected days similar aerosol AOD were measured. Also the diurnal variation of the SZA was similar because of the proximity to summer solstice. However, also many differences are found for the two days, which are discussed below.

a) temperature, pressure, wind:
On 18 June surface pressure was lower by about 13 hPa and surface temperature was higher by about 7K than on 8 June, respectively. These differences were explicitly taken into account in the calculation of the O_4 profiles / VCDs, the radiative transfer simulations and the interpretation of the spectral analyses. Thus they can very probably not explain the different comparison results on the two days.
On both days, wind was mainly blowing from East-North-East, but on 18 June it was blowing from West before about 08:00 and after 20:00 UTC. Wind speeds were lower on 18 June (between 1 and 2 m/s) than on 8 July (between 1 and 3 m/s).

b) aerosol properties:
The in situ aerosol measurements show very different abundances and properties of aerosols close to the ground for the selected days. On 18 June much larger concentrations of larger aerosol particles are found, which cannot be measured by the ceilometer, because the lowest detecting altitude is 180m. Thus it can be concluded that the enhanced aerosol concentration on 18 June is confined to a shallow layer at the surface. In general the aerosol concentrations close to the surface are more variable on 18 June than on 8 July. The high aerosol concentrations close to the surface probably also affect the LIDAR ratio, which is thus more variable on 18 June. Since a constant LIDAR ratio is used for the extraction of the aerosol extinction profiles, also the uncertainties of the aerosol profile are probably larger on 18 June. Similarly, also the phase function derived from the sun photometer (for the integrated aerosol profile) is probably less representative for the low elevation angles on 18 June because different aerosol size distributions probably existed at different altitudes. Finally, the Ångström parameter derived from AERONET observations is different for both days, especially for large wavelengths, which is in qualitative agreement with the higher in situ aerosol concentrations of large particles on 18 June. Also a larger forward peak of the derived aerosol phase function is found for 18 June. Both effects probably cause larger uncertainties on 18 June.

c) spectral analysis
Larger uncertainties of the spectral analysis are found for 18 June compared to 8 July. This finding was surprising, but was also partly reproduced by the analysis of the synthetic spectra. One possible explanation is the smaller wavelength dependence of aerosol scattering at low altitudes on 18 June, which mainly affects measurements at low elevation angles. When analysed versus a zenith reference, for which the broad band wavelength dependency is much stronger (because of the larger contribution from Rayleigh scattering), larger deviations can be expected (e.g. because of differences of instrumental straylight, or the different detector saturation levels). On 18 June also higher (about doubled) NO_2 and HCHO concentrations are present compared to 8 July possibly leading to increased spectral interferences with the O_4 absorption, but this effect is expected to be small.

5.2 Recommendations

~~Based on the findings of this comparison study, recommendations for similar future studies are derived. Part of them are also of interest for the interpretation of O_4 measurements in general.~~

~~a) VCD calculation~~

~~Temperature and pressure profiles representative for individual days should be used. If such profiles are not available, also profiles extrapolated from surface measurements can be used. They are not 'perfect' but usually the associated errors are at the percent level. The vertical grid for the integration of the O_4 profile should not be coarser than 100m. The integration should be carried out up to an altitude of at least 30 km. The exact height of the instrument position needs to be taken into account.~~

~~b) Radiative transfer simulations~~

~~If available appropriate phase functions (e.g. from Mie calculations) should be used. Here it is important to note that even if appropriate asymmetry parameters are available, the often used HG parameterisation becomes very imprecise for forward scattering geometries.~~

~~c) Spectral analysis~~

~~The spectral range should cover the two O_4 bands at 360 and 380 nm. An intensity offset should be included in the analysis. If the surface temperature differs strongly (more than 25K) from 300K the effect of the temperature dependence of the O_4 absorption should be considered.~~

~~d) Preferred scenarios for future studies~~

~~In particular the uncertainties related to aerosols should be minimised. For example, measurements at rather low AOD (≤0.1) and with low temporal variability should be selected. Aerosol profiles should be derived from LIDARs/ceilomters which are sensitive down to very shallow altitudes (low overlap ranges). If possible, Raman LIDARs or high spectral-resolution LIDARs (HSRL) should be used, because from such observations the aerosol extinction profile can be derived without the assumption of a LIDAR ratio. Also sun photometer measurements should be available. Besides AOD and the Ångström parameter also information on the phase function and single scattering albedo from these measurements should be used.~~

~~It would be interesting to cover other meteorological conditions (e.g. low temperatures), viewing geometries (e.g. low SZA), surface albedos (e.g. snow and ice) and wavelengths (e.g. 477, 577, and 630 nm).~~

~~In order to minimise the effects of instrumental properties, the instruments should be well calibrated and should have low straylight levels. At least two instruments should be operated at the same site. Based on the above criteria, measurements during the CINDI-2 campaign are probably well suited for a similar study.~~

Acknowledgments

We are thankful for several external data sets which were used in this study: Temperature and pressure profiles from the ERAInterim reanalysis data set were provided by the European Centre for Medium-Range Weather Forecasts. In situ measurements of trace gas and aerosol concentrations as well as meteorological data were performed by the environmental monitoring services of the States of Rhineland-Palatinate and Hesse (http://www.luft-rlp.de and https://www.hlnug.de/themen/luft/luftmessnetz.html). We thank M. O. Andreae and

Günther Schebeske for operating the Ceilometer and the AERONET instrument at the Max Planck Institute for Chemistry. We also would like to thank Rainer Volkamer, Theodore K. Koenig and Ivan Ortega for very helpful comments.

Tables

Table 1 Overview on studies which did not apply a scaling factor (upper part) or did apply a scaling factor (lower part) to the measured O_4 dSCDs. Besides the initial studies proposing a scaling factor (Wagner et al., 2009; Clémer et al., 2010) only studies after 2010 are listed.

Reference	Measurement type	Location and period	O_4 band (nm)	Scaling factor
Studies which did not apply a scaling factor*				
Thalmann and Volkamer, 2010	CE-DOAS	Laboratory	477	1
Frieß et al., 2011	MAX-DOAS	Barrow, Alaska (Feb-Apr 2009)	360	1
Peters et al., 2012a	MAX-DOAS	Western Pacific Ocean (Oct 2009)	360, 477	1
Spinei et al. 2015	Direct sun DOAS	JPL, USA (Jul 2007) Pullman, USA (Sep – Nov 2007, Jul – Nov 2011) Fairbanks, USA (Mar-Apr 2011) Huntsville, USA (Aug 2008) Richland, USA (Apr-Jun 2008) Greenbelt, USA (May 2007, 2012-2014) Cabauw, The Netherlands (Jun-Jul 2009)	360, 477	1
Spinei et al., 2015 /	Airborne DOAS	Subtropical Pacific Ocean (Jan 2012)	360, 477	1
Volkamer et al., 2015	Airborne DOAS	Subtropical Pacific Ocean (Jan 2012)	360, 477	1

Ortega et al., 2016	MAX-DOAS	Cape Cod, USA (Jul 2012)	360, 477	1
Schreier et al., 2016	MAX-DOAS	Zugspitze, Germany (Apr-Jul 2003) Pico Espeio, Venezuela (2004 - 2009)	360	1
Seyler et al., 2017	MAX-DOAS	German Bight (2013-2016)	360, 477	1
Wang et al., 2017a,b	MAX-DOAS	Wuxi, China (2011 - 2014)	360	1
Gielen et al., 2017	MAX-DOAS	Bujumbura, Burundi (2013-2015)	360, 477	1
Franco et al., 2015	MAX-DOAS	Jungfraujoch (2010 –2012)	360	1
Studies which did apply a scaling factor				
Wagner et al., 2009	MAX-DOAS	Milano, Italy Sep 2013 (FORMAT II)	360	0.81
Clemer et al., 2010	MAX-DOAS	Beijing, China Jul 2008 – Apr 2009	360, 477, 577, 630	0.80
Irie et al., 2011	MAX-DOAS	Cabauw, The Netherlands Jul-Jun 2009 (CINDI-I)	360, 477	0.75±0.1
Merlaud et al., 2011	Airborne DOAS	Arctic Apr 2008 POLARCAT)	360	0.89
Vlemmix et al., 2011	MAX-DOAS	Cabauw, The Netherlands Jul-Oct 2009 (CINDI-I)	477	0.8
Zieger et al., 2011	Overview on MAX-DOAS	Cabauw, The Netherlands Jul-Oct 2009 (CINDI-I)	360 (MPIC) 477 (BIRA) 477 (IUPHD) 477 (JAMSTEC)	0.83 0.75 0.8 0.8*
Wang et al., 2014	MAX-DOAS	Xianghe, China (2010 - 2013)	360	0.8
Kanaya et al., 2014	MAX-DOAS	Cape Hedo, Japan (2007 – 2012) Fukue, Japan (2008 – 2012) Yokosuda, Japan (2007 – 2012) Gwangju, Korea (2008 – 2012) Hefei, China (2008 – 2012) Zvenigorod; Russia (2009 – 2012)	477 477 477 477 477 477	0.8 0.8 0.8 0.8 0.8 0.8
Hendrick et al., 2014	MAX-DOAS	Beijing, China (2008 - 2009) Xianghe, China (2010 – 2012)	360	0.8
Vlemmix et al., 2015	MAX-DOAS	Beijing, China (2008 - 2009) Xianghe, China (2010 – 2012)	360, 477	0.8
Irie et al., 2015	MAX-DOAS	Tsukuba, Japan (Oct 2010)	477	elevation dependent scaling factor**
Wang et al., 2016	MAX-DOAS	Madrid, Spain (Mar – Sep 2015)	360	0.83
Friess et al., 2016	MAX-DOAS	Cabauw, The Netherlands Jul-Jul 2009 (CINDI-I)	477 (AOIFM) 477 (BIRA) 477 (IUPHD) 477 (JAMSTEC) 360 (MPIC)	0.8 0.8 1 0.8*** 0.77

*The authors of part of these studies were probably not aware that a scaling factor was applied by other groups.
**SF = 1 / (1 + EA/60)
***SF is varied during profile inversion

Table 2 Periods on both selected days, which are used for the comparisons.

day	1st period	2nd period	3rd period
18 June 2013	8:00 – 11:00 UTC	11:00 – 14:00 UTC	14:00 – 19:00 UTC

8 July 2013	4:00 – 7:00 UTC	7:00 – 11:00 UTC	11:00 – 19:00 UTC

Table 3 Participation of the different groups in the different analysis steps

Abreviation	Institution	Determination of the O_4 profile and VCD	Extraction of aerosol profiles	Radiative transfer simulations	Spectral analysis
BIRA	BIRA/IASB, Brussels, Belgium				•
CMA	Meteorological Observation Center, Beijing, China			•	•
CSIC	Department of Atmospheric Chemistry and Climate, Institute of Physical Chemistry Rocasolano (CSIC), Spain.	•			•
INTA	Instituto Nacional de Tecnica Aeroespacial, Spain	•	•	•	•
IUP-B	University of Bremen, Germany		•	•	•
IUP-HD	University of Heidelberg, Germany				•
LMU	Ludwig-Maximilians-Universität München, Germany	•	•		
MPIC	MPI for chemistry, Mainz, Germany	•	•	•	•

Table 4 Overview on properties of MAX-DOAS instruments participating in this study

Institute / Instrument type	**Spectral range (nm)**	**Spectral resolution (FWHM, nm)**	**Spectral range per detector pixel (nm)**	**Detector type / temperature**	**Integration time of individual spectra (s)**	**Reference**
BIRA / 2-D scanning MAX-DOAS	300 - 386	0.49	0.04	2-D back-illuminated CCD, 2048 x 512 pixels / -40 °C	60	Clémer et al., 2010
IUP-Bremen / 2-D scanning MAX-DOAS	308 - 376	0.43	0.05	2-D back-illuminated CCD, 1340 x 400 pixels / -35 °C	20	Peters et al., 2012b
IUP-Heidelberg / 1-D scanning	294 - 459	0.59	0.09	AvaSpec-ULS 2048 pixels back-thinned Hamamatsu CCD	60	Lampel et al., 2015

MAX-DOAS				S11071-1106 / 20°C		
MPIC / 4-azimuth MAX-DOAS	320 – 457	0.67	0.14	2-D back-illuminated CCD, 1024 x 255 Pixels / -30°C	10 s	Krautwurst, 2010

Table 5 Independent data sets used to constrain the atmospheric properties during both selected days.

Measurement / data set	**Measured quantities**	**Derived quantities**	**Temporal / spatial resolution**	**Source / reference**
Ceilometer	Attenuated backscatter profiles* at 1064 nm	Aerosol extinction pofiles at 360 nm	30s** / 15 m	Wiegner and Geiß, 2012
AERONET sun photometer	Solar irradiances, Sky radiances	Aerosol optical depth, single scattering albedo, phase function	Typical integration time: 2 to 15 min	Holben et al., 2001, https://aeronet.gsfc.nasa.gov/
Surface measurements air quality stations in Mainz	temperature, pressure, rel. humidity		1h	http://www.luft-rlp.de

Mombach				
Surface measurements air quality stations in Mainz and Wiesbaden	$pm_{2.5}$ pm_{10}		1h (Mainz stations) 30 min (Wiesbaden stations)***	http://www.luft-rlp.de https://www.hlnug.de/themen/luft/luftmessnetz.html
ECMWF ERA-Interim reanalysis	temperature, Pressure, rel. humidity		Average over the area 49.41°-50.53° N, 7.88°-9.00° E, every 6 h	(Dee et al., 2011)

*no useful signal below 180m due to limited overlap
**Here 15 min averages are used.
***Stations in Mainz: Parcusstrasse, Zitadelle, Mombach; Stations in Wiesbaden: Schierstein, Ringkirche, Süd

Table 6 Standard settings for the radiative transfer simulations

Parameter	**Standard setting**
Temperature and pressure profile	MPIC extraction
O_4 profile	MPIC extraction
Surface albedo	5 %
Aerosol single scattering albedo	0.95
Aerosol phase function	HG model with asymmetry parameter of 0.68
Aerosol extinction profile	MPIC extraction with linear interpolation < 180 m
Polarisation	Not considered
Raman scattering	Partly considered for synthetic spectra

Table 7 Standard settings for the DOAS analysis of O_4.

Parameter	**Value, Remark / Reference**
Spectral range	352 – 387 nm
Degree of DOAS polynomial	5
Degree of intensity offset polynomial	2
Fraunhofer reference spectrum	08 July, 10:05:35, SZA: 32.37°, elevation angle: 90° (this spectrum is used for both days)
Wavelength calibration	Fit to high resolution solar spectrum using Gaussian slit function
Shift / squeeze	The measured spectrum is shifted and squeezed against all other spectra
Ring spectrum 1	Normal Ring spectrum calculated from DOASIS
Ring spectrum 2	Ring spectrum 1 multiplied by λ^{-4}
O_3 cross section	223 K, Bogumil et al. (2003)
NO_2 cross section	294 K, Vandaele et al. (1997)
BrO cross section	223 K, Fleischmann et al. (2004)
O_4 cross section	293 K, Thalman and Volkamer (2013)

Table 8 Average ratios (simulation results divided by measurements) of the O_4 (d)AMFs for both middle periods of the selected days.

Period	**18.06.2013, 11:00 – 14:00**	**08.07.2013, 7:00 – 11:00**
AMF ratio	0.97	0.83
DAMF ratio	0.94	0.69

Table 9 Summary of uncertainties of the simulated O_4 (d)AMFs for the middle periods of both selected days. The two numbers left and right of the '/' indicate the minimum and maximum deviations. The columns with label 'Optimum' indicate the uncertainties which could be reached if optimum information on the measurement conditions was available (e.g. height profiles of temperature, pressure and aerosol extinction as well as well aerosol microphysical or optical properties).

	O_4 AMF				O_4 dAMF		
	18 June	8 July	Optimum settings		18 June	8 July	Optimum settings
Effects of RTM							
Radiative transfer model	-1% / +2%	0% / +1%	±1%		-1% / +5%	0% / +3%	±1%
Polarisation	0% / 0%	0% / 0%	0%		0% / 0%	0% / +1%	0%
Effects of input parameters							
O_4 profile extraction	0% / + 2%	0% / + 1%	±1%		0% / + 4%	0% / + 2%	±1%
Single scattering albedo	-1% / + 3%	-1% / + 1%	0%		-1% / + 3%	-1% / + 1%	0%
Phase function	-3% / +3%	-2% / 0%	±1%		-5% / + 9%	-5% / +2%	±1.5%
Aerosol profile extraction	-1% / + 1%*	-2% / + 2%	±1%		-2% / + 1%*	-4% / + 4%	±1.5%
Extrapolation below 180 m	0% / + 2%	-1% / + 1%	0%		-1% / + 4%	-2% / + 2%	0%
LIDAR ratio & wrong wavelength	?	+5% / +6%	±2%**		?	+13% / +18%	±3%**
Surface albedo	0% / + 2%	0% / + 1%	0%		0% / + 2%	-1% / + 0%	0%
Total uncertainty							
Average deviation (from results for standard settings)	+4.5%	+~~0.5~~6%			+8.5%	+16.5%	
Range of uncertainty	±4.4%*	±2.8%	±2.8%**		±8.7%*	±6.~~1~~4%	±3.8%**

*this uncertainty does not contain the contribution from variation of aerosol properties with altitude, see text

**if LIDAR profiles at the same wavelength and without gaps in the troposphere were available.

Table 10 Summary of uncertainties of the measured O_4 (d)AMFs for the middle periods of both selected days. The two numbers left and right of the '/' indicate the minimum and maximum deviations. The columns with label 'Optimum' indicate the uncertainties which could be reached if optimum instrumental performance was ensured and optimum cross section were availble.

	O_4 AMF			O_4 dAMF		
	18 June	8 July	Optimum	18 June	8 July	Optimum
Consistency spectral analysis versus RTM						
Analysis of synthetic spectra	-1% / +1%	-1% / 0%	±1%	0% / 0%	0% / +1%	±1%
Fit settings						
Spectral range	-7% / -3%	-3% / 0%	±1%	-12% / -1%	-6% / -1%	±1%
Degree of polynomial	+0% / +4%	0% / + 3%	±1%	0% / +6%	0% / +6%	±1%
Intensity offset*	+1% / +5%	+1% / +3%	±1%	+3% / +11%	+2% / +4%	±1.5%
Ring	+1% / +2%	-1% / +1%	±1%	+1% / +1%	-1% / +1%	±1.5%
Temperature dependence of NO_2 absorption	0% / 0%	0% / 0%	0%	0% / 0%	0% / 0%	0% / 0%
Wavelength dependence of NO_2 absorption	-1% / 0%	0% / 0%	0%	-2% / -1%	-1% / 0%	0%
Wavelength dependence of O_4 absorption	-1% / 0%	-1% / -1%	0%	0% / +1%	-1% / -1%	0%
Including H_2O cross section	0% / 0%	0% / 0%	0%	+1% / +1%	+1% / +1%	0%
Including HCHO cross section	-3% / 0%	-1% / 0%	0%	-6% / -4%	-3% / -2%	0%
Different O_4 cross sections♣	-2% / +1%	-2% / +1%	±2%	-3% / +3%	-3% / +3%	±2%
Temperature dependence of the O_4 absorption						
Analysis using two O_4 cross sections for different temperatures♥	0% / 0%	+2% / +2%	±1%	+4% / +4%	+1% / +1%	±1.5%
Analysis of synthetic spectra	-1% / 0%	-1% / +2%		+4% / +4%	+1% / +1%	

for different surface temperatures							
Analysis from different instruments and groups							
Different groups and analyses♦	-6% / + 5%	-6% / + 5%	±3%♠		-12% / +7%	-12% / +7%	±4.5%
Total uncertainty							
Average deviation (from results for standard settings)	-4.5%	-0.5%			+1%	-1.5%	
Range of uncertainty	±7.0%	±6.5%	±4.2%		±12.5%	±10.8%	±5.7%

*here the case 'no offset' is not considered

♣here the case of the non-shifted Greenblatt O_4 cross section is not considered

♥here only the results for the measured spectra in the spectral range 352 – 387 nm are considered. (temperatures on 18 June: 27–31 °C; 8 July: 20–30 °C)

♦The results for 18 June are also taken for 8 July due to the lack of measurements on 8 July

♠see Kreher et al., 2019

References

Acarreta, J. R., De Haan, J. F., and Stammes, P.: Cloud pressure retrieval using the O2-O2 absorption band at 477 nm, J. Geophys. Res., 109, D05204, doi:10.1029/2003JD003915, 2004.

Bogumil, K., J. Orphal, T. Homann, S. Voigt, P.Spietz, O.C. Fleischmann, A. Vogel, M. Hartmann, H. Bovensmann, J. Frerik and J.P. Burrows, 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, J. Photochem. Photobiol. A., 157, 167-184, 2003.

Chance, K.V., and R.L. Kurucz, An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared, J. Quant. Spectrosc. Radiat. Transfer, 111, 1289-1295, 2010.

Chandrasekhar S. Radiative Transfer. New York: Dover Publications Inc.; 1960.

Chandrasekhar S. Selected papers, vol. 2. New York: University of Chicago Press, 1989.

Clémer, K., Van Roozendael, M., Fayt, C., Hendrick, F., Hermans, C., Pinardi, G., Spurr, R., Wang, P., and De Mazière, M.: Multiple wavelength retrieval of tropospheric aerosol optical properties from MAXDOAS measurements in Beijing, Atmos. Meas. Tech., 3, 863-878, doi:10.5194/amt-3-863-2010, 2010.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., H´olm, E. V., Isaksen, L., K°allberg, P., K¨ohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Th´epaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Deutschmann, T., Beirle, S., Frieß, U., Grzegorski, M., Kern, C., Kritten, L., Platt, U., Pukite, J., Wagner, T., Werner, B., and Pfeilsticker, K.: The Monte Carlo Atmospheric Radiative Transfer Model McArtim: Introduction and Validation of Jacobians and 3D Features, J. Quant. Spectrosc. Ra., 112, 1119–1137, doi:10.1016/j.jqsrt.2010.12.009, 2011.

Dubovik, O., Holben, B. N., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanr´e, D., and Slutsker, I.: Variability of absorption and optical properties of key aerosol types observed in worldwide locations, J. Atmos. Sci., 59, 590–608, 2002.

Erle F., K. Pfeilsticker, and U. Platt, On the influence of tropospheric clouds on zenith-scattered-light measurements of stratospheric species, *Geophys. Res. Lett., 22,* 2725- 2728, 1995.

Fleischmann, O. C., Hartmann, M., Burrows, J. P., and Orphal, J.: New ultraviolet absorption cross-sections of BrO at atmospheric temperatures measured by time-windowing Fourier transform spectroscopy, J. Photoch. Photobio. A, 168, 117–132, 2004.

Franco, B., Hendrick, F., Van Roozendael, M., Müller, J.-F., Stavrakou, T., Marais, E. A., Bovy, B., Bader, W., Fayt, C., Hermans, C., Lejeune, B., Pinardi, G., Servais, C., and Mahieu, E.: Retrievals of formaldehyde from ground-based FTIR and MAX-DOAS observations at the Jungfraujoch station and comparisons with GEOS Chem and IMAGES model simulations, Atmos. Meas. Tech., 8, 1733-1756, https://doi.org/10.5194/amt-8-1733-2015, 2015.

Frieß, F., P. S. Monks, J. J. Remedios, A. Rozanov, R. Sinreich, T. Wagner, and U. Platt, MAX-DOAS O4 measurements: A new technique to derive information on atmospheric aerosols. (II) Modelling studies, J. Geophys. Res., 111, D14203, doi:10.1029/2005JD006618.2006.

Frieß, U., Sihler, H., Sander, R., Pöhler, D., Yilmaz, S., and Platt, U.: The vertical distribution of BrO and aerosols in the Arctic: measurements by active and passive differential optical absorption sprectroscopy, J. Geophys. Res., 116, D00R04, doi:10.1029/2011JD015938, 2011.

Frieß, U., Klein Baltink, H., Beirle, S., Clémer, K., Hendrick, F., Henzing, B., Irie, H., de Leeuw, G., Li, A., Moerman, M. M., van Roozendael, M., Shaiganfar, R., Wagner, T., Wang, Y., Xie, P., Yilmaz, S., and Zieger, P.: Intercomparison of aerosol extinction profiles retrieved from MAX-DOAS measurements, Atmos. Meas. Tech., 9, 3205-3222, https://doi.org/10.5194/amt-9-3205-2016, 2016.

Gielen, C., Hendrick, F., Pinardi, G., De Smedt, I., Fayt, C., Hermans, C., Stavrakou, T., Bauwens, M., Müller, J.-F., Ndenzako, E., Nzohabonayo, P., Akimana, R., Niyonzima, S., Van Roozendael, M., and De Mazière, M.: Characterisation of Central-African aerosol and

trace-gas emissions based on MAX-DOAS measurements and model simulations over Bujumbura, Burundi, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2016-1104, in review, 2017.

Greenblatt G.D., Orlando, J.J., Burkholder, J.B., and Ravishankara, A.R.: Absorption measurements of oxygen between 330 and 1140 nm, J. Geophys. Res., 95, 18577-18582, 1990.

Hendrick, F., Van Roozendael, M., Kylling, A., Petritoli, A., Rozanov, A., Sanghavi, S., Schofield, R., von Friedeburg, C., Wagner, T., Wittrock, F., Fonteyn, D., and De Mazière, M.: Intercomparison exercise between different radiative transfer models used for the interpretation of ground-based zenith-sky and multi-axis DOAS observations, Atmos. Chem. Phys., 6, 93-108, doi:10.5194/acp-6-93-2006, 2006.

Hendrick, F., Müller, J.-F., Clémer, K., Wang, P., De Mazière, M., Fayt, C., Gielen, C., Hermans, C., Ma, J. Z., Pinardi, G., Stavrakou, T., Vlemmix, T., and Van Roozendael, M.: Four years of ground-based MAX-DOAS observations of HONO and NO_2 in the Beijing area, Atmos. Chem. Phys., 14, 765-781, https://doi.org/10.5194/acp-14-765-2014, 2014.

Heue, K.-P., Riede, H., Walter, D., Brenninkmeijer, C. A. M., Wagner, T., Frieß, U., Platt, U., Zahn, A., Stratmann, G., and Ziereis, H.: CARIBIC DOAS observations of nitrous acid and formaldehyde in a large convective cloud, Atmos. Chem. Phys., 14, 6621-6642, https://doi.org/10.5194/acp-14-6621-2014, 2014.

Hönninger, G., von Friedeburg, C., and Platt, U.: Multi Axis Differential Optical Absorption Spectroscopy (MAX-DOAS), Atmos. Chem. Phys., 4, 231–254, 2004.

Holben, B. N., Tanre, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., Newcomb, W. W., Schafer, J., Chatenet, B., Lavenue, F., Kaufman, Y. J., Vande Castle, J., Setzer, A., Markham, B., Clark, D., Frouin, R., Halthore, R., Karnieli, A., O'Neill, N. T., Pietras, C., Pinker, R. T., Voss, K., and Zibordi, G.: An emerging ground-based aerosol climatology: Aerosol Optical Depth from AERONET, J. Geophys. Res., 106, 12067–12097, 2001.

Irie, H., Kanaya, Y., Akimoto, H., Iwabuchi, H., Shimizu, A., and Aoki, K.: First retrieval of tropospheric aerosol profiles using MAX-DOAS and comparison with lidar and sky radiometer measurements, Atmos. Chem. Phys., 8, 341–350, doi:10.5194/acp-8-341-2008, 2008.

Irie, H., Takashima, H., Kanaya, Y., Boersma, K. F., Gast, L., Wittrock, F., Brunner, D., Zhou, Y., and Van Roozendael, M.: Eight-component retrievals from ground-based MAX-DOAS observations, Atmos. Meas. Tech., 4, 1027-1044, https://doi.org/10.5194/amt-4-1027-2011, 2011.

Irie, H., Nakayama, T., Shimizu, A., Yamazaki, A., Nagai, T., Uchiyama, A., Zaizen, Y., Kagamitani, S., and Matsumi, Y.: Evaluation of MAX-DOAS aerosol retrievals by coincident observations using CRDS, lidar, and sky radiometer inTsukuba, Japan, Atmos. Meas. Tech., 8, 2775-2788, https://doi.org/10.5194/amt-8-2775-2015, 2015.

Kanaya, Y., Irie, H., Takashima, H., Iwabuchi, H., Akimoto, H., Sudo, K., Gu, M., Chong, J., Kim, Y. J., Lee, H., Li, A., Si, F., Xu, J., Xie, P.-H., Liu, W.-Q., Dzhola, A., Postylyakov, O., Ivanov, V., Grechko, E., Terpugova, S., and Panchenko, M.: Long-term MAX-DOAS

network observations of NO_2 in Russia and Asia (MADRAS) during the period 2007–2012: instrumentation, elucidation of climatology, and comparisons with OMI satellite observations and global model simulations, Atmos. Chem. Phys., 14, 7909-7927, https://doi.org/10.5194/acp-14-7909-2014, 2014.

Kreher, K., M. Van Roozendael, F. Hendrick, A. Apituley, E. Dimitropoulou, U. Friess, A. Richter, T. Wagner, L. Ang, M. Anguas, A. Bais, N. Benavent, K. Bognar, A. Borovski, I. Bruchkovsky, A. Cede, K.L. Chan, S. Donner, T. Drosoglou, C. Fayt, H. Finkenzeller, N. Hao, C. Hermans, S. Hoque, H. Irie, J. Jin, P. Johnston, J. Khayyam Butt, F. Khokhar, T. Koenig, J. Ma, A. K. Mishra, M. Navarro-Comas, A. Pazmino, E. Peters, M. Pinharanda, A. Piters, O. Postylyakov, C. Prados, O. Rodriguez, R. Querel, A. Saiz-Lopez, S. Schreier, A. Seyler, E. Spinei, K. Strong, M. Tiefengraber, J.-L. Tirpitz4, V. Kumar, R. Volkamer, M. Wenig, P. Xie, J. Xu, M. Yela, X. Zhao, W. Zhuoru, Intercomparison of NO_2, O_4, O_3 and HCHO slant column measurements by MAX-DOAS and zenith-sky UV-Visible spectrometers, to be submitted to Atmos. Meas. Tech, 2019.

Krautwurst, S.: Charakterisierung eines neu aufgebauten MAXDOAS-Systems und Interpretation von ersten Messergebnissen zu dem Spurenstoff NO2, Diplomarbeit, Fachhochschule Coburg, Coburg, Germany, 2010.

Lampel, J., Frieß, U., and Platt, U.: The impact of vibrational Raman scattering of air on DOAS measurements of atmospheric trace gases, Atmos. Meas. Tech., 8, 3767–3787, https://doi.org/10.5194/amt-8-3767-2015, 2015.

Lampel, J., Pöhler, D., Polyansky, O. L., Kyuberis, A. A., Zobov, N. F., Tennyson, J., Lodi, L., Frieß, U., Wang, Y., Beirle, S., Platt, U., and Wagner, T.: Detection of water vapour absorption around 363 nm in measured atmospheric absorption spectra and its effect on DOAS evaluations, Atmos. Chem. Phys., 17, 1271-1295, https://doi.org/10.5194/acp-17-1271-2017, 2017.

Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., Van Roozendael, M., Wang, Y., Wagner, T., Beirle, S., Lin, J.-T., Krotkov, N., Stammes, P., Wang, P., Eskes, H. J., and Krol, M.: Structural uncertainty in air mass factor calculation for NO_2 and HCHO satellite retrievals, Atmos. Meas. Tech., 10, 759-782, https://doi.org/10.5194/amt-10-759-2017, 2017.

Meller, R. and G. K. Moortgat, Temperature dependence of the absorption cross sections of formaldehyde between 223 and 323 K in the wavelength range 225-375 nm, J. Geophys. Res., 105, 7089-7101, 2000.

Merlaud, A., Van Roozendael, M., Theys, N., Fayt, C., Hermans, C., Quennehen, B., Schwarzenboeck, A., Ancellet, G., Pommier, M., Pelon, J., Burkhart, J., Stohl, A., and De Mazière, M.: Airborne DOAS measurements in Arctic: vertical distributions of aerosol extinction coefficient and NO2 concentration, Atmos. Chem. Phys., 11, 9219–9236, doi:10.5194/acp-11-9219-2011, 2011.

Ortega, I., Berg, L. K., Ferrare, R. A., Hair, J. W., Hostetler, C. A., and Volkamer, R.: Elevated aerosol layers modify the O2-O2 absorption measured by ground-based MAX-DOAS, J. Quant. Spectrosc. Ra., 176, 34–49, doi:10.1016/j.jqsrt.2016.02.021, 2016.

Paur, R. J. and Bass, A. M.: The Ultraviolet Cross-Sections of Ozone: II. Results and temperature dependence, in: Atmospheric ozone; Proc. Quadrennial Ozone Symposium, edited by: Zeferos, C. S. and Ghazi, A., Halkidiki Greece, 1984, Dordrecht: Reidel, D., 611–615, 1984.

Peters, E., Wittrock, F., Großmann, K., Frieß, U., Richter, A., and Burrows, J. P.: Formaldehyde and nitrogen dioxide over the remote western Pacific Ocean: SCIAMACHY and GOME-2 validation using ship-based MAX-DOAS observations, Atmos. Chem. Phys., 12, 11179-11197, https://doi.org/10.5194/acp-12-11179-2012, 2012a.

Peters, E., Wittrock, F., Großmann, K., Frieß, U., Richter, A., and Burrows, J. P.: Formaldehyde and nitrogen dioxide over the remote western Pacific Ocean: SCIAMACHY and GOME-2 validation using ship-based MAX-DOAS observations, Atmos. Chem. Phys., 12, 11179–11197, https://doi.org/10.5194/acp-12-11179-2012, 2012b.

Pinardi, G., Van Roozendael, M., Abuhassan, N., Adams, C., Cede, A., Clémer, K., Fayt, C., Frieß, U., Gil, M., Herman, J., Hermans, C., Hendrick, F., Irie, H., Merlaud, A., Navarro Comas, M., Peters, E., Piters, A. J. M., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vrekoussis, M., Wagner, T., Wittrock, F., and Yilmaz, S.: MAX-DOAS formaldehyde slant column measurements during CINDI: intercomparison and analysis improvement, Atmos. Meas. Tech., 6, 167-185, https://doi.org/10.5194/amt-6-167-2013, 2013.

Polyansky, O.L., A.A. Kyuberis, N.F. Zobov, J. Tennyson, S.N. Yurchenko and L. Lodi ExoMol molecular line lists XXX: a complete high-accuracy line list for water, Mon. Not. R. astr. Soc., submitted, 2018.

Prados-Roman, C., Butz, A., Deutschmann, T., Dorf, M., Kritten, L., Minikin, A., Platt, U., Schlager, H., Sihler, H., Theys, N., Van Roozendael, M., Wagner, T., and Pfeilsticker, K.: Airborne DOAS limb measurements of tropospheric trace gas profiles: case studies on the profile retrieval of O_4 and BrO, Atmos. Meas. Tech., 4, 1241-1260, doi:10.5194/amt-4-1241-2011, 2011.

Puķīte, J., Kühl, S., Deutschmann, T., Platt, U., and Wagner, T.: Extending differential optical absorption spectroscopy for limb measurements in the UV, Atmos. Meas. Tech., 3, 631-653, doi:10.5194/amt-3-631-2010, 2010.

Schreier, S. F., Richter, A., Wittrock, F., and Burrows, J. P.: Estimates of free-tropospheric NO_2 and HCHO mixing ratios derived from high-altitude mountain MAX-DOAS observations at midlatitudes and in the tropics, Atmos. Chem. Phys., 16, 2803-2817, https://doi.org/10.5194/acp-16-2803-2016, 2016.

Serdyuchenko, A., Gorshelev, V., Weber, M., Chehade, W., and Burrows, J. P.: 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, 2014.

Seyler, A., Wittrock, F., Kattner, L., Mathieu-Üffing, B., Peters, E., Richter, A., Schmolke, S., and Burrows, J. P.: Monitoring shipping emissions in the German Bight using MAX-DOAS measurements, Atmos. Chem. Phys., 17, 10997-11023, https://doi.org/10.5194/acp-17-10997-2017, 2017.

Sneep, M., de Haan, J. F., Stammes, P., Wang, P., Vanbauce, C., Joiner, J., Vasilkov, A. P., and Levelt, P. F.: Three-way comparison between OMI and PARASOL cloud pressure products, J. Geophys. Res., 113, D15S23, doi:10.1029/2007JD008694, 2008.

Solomon, S., A. L. Schmeltekopf, and R. W. Sanders, On the interpretation of zenith sky absorption measurements, *J. Geophys. Res, 92,* 8311-8319, 1987.

Spinei, E., Cede, A., Herman, J., Mount, G. H., Eloranta, E., Morley, B., Baidar, S., Dix, B., Ortega, I., Koenig, T., and Volkamer, R.: Ground-based direct-sun DOAS and airborne MAX-DOAS measurements of the collision-induced oxygen complex, O_2O_2, absorption with significant pressure and temperature differences, Atmos. Meas. Tech., 8, 793-809, https://doi.org/10.5194/amt-8-793-2015, 2015.

Spurr RJD, Kurosu TP, Chance KV. A linearized discrete ordinate radiative transfer model for atmospheric remote sensing retrieval. JQSRT 2001;68:689–735.

Spurr, R., LIDORT and VLIDORT: Linearized Pseudo-Spherical Scalar and Vector Discrete Ordinate Radiative Transfer Models for Use in Remote Sensing Retrieval Problems, Light Scattering Reviews, Vol. 3, edited by: Kokhanovsky, A., Springer, Berlin Heidelberg, Germany, 2008.

Stamnes K, Tsay S-C, Wiscombe W, Jayaweera K. Numerically stable algorithm for discrete ordinate method radiative transfer in multiple scattering and emitting layered media. Appl Opt 1988; 27:2502-9.

Thalman, R. and Volkamer, R.: Inherent calibration of a blue LED-CE-DOAS instrument to measure iodine oxide, glyoxal, methyl glyoxal, nitrogen dioxide, water vapour and aerosol extinction in open cavity mode, Atmos. Meas. Tech., 3, 1797-1814, 2010.

Thalman, R. and Volkamer, R.: Temperature dependent absorption cross-sections of O_2–O_2 collision pairs between 340 and 630 nm and at atmospherically relevant pressure, Phys. Chem. Chem. Phys., 15, 15371, doi:10.1039/c3cp50968k, 2013.

United States Committee on Extension to the Standard Atmosphere: U.S. Standard Atmosphere, 1976, National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, United States Air Force, Washington D.C., 1976.

Vandaele, A. C., C. Hermans, P. C. Simon, M. Carleer, R. Colin, S. Fally, M.-F. Mérienne, A. Jenouvrier, and B. Coquart, Measurements of the NO2 Absorption Cross-section from 42000 cm-1 to 10000 cm-1 (238-1000 nm) at 220 K and 294 K, J. Quant. Spectrosc. Radiat. Transfer, 59, 171-184, 1997.

Vlemmix, T., Piters, A. J. M., Berkhout, A. J. C., Gast, L. F. L., Wang, P., and Levelt, P. F.: Ability of the MAX-DOAS method to derive profile information for NO_2: can the boundary layer and free troposphere be separated?, Atmos. Meas. Tech., 4, 2659-2684, https://doi.org/10.5194/amt-4-2659-2011, 2011.

Vlemmix, T., Hendrick, F., Pinardi, G., De Smedt, I., Fayt, C., Hermans, C., Piters, A., Wang, P., Levelt, P., and Van Roozendael, M.: MAX-DOAS observations of aerosols, formaldehyde and nitrogen dioxide in the Beijing area: comparison of two profile retrieval approaches, Atmos. Meas. Tech., 8, 941-963, https://doi.org/10.5194/amt-8-941-2015, 2015.

Volkamer, R., Baidar, S., Campos, T. L., Coburn, S., DiGangi, J. P., Dix, B., Eloranta, E. W., Koenig, T. K., Morley, B., Ortega, I., Pierce, B. R., Reeves, M., Sinreich, R., Wang, S., Zondlo, M. A., and Romashkin, P. A.: Aircraft measurements of BrO, IO, glyoxal, NO_2, H_2O, O_2–O_2 and aerosol extinction profiles in the tropics: comparison with aircraft-/ship-based in situ and lidar measurements, Atmos. Meas. Tech., 8, 2121-2148, 2015.

Wagner, T., F. Erle, L. Marquard, C. Otten, K. Pfeilsticker, T. Senne, J. Stutz, and U. Platt, Cloudy sky optical paths as derived from differential optical absorption spectroscopy observations, *J. Geophys. Res., 103, 25307-25321, 1998.*

Wagner, T., B. Dix, C.v. Friedeburg, U. Frieß, S. Sanghavi, R. Sinreich, and U. Platt MAX-DOAS O4 measurements – a new technique to derive information on atmospheric aerosols. (I) Principles and information content, J. Geophys. Res., 109, doi: 10.1029/2004JD004904, 2004.

Wagner, T., J. P. Burrows, T. Deutschmann, B. Dix, C. von Friedeburg, U. Frieß, F. Hendrick, K.-P. Heue, H. Irie, H. Iwabuchi, Y. Kanaya, J. Keller, C. A. McLinden, H. Oetjen, E. Palazzi, A. Petritoli, U. Platt, O. Postylyakov, J. Pukite, A. Richter, M. van Roozendael, A. Rozanov, V. Rozanov, R. Sinreich, S. Sanghavi, F. Wittrock, Comparison of Box-Air-Mass-Factors and Radiances for Multiple-Axis Differential Optical Absorption Spectroscopy (MAX-DOAS) Geometries calculated from different UV/visible Radiative Transfer Models, Atmos. Chem. Phys., 7, 1809-1833, 2007.

Wagner, T., Deutschmann, T., and Platt, U.: Determination of aerosol properties from MAX-DOAS observations of the Ring effect, Atmos. Meas. Tech., 2, 495-512, 2009.

Wagner, T., Beirle, S., Deutschmann, T., and Penning de Vries, M.: A sensitivity analysis of Ring effect to aerosol properties and comparison to satellite observations, Atmos. Meas. Tech., 3, 1723-1751, doi:10.5194/amt-3-1723-2010, 2010.

Wang, T., Hendrick, F., Wang, P., Tang, G., Clémer, K., Yu, H., Fayt, C., Hermans, C., Gielen, C., Müller, J.-F., Pinardi, G., Theys, N., Brenot, H., and Van Roozendael, M.: Evaluation of tropospheric SO_2 retrieved from MAX-DOAS measurements in Xianghe, China, Atmos. Chem. Phys., 14, 11149-11164, https://doi.org/10.5194/acp-14-11149-2014, 2014.

Wang, S., Cuevas, C. A., Frieß, U., and Saiz-Lopez, A.: MAX-DOAS retrieval of aerosol extinction properties in Madrid, Spain, Atmos. Meas. Tech., 9, 5089-5101, https://doi.org/10.5194/amt-9-5089-2016, 2016.

Wang, Y., Beirle, S., Lampel, J., Koukouli, M., De Smedt, I., Theys, N., Li, A., Wu, D., Xie, P., Liu, C., Van Roozendael, M., Stavrakou, T., Müller, J.-F., and Wagner, T.: Validation of OMI, GOME-2A and GOME-2B tropospheric NO_2, SO_2 and HCHO products using MAX-DOAS observations from 2011 to 2014 in Wuxi, China: investigation of the effects of priori profiles and aerosols on the satellite products, Atmos. Chem. Phys., 17, 5007-5033, https://doi.org/10.5194/acp-17-5007-2017, 2017a.

Wang, Y., Lampel, J., Xie, P., Beirle, S., Li, A., Wu, D., and Wagner, T.: Ground-based MAX-DOAS observations of tropospheric aerosols, NO2, SO2 and HCHO in Wuxi, China,

from 2011 to 2014, Atmos. Chem. Phys., 17, 2189–2215, doi:10.5194/acp-17-2189-2017, 2017b.

Wang, Y., Beirle, S., Hendrick, F., Hilboll, A., Jin, J., Kyuberis, A. A., Lampel, J., Li, A., Luo, Y., Lodi, L., Ma, J., Navarro, M., Ortega, I., Peters, E., Polyansky, O. L., Remmers, J., Richter, A., Puentedura, O., Van Roozendael, M., Seyler, A., Tennyson, J., Volkamer, R., Xie, P., Zobov, N. F., and Wagner, T.: MAX-DOAS measurements of HONO slant column densities during the MAD-CAT campaign: inter-comparison, sensitivity studies on spectral analysis settings, and error budget, Atmos. Meas. Tech., 10, 3719-3742, https://doi.org/10.5194/amt-10-3719-2017, 2017c.

Wiegner, M. and Geiß, A.: Aerosol profiling with the Jenoptik ceilometer CHM15kx, Atmos. Meas. Tech., 5, 1953-1964, https://doi.org/10.5194/amt-5-1953-2012, 2012.

Winterrath, T., T. P. Kurosu, A. Richter and J. P. Burrows, Enhanced O3 and NO2 in thunderstorm clouds: convection or production?, Geophys. Res. Lett., No. 26, pp. 1291-1294, 1999.

Wittrock, F., H. Oetjen, A. Richter, S. Fietkau, T. Medeke, A. Rozanov, J. P. Burrows MAX-DOAS measurements of atmospheric trace gases in Ny-Ålesund - Radiative transfer studies and their application, Atmos. Chem. Phys., 4, 955-966, 2004

Zieger, P., Weingartner, E., Henzing, J., Moerman, M., de Leeuw, G., Mikkilä, J., Ehn, M., Petäjä, T., Clémer, K., van Roozendael, M., Yilmaz, S., Frieß, U., Irie, H., Wagner, T., Shaiganfar, R., Beirle, S., Apituley, A., Wilson, K., and Baltensperger, U.: Comparison of ambient aerosol extinction coefficients obtained from in-situ, MAX-DOAS and LIDAR measurements at Cabauw, Atmos. Chem. Phys., 11, 2603-2624, https://doi.org/10.5194/acp-11-2603-2011, 2011.

Figures

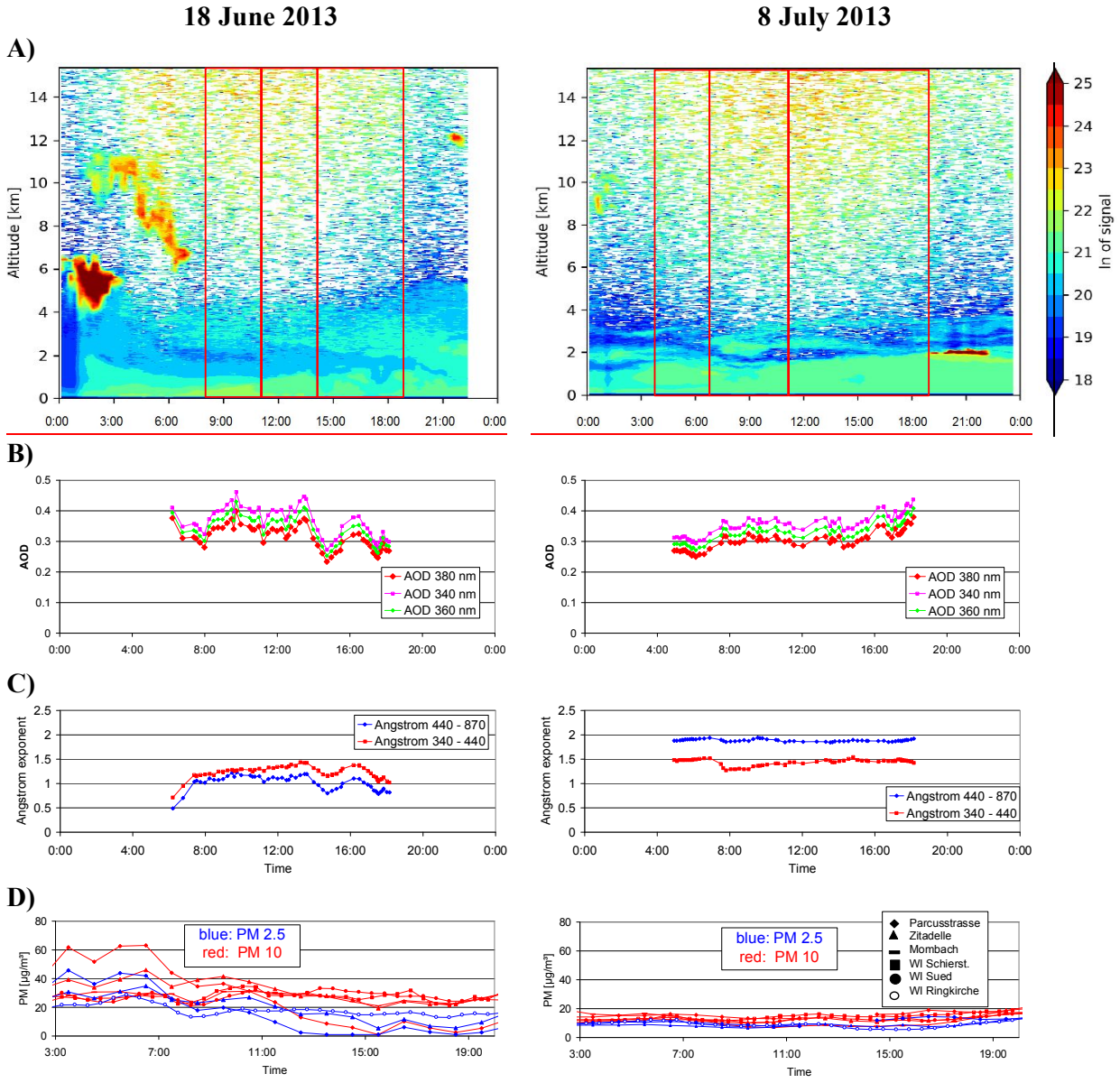


Fig. 1 Various aerosol properties on the two selected days (left: 18 June 2013; right: 8 July 2013). A) Aerosol backscatter profiles from ceilometer measurements; B) AOD at 340, 360, and 380 nm (360 values are interpolated from 340 and 380 nm) from AERONET sun photometer measurements; C) Ångström parameters for two wavelength pairs (340 – 440 nm and 440 – 870 nm) from AERONET sun photometer measurements; D) Surface in situ measurements of $PM_{2.5}$ and PM_{10} measured at different air quality monitoring stations in Mainz and the nearby city of Wiesbaden .

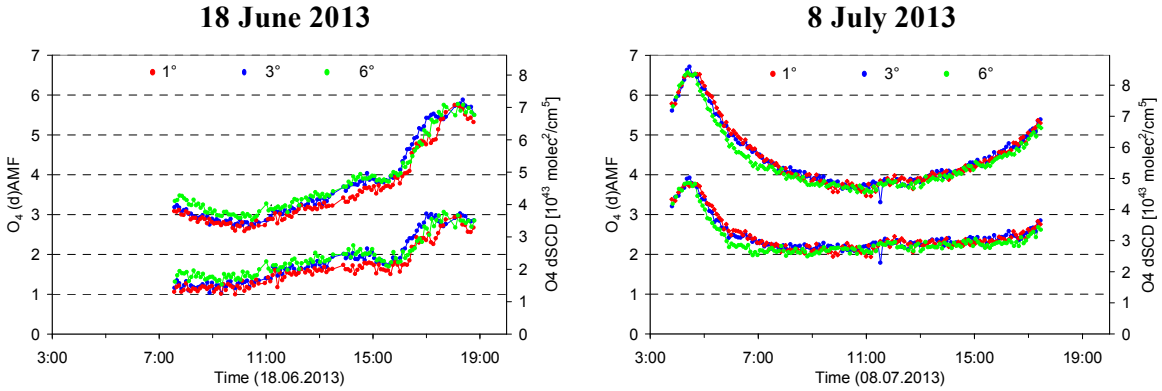


Fig. 2 O_4 AMFs (upper lines) and dAMFs (lower lines) for 1°, 3°, and 6° elevation angles derived from the MPIC MAX-DOAS measurements on the two selected days. Interestingly, on 18 June the lowest values are in general found for the lowest elevation angles, which is an indication for the high aerosol load close to the surface. The y-axis on the right side shows the corresponding O_4 (d)SCDs for O_4 VCDs of $1.23 \cdot 10^{43}$ molec²/cm^5 and of $1.28 \cdot 10^{43}$ molec²/cm^5 for 18 June and 08 July, respectively (see section 4.1.2).

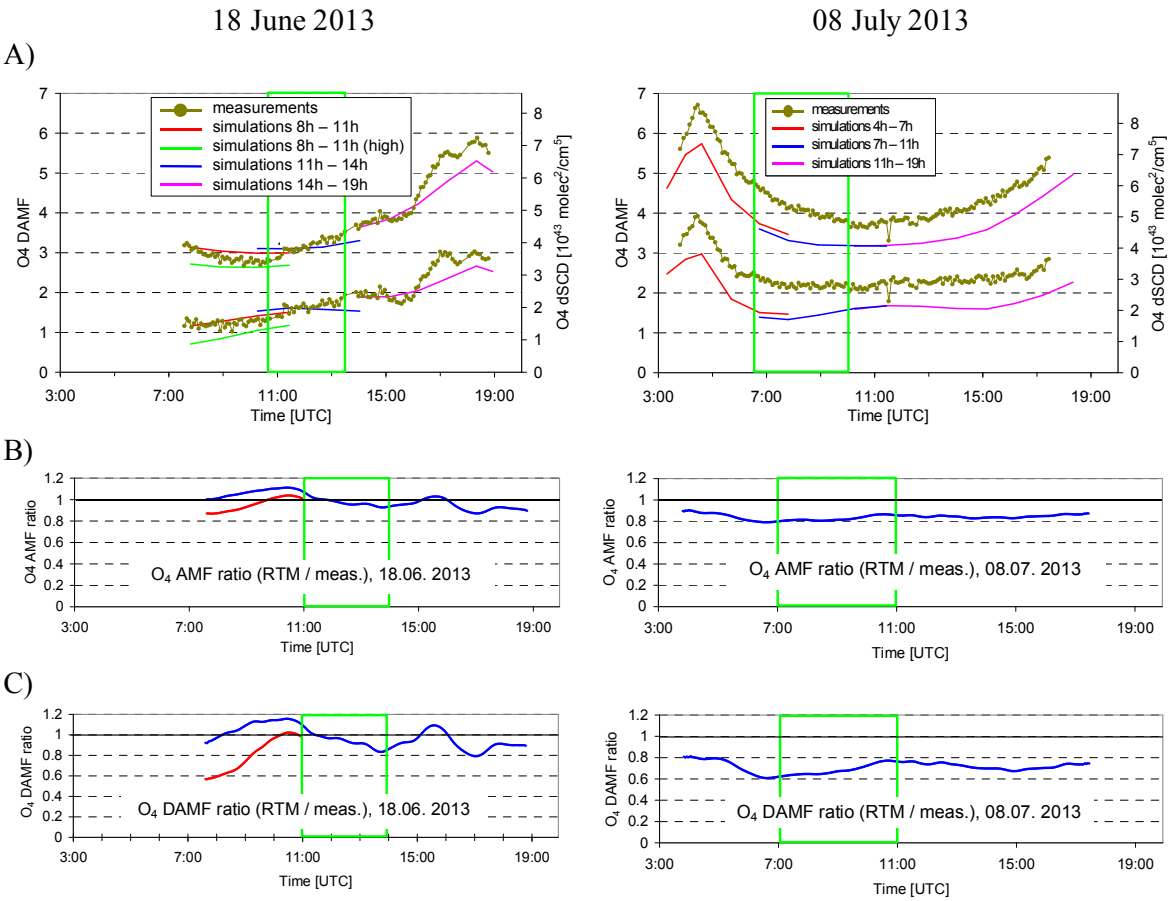


Fig. 3 A) Comparison of O_4 (d)AMFs from MAX-DOAS measurements and forward model simulations for the two selected days. The green rectangle indicates the middle periods on both days, which are the focus of the quantitative comparison. The green line on 18 June

represents forward model results for a modified aerosol profile (see text). The y-axis on the right side shows the corresponding O_4 (d)SCDs for O_4 VCDs of $1.23 \cdot 10^{43}$ molec²/cm^5 and of $1.28 \cdot 10^{43}$ molec²/cm^5 for 18 June and 08 July, respectively (see section 4.1.2). In B) and C) the ratios of the simulated and measured AMFs and dAMFs are shown, respectively. The red line on 18 June represents the ratios for the modified aerosol scenario.

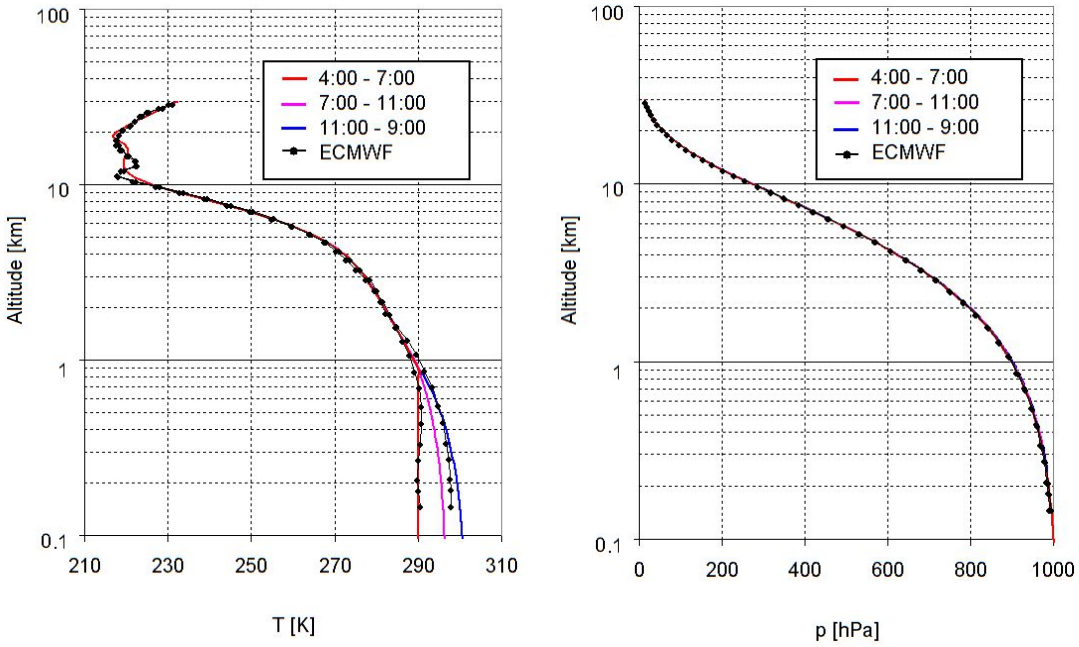


Fig. 4 Extracted temperature (left) and pressure (right) profiles for the three periods on 8 July 2013. Also shown are ECMWF profiles above Mainz for 6:00 and 18:00. To better account for the diurnal variation of the temperatures near the surface, below 1 km the temperature is linearly interpolated between the surface measurements and the ECMWF temperatures at 1 km (for details see text). Note that the altitude is given relative to the height of the measurement site (150 m).

18 June 14:00 – 19:00 **8 July 4:00 – 7:00**

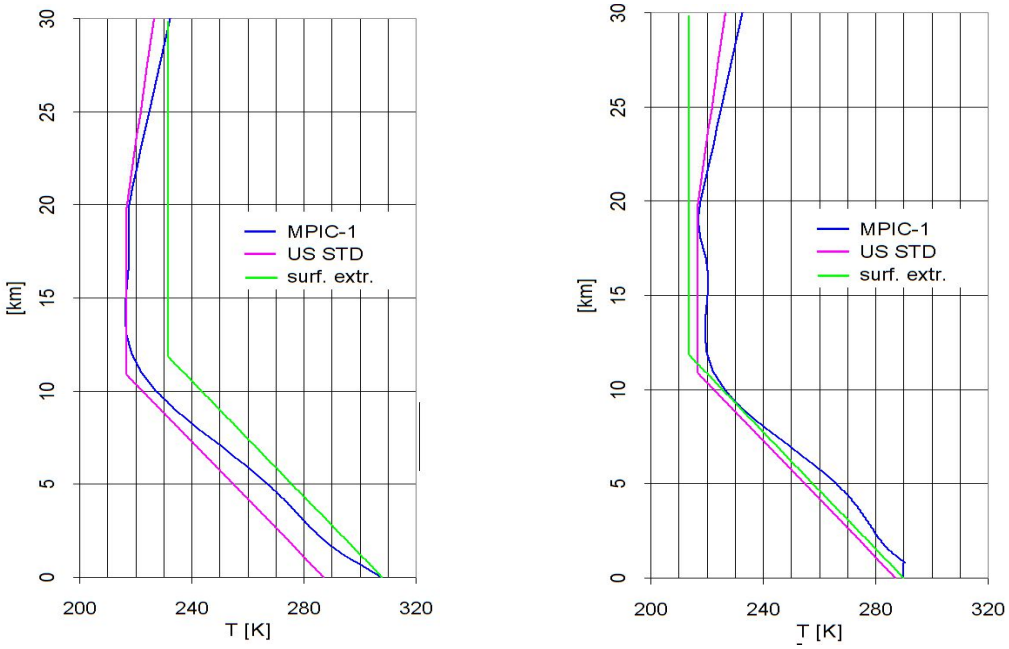


Fig. 5 Temperature profiles extracted in different ways for two periods (Left: 18 June 14:00 – 19:00; right: 8 July 4:00 – 7:00). The blue profiles are extracted from in situ measurements and ECMWF profiles as described in the text. The green profiles are extracted from the surface temperatures and assuming a constant lapse rate of –6.5K / km up to 12 km and a constant temperature above. The pink curves represent the temperature profile from the US standard atmosphere.

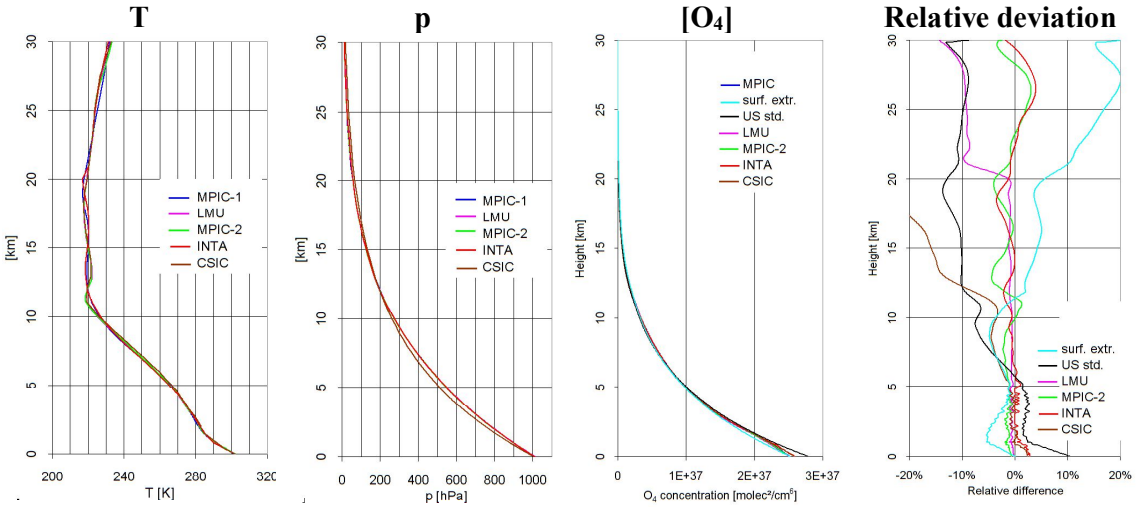


Fig. 6 Comparison of the vertical profiles of temperature, pressure and O_4 concentration (expressed as the square of the O_2 concentration) for 8 July, 11:00 – 19 :00, extracted by the different groups. In the right figure the relative deviations of the O_4 concentration compared to the MPIC standard extraction are shown. There, also the profiles derived from the extrapolation from the surface values and the US standard atmosphere are included.

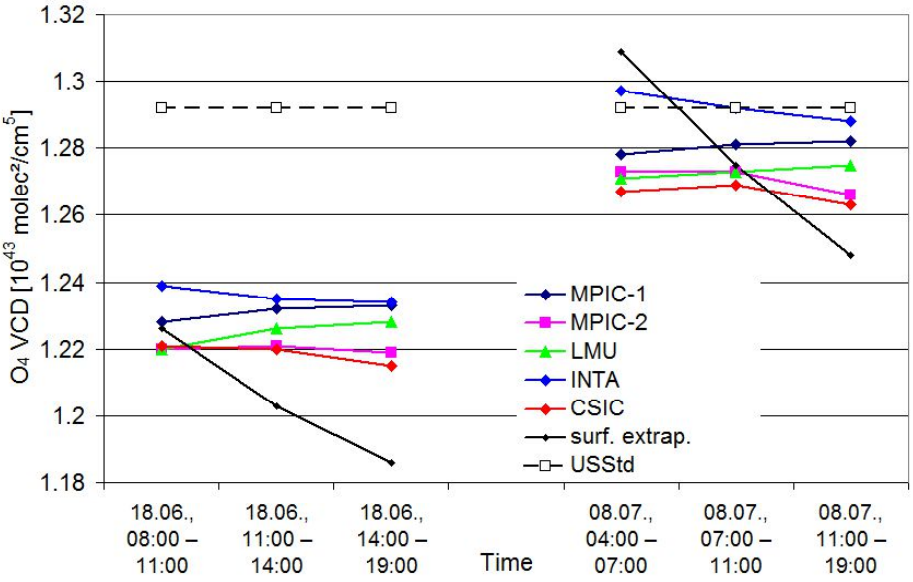


Fig. 7 Comparison of the O_4 VCDs for the selected periods on both days calculated from the profiles extracted by the different groups. Also the results for the profiles extrapolated from the surface values and the US standard atmosphere are shown.

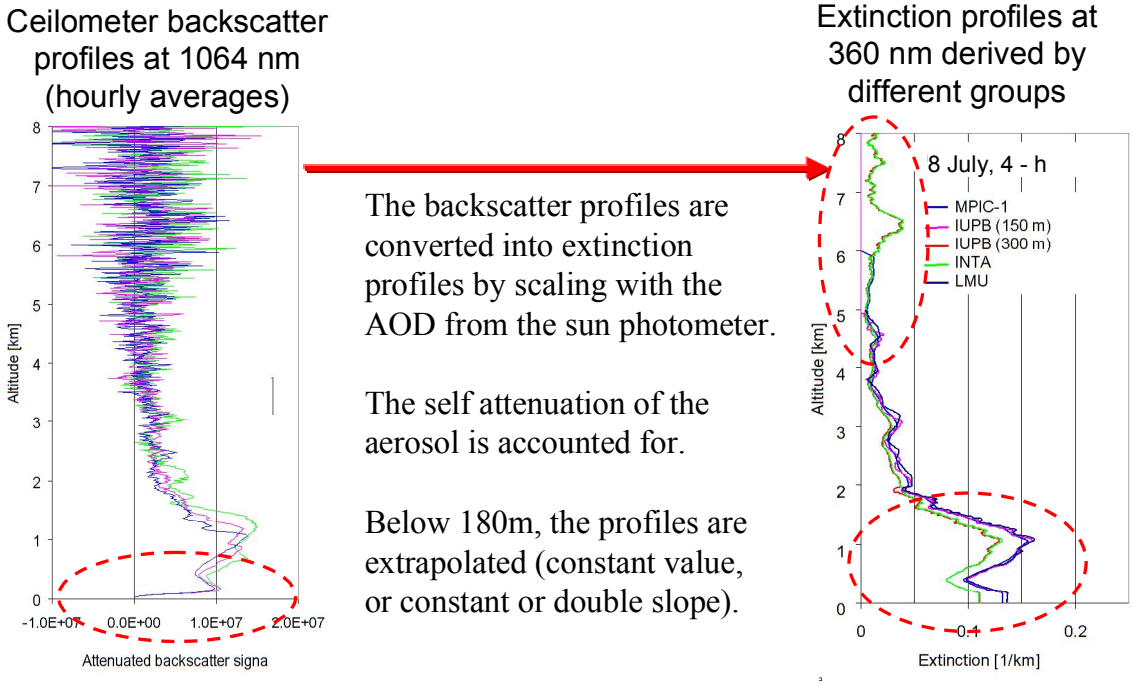
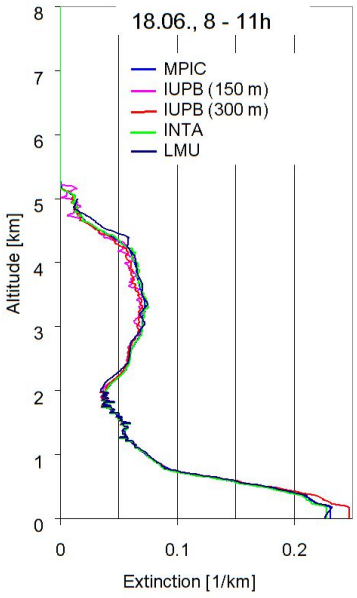


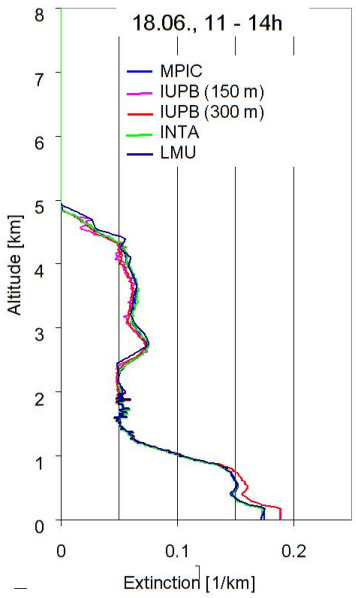
Fig. 8 Left: Hourly averaged backscatter profiles from the ceilometer measurements for the period 4:00 – 7:00 on 8 July 2013. Below 180 m the values rapidly decrease to zero due to the

missing overlap between the outgoing beam and the field of view of the telescope. Right: Aerosol extinction profiles extracted by the different groups from the ceilometer profiles (assuming a constant extinction below 180 m). The red circles indicate the height intervals with the larges deviations (IUPB 150 m and IUPB 300 m indicate profile extractions with different widths of the smoothing kernels: Hanning windows of 150 and 300 m, respectively).

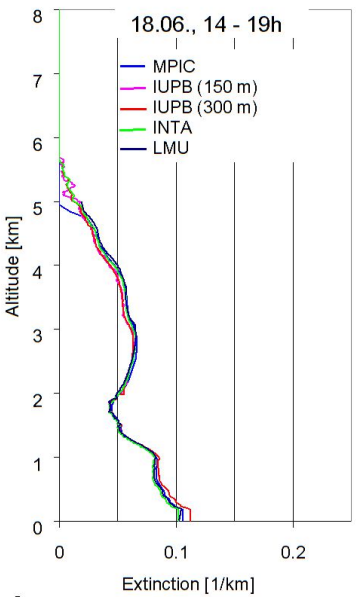
18 June 08:00 to 11:00



18 June 11:00 – 14:00



18 June 14:00 – 19:00



8 July 04:00 to 07:00

8 July 07:00 – 11:00

8 July 11:00 – 19:00

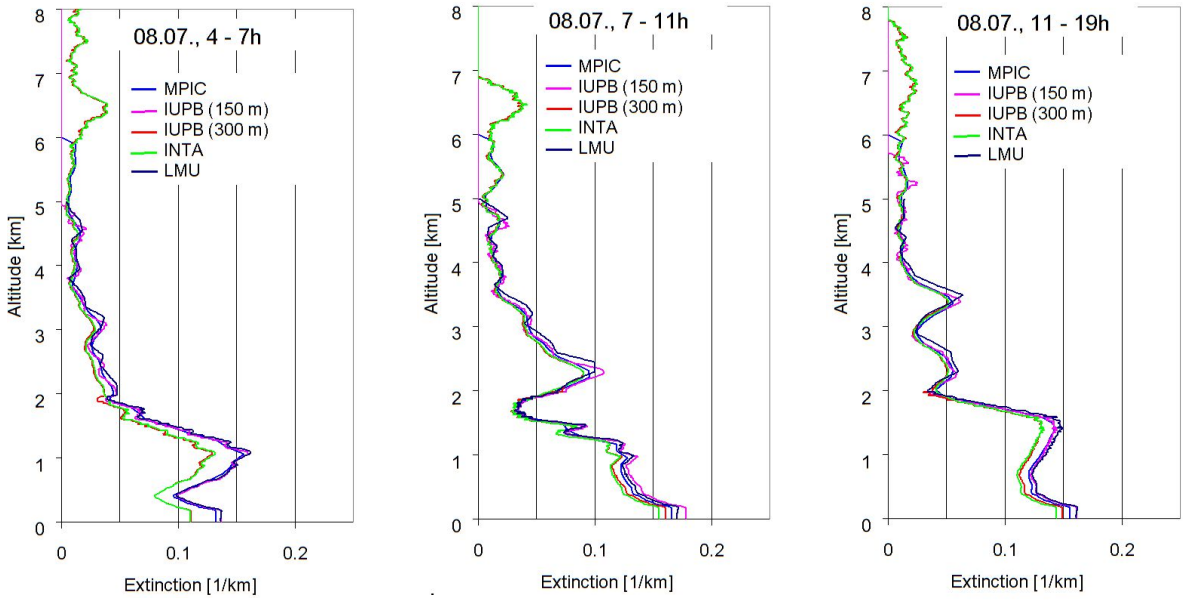


Fig. 9 Comparison of the aerosol extinction profiles extracted by the different groups for all three periods on both days.

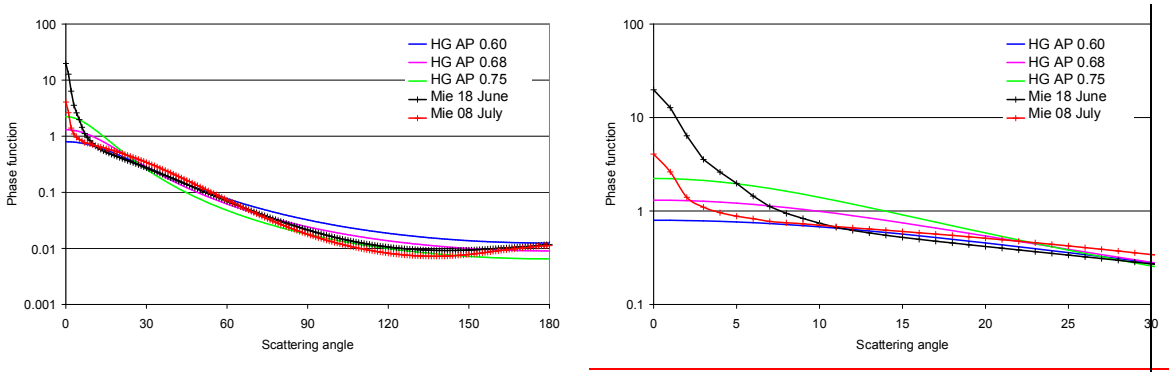


Fig. 10 Comparison of different aerosol phase functions used in the radiative transfer simulations. The right figure is a zoom of the left figure.

Real measurements	Synthetic spectra with noise	Synthetic spectra without noise
$2.71 \cdot 10^{43}$ molec²/cm^5	$2.00 \cdot 10^{43}$ molec²/cm^5	$1.84 \cdot 10^{43}$ molec²/cm^5

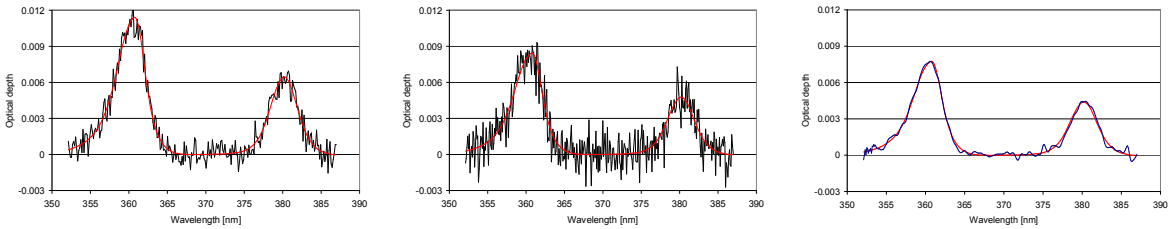


Fig. 11 Spectral analysis results for a real measurement from the MPIC instrument (left) and a synthetic spectrum with and without noise. Spectra are taken from 8 July 2013 at 11:26 (elevation angle = 1°). The derived O_4 dSCD is shown above the individual plots.

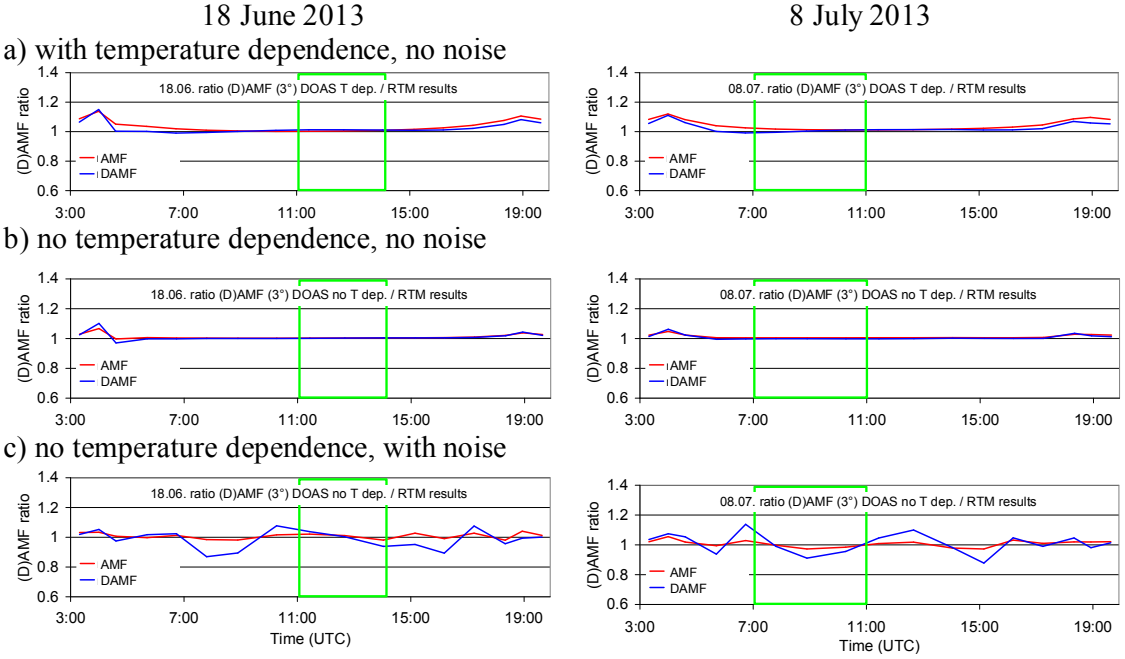


Fig. 12 Ratio of the O_4 (d)AMFs derived from synthetic spectra versus those obtained from radiative transfer simulations at 360 nm for both selected days.

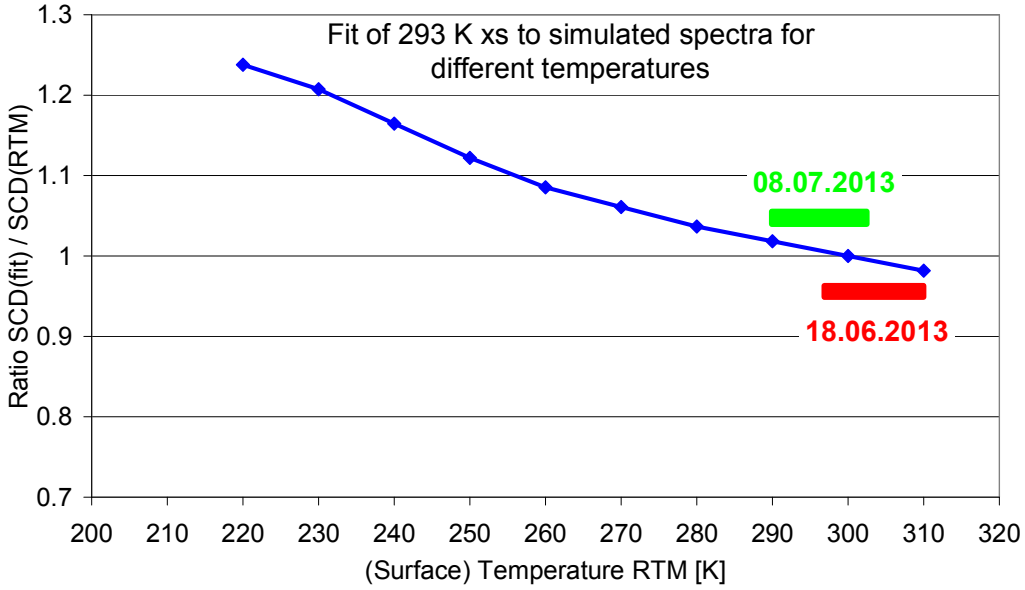
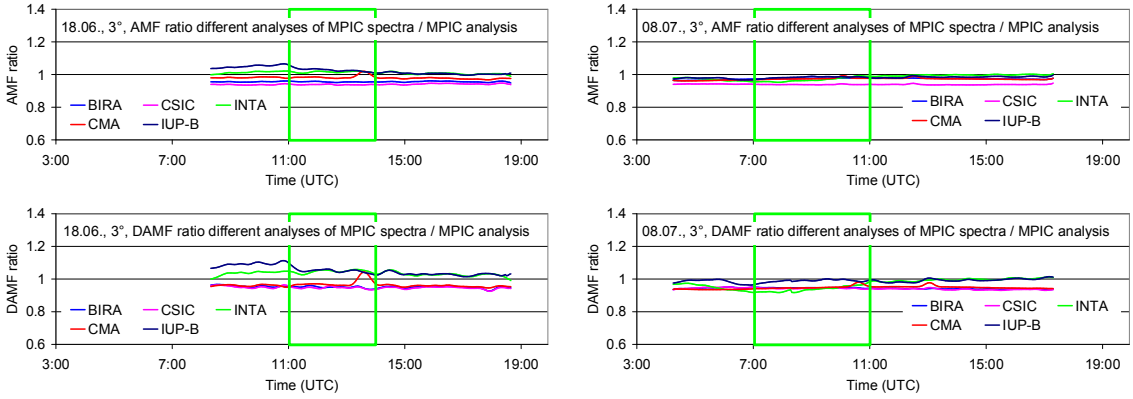


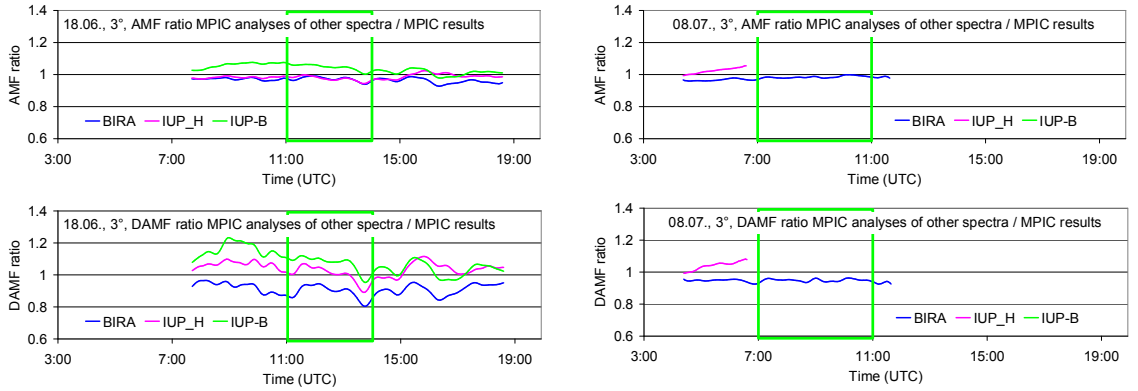
Fig. 13 Ratio of the O_4 dAMF obtained from simulated spectra for different surface temperatures by the corresponding O_4 dAMFs derived from radiative transfer simulations. The results represent MAX-DOAS observations at low elevation angles (2° to 3°).

18 June 2013 8 July 2013

a) Spectra from MPIC analysed by other groups



b) Spectra from other groups analysed by MPIC (all analyses for 335 – 374 nm)



c) Spectra from other groups analysed by the same groups

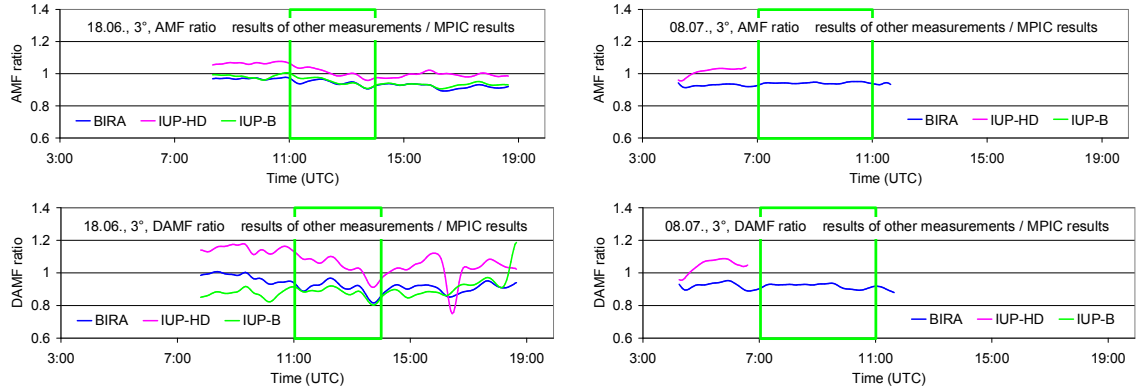


Fig. 14 a) Ratio of the O_4 (d)AMFs derived from MPIC spectra when analysed by other groups versus those analysed by MPIC for both selected days; b) Ratio of the O_4 (d)AMFs derived from spectra measured and analysed by other groups (using different wavelength ranges and settings) versus those for the MPIC instrument analysed by MPIC; c) Ratio of the O_4 (d)AMFs derived from spectra measured by other groups but analysed by MPIC versus those for the MPIC instrument analysed by MPIC (using the spectral range 335 – 374 nm for all instruments).

18 June 2013 8 July 2013

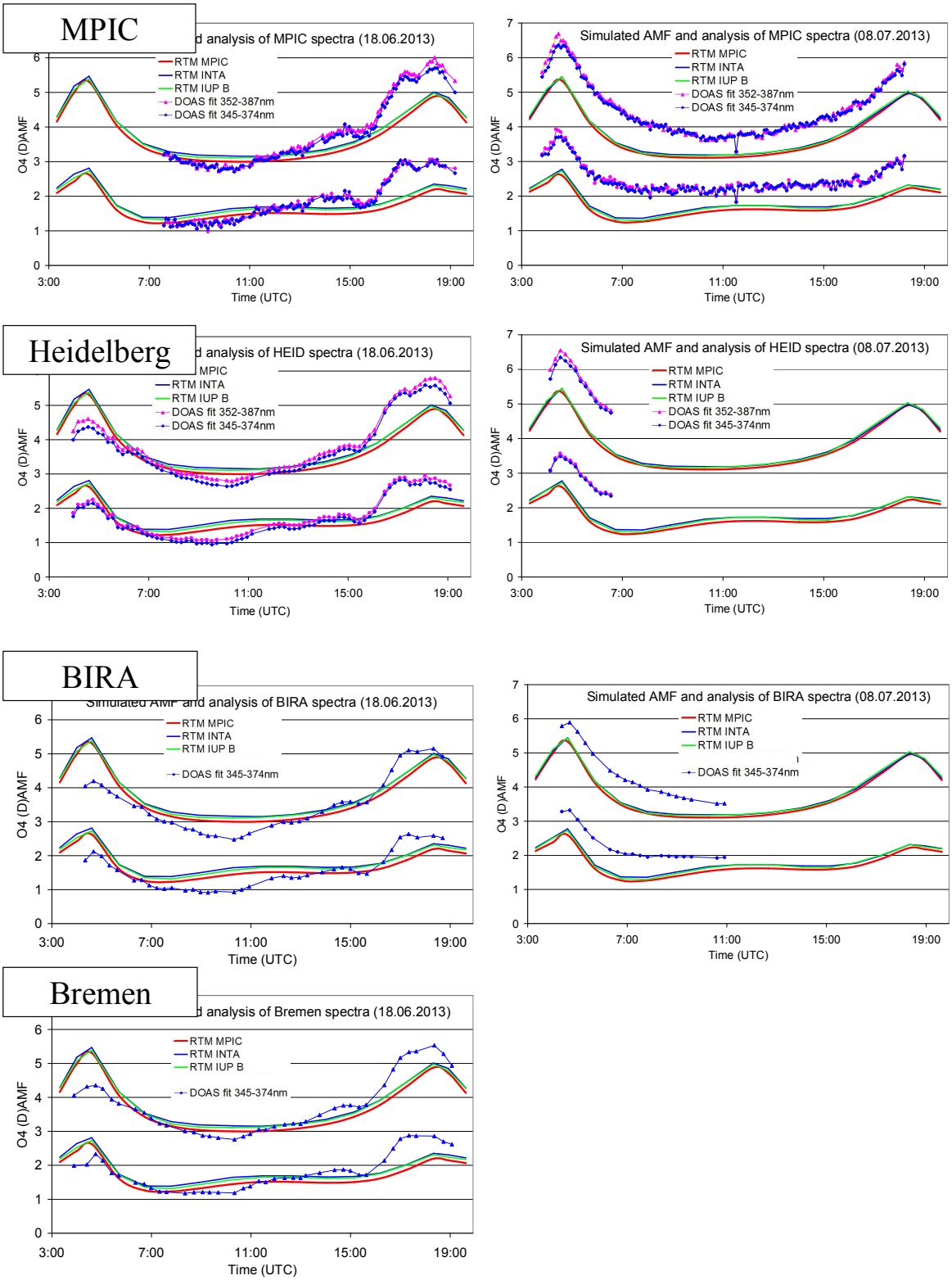


Fig. 15 Comparison of measured and simulated O_4 (d)AMFs for both selected days. Measurements are from 4 different instruments, but analysed by MPIC using the standard settings (see Table 7). Simulations are performed by three different groups using Mie phase functions and otherwise the standard settings (see Table 6).

Appendix A1 Settings used for the simulation of synthetic spectra

Table A1 Vertical resolution used in radiative transfer simulations for different altitude ranges.

Lower boundary [km]	Upper boundary [km]	Vertical resolution [km]
0	0.5	0.02
0.5	2	0.1
2	12	0.2
12	25	1
25	45	2
45	100	5
100	1000	900

Table A2 Dependence of SZA and relative azimuth angle on time (UTC) for the standard viewing direction (51° with respect to North).

Time (UTC)	SZA	RAZI
03:19	90	-0.1
04:00	85	7.7
04:36	80	14.2
05:42	70	26
06:44	60	37.5
07:48	50	50.1
08:54	40	66.2
10:16	30	94.6
11:26	26	129
12:40	30	163.3
14:02	40	191.8
15:09	50	207.9
16:11	60	220.5
17:14	70	232
18:20	80	243.8
18:56	85	250.3
19:38	90	258

Table A3 Trace gas profiles and cross sections used for the simulation of the synthetic spectra.

Trace gas	Vertical profile	Cross section (reference and T)
O_4	Derived from temperature and pressure profiles during. 18.06.: average profiles 11:00 – 14:00 08.07.: average profiles 7:00 – 11:00	Thalman and Volkamer (2013) (203, 232~~2~~3, 253, 273, 293 K)*
HCHO	18.06.: 0-1000m, constant concentration of $2 \cdot 10^{11}$ molec/cm³ (about 8 ppb) 08.07.: 0-1000m, constant concentration of $1 \cdot 10^{11}$ molec/cm³ (about 4 ppb)	Meller and Moortgat (2000) (298 K)
NO_2	Troposphere 18.06.: 0-500m, constant concentration of $4 \cdot 10^{11}$ molec/cm³ (about 16 ppb) 08.07.: 0-500m, constant concentration of $2 \cdot 10^{11}$ molec/cm³ (about 8 ppb) Stratosphere: Gaussian profile with maximum at 25 km, and FWHM of 16 km, VCD = $5 \cdot 10^{15}$ molec/cm²	Vandaele et al. (1997) (220, 294 K)
O_3	Troposphere (0-8km): constant concentration $6 \cdot 10^{11}$ molec/cm³ (about 24 ppb) Stratosphere: Gaussian profile with maximum at 22 km, and FWHM of 15 km, VCD = 314 DU	Serdyuchenko et al. (2014) (193 – 293 K in steps of 10 K)**

*The temperature dependence is either considered or a constant temperature of 293 K is assumed (see text for details).

**The temperature dependence was parameterised according to Paur and Bass (1984).

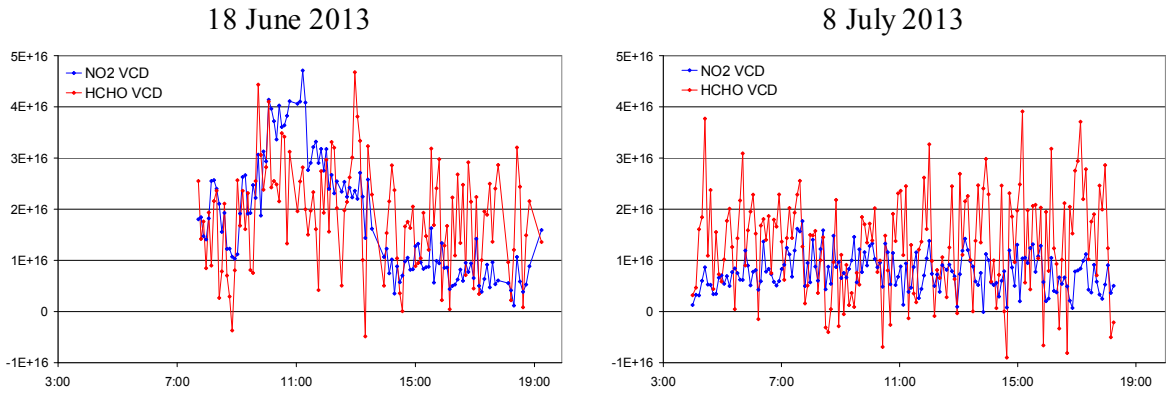


Fig. A1 Tropospheric VCDs of NO_2 (blue) and HCHO (red) derived from measurements at 30° elevation using the geometric approximation.

Appendix A2 Comparison of measured and simulated O_4 (d)AMFs for all azimuth and elevation angles of the MPIC MAX-DOAS measurements.

The settings for the simulation of the synthetic spectra are given in Table 6 and Tables A1, A2, and A3 in appendix 1. Measurements are analysed using the standard settings (see Table 7).

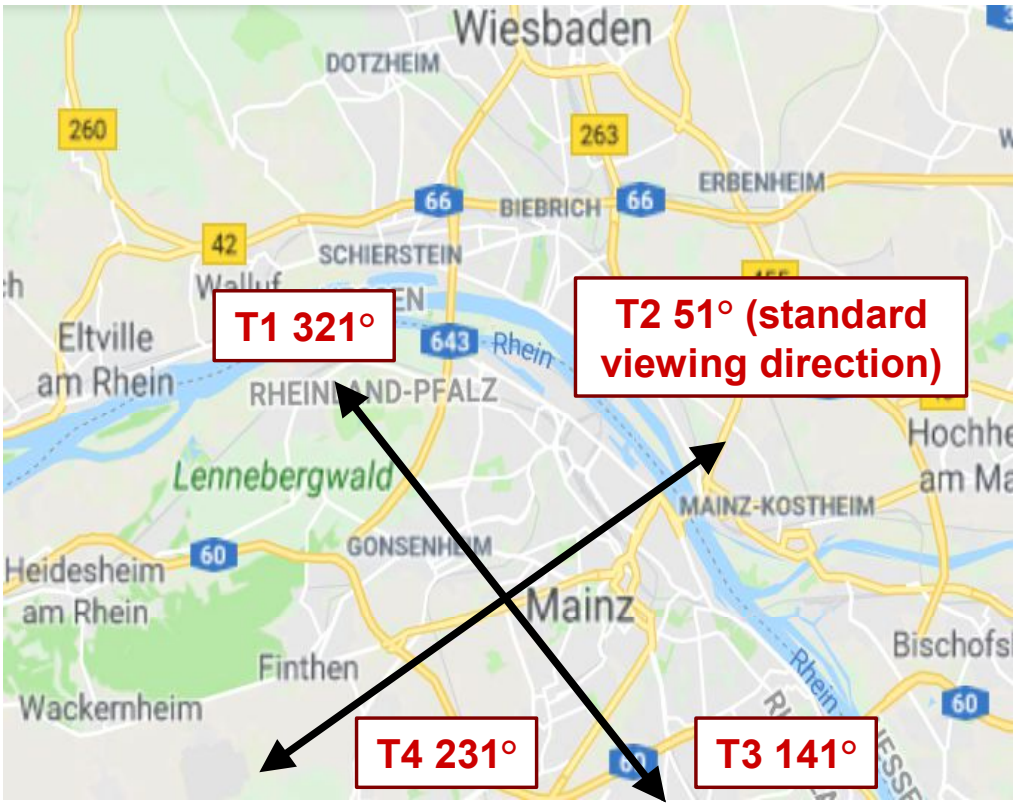
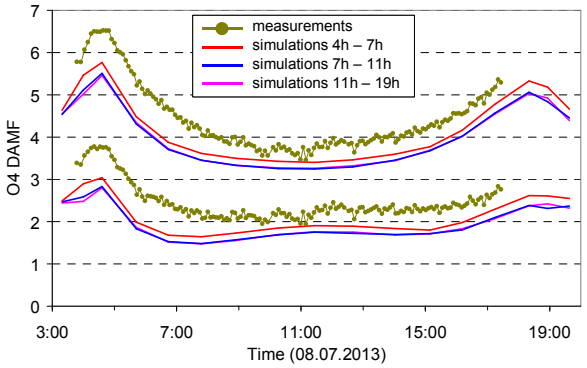


Fig. A2 Azimuth viewing directions of the 4 telescopes (T1 to T4) of the MPIC MAX-DOAS instrument. The azimuth angles are defined with respect to North (map: © google maps).

T1 North-West

For T1 and T4 azimuth direction, no measurements at 1° elevation were possible due to obstacles.



T4 South-West

For T1 and T4 azimuth direction, no measurements at 1° elevation were possible due to obstacles.

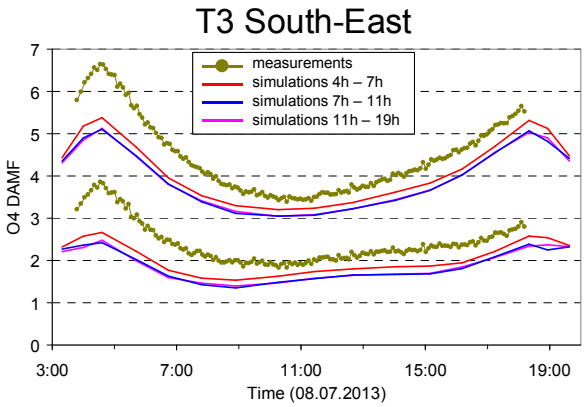


Fig. A3a Comparison results for 1° elevation angles on 8 July 2013. The upper lines indicate the O_4 AMFs, the lower lines the O_4 dAMFs (see also Fig. 2 and 3).

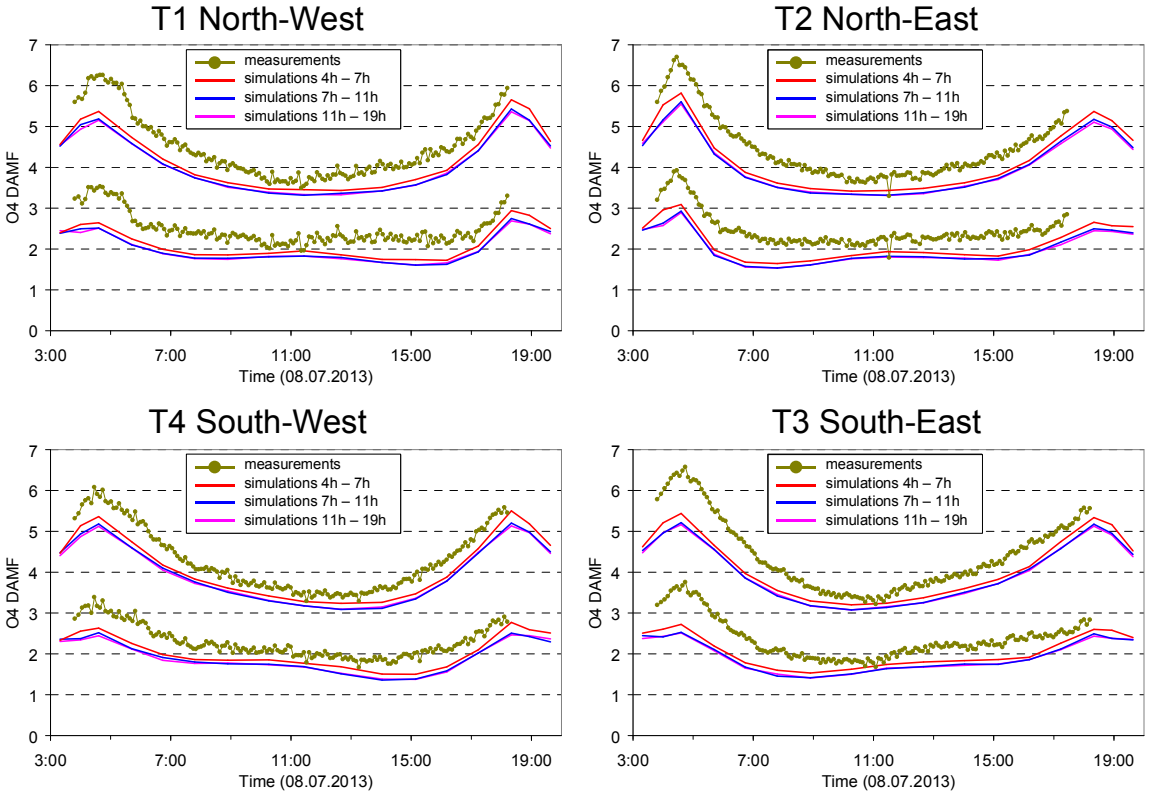


Fig. A3b Comparison results for 3° elevation angles on 8 July 2013.

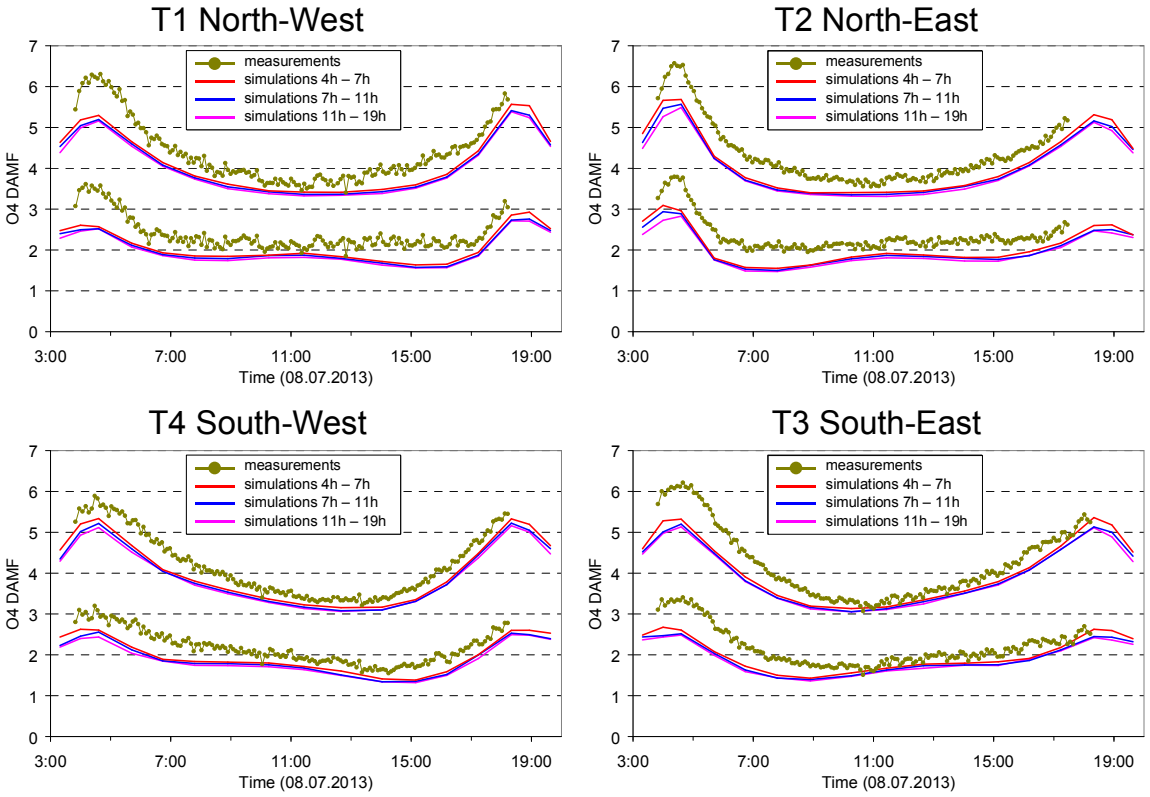


Fig. A3c Comparison results for 6° elevation angles on 8 July 2013.

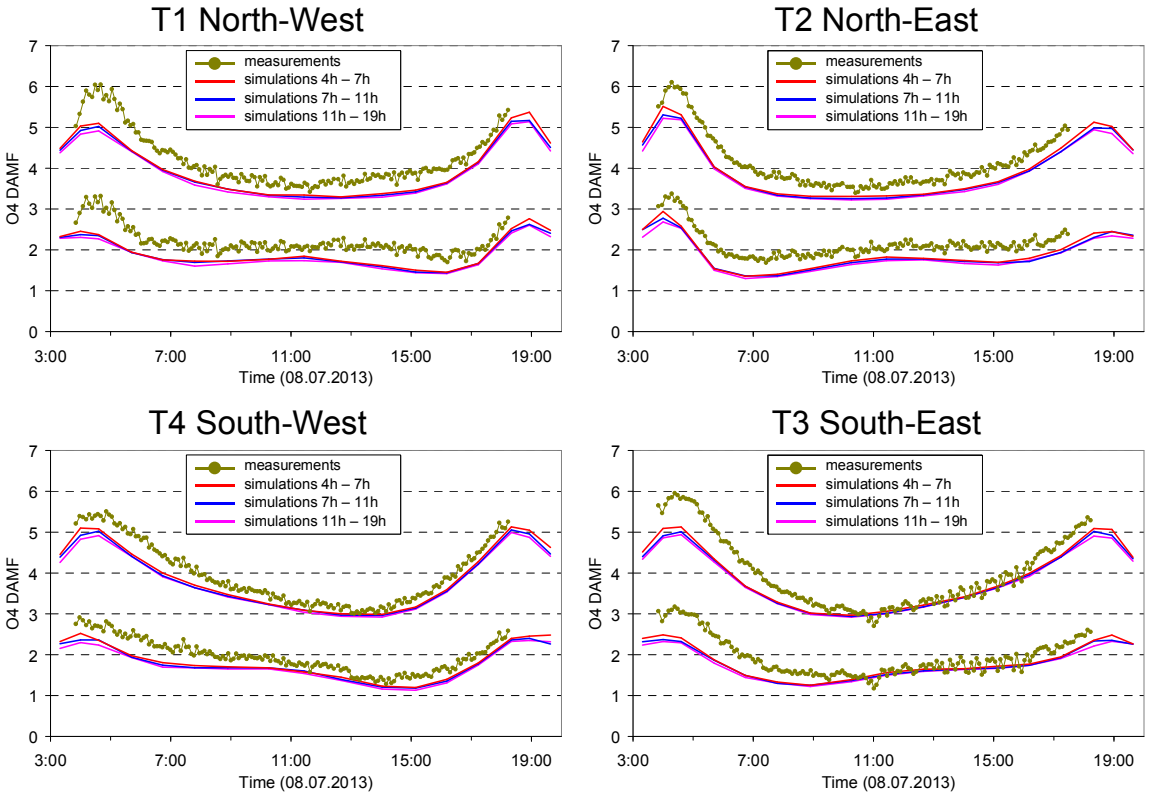


Fig. A3d Comparison results for 10° elevation angles on 8 July 2013.

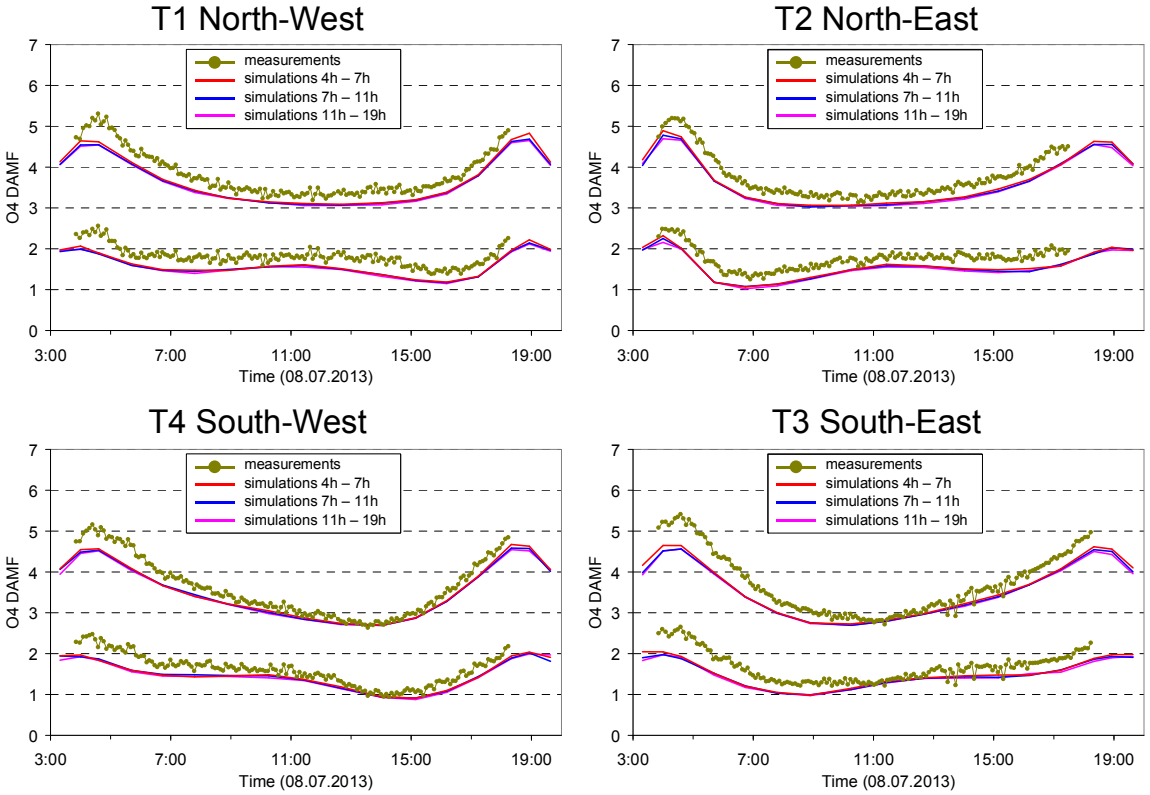


Fig. A3e Comparison results for 15° elevation angles on 8 July 2013.

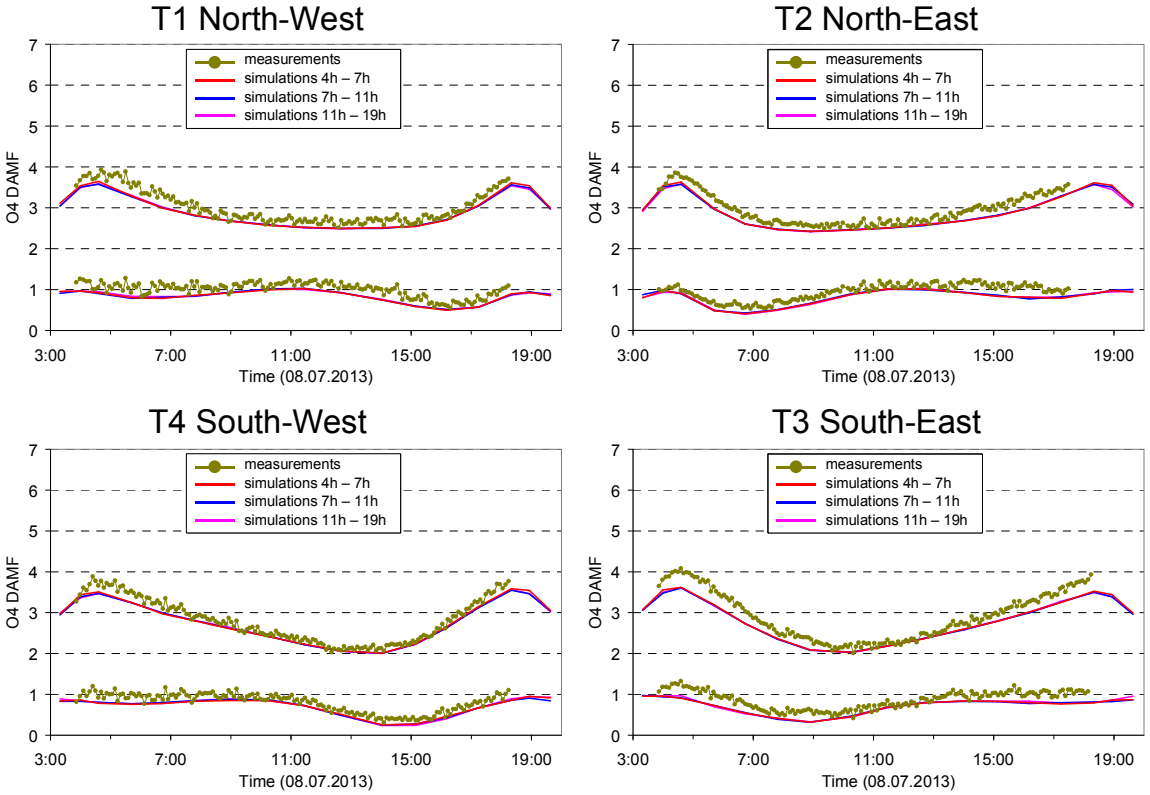


Fig. A3f Comparison results for 30° elevation angles on 8 July 2013.

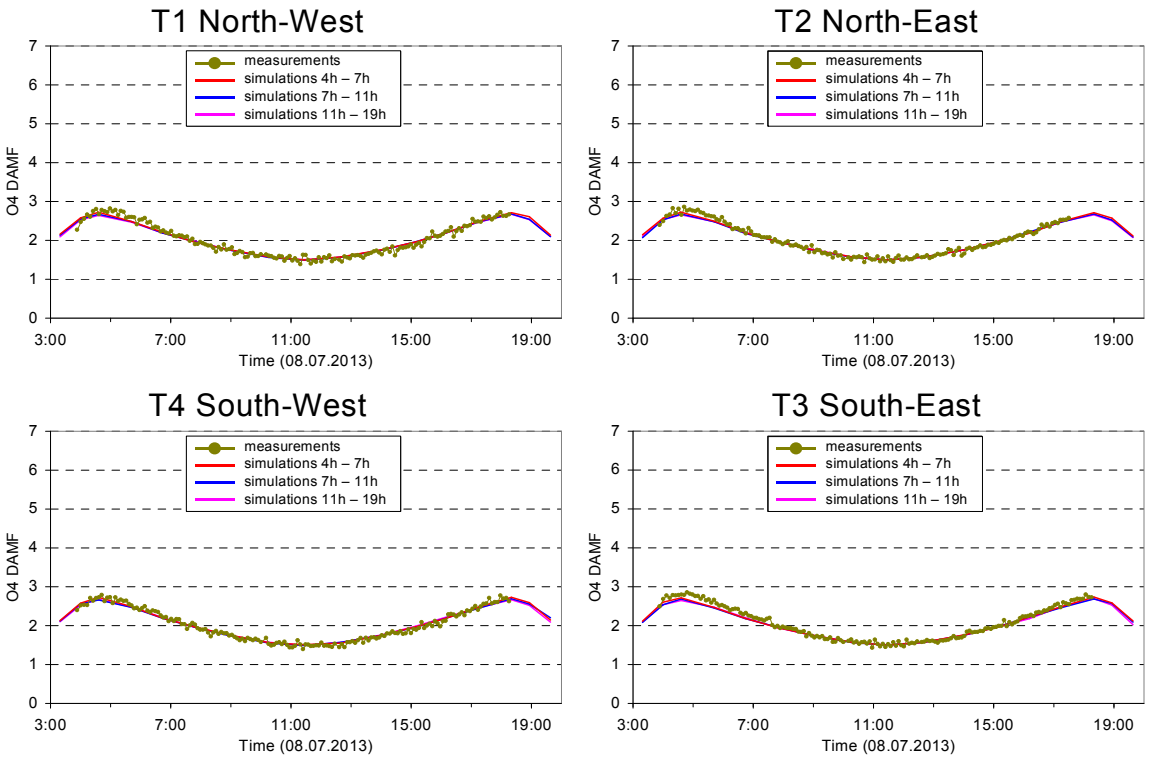


Fig. A3g Comparison results (only O_4 AMFs) for 90° elevation angles on 8 July 2013.

T1 North-West

For T1 and T4 azimuth direction, no measurements at 1° elevation were possible due to obstacles.

In the following figures, the green lines indicate the results for the modified aerosol profile

T4 South-West

For T1 and T4 azimuth direction, no measurements at 1° elevation were possible due to obstacles.

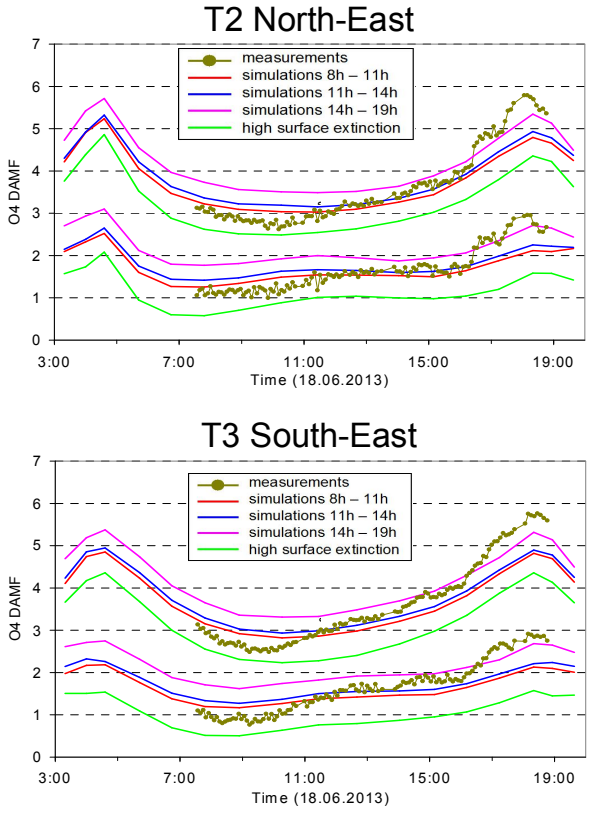


Fig. A4a Comparison results for 1° elevation angles on 18 June 2013 including the RTM results for the modified aerosol extinction profile (green line).

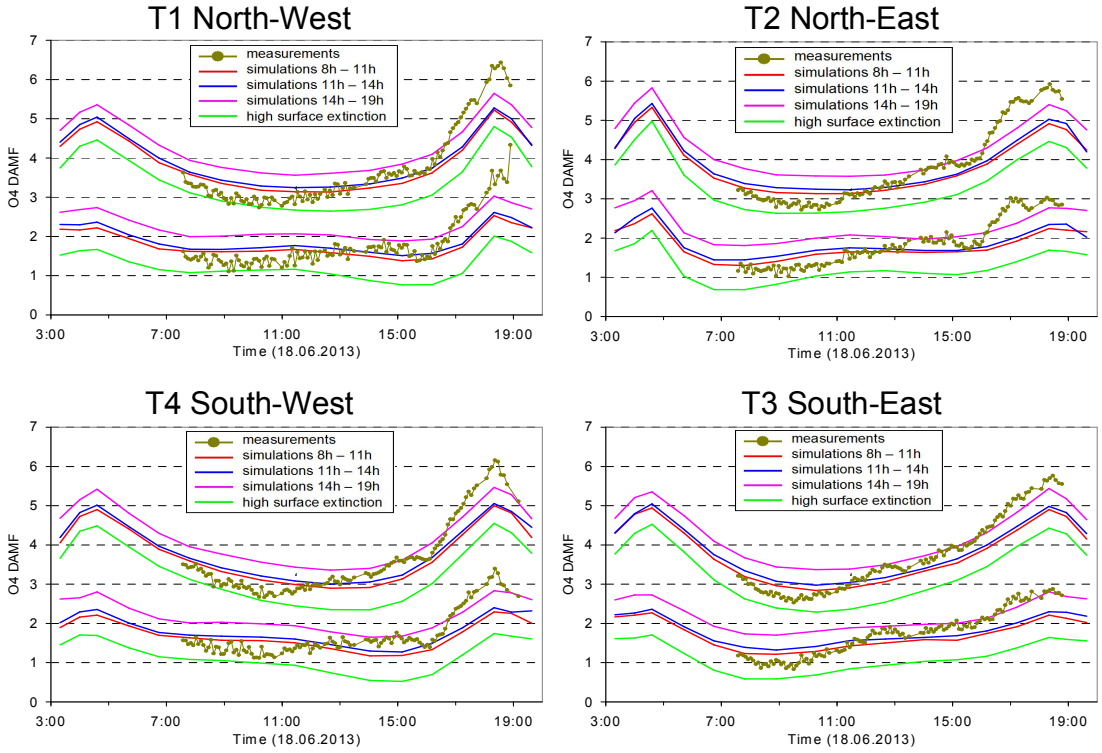


Fig. A4b Comparison results for 3° elevation angles on 18 June 2013 including the RTM results for the modified aerosol extinction profile (green line)..

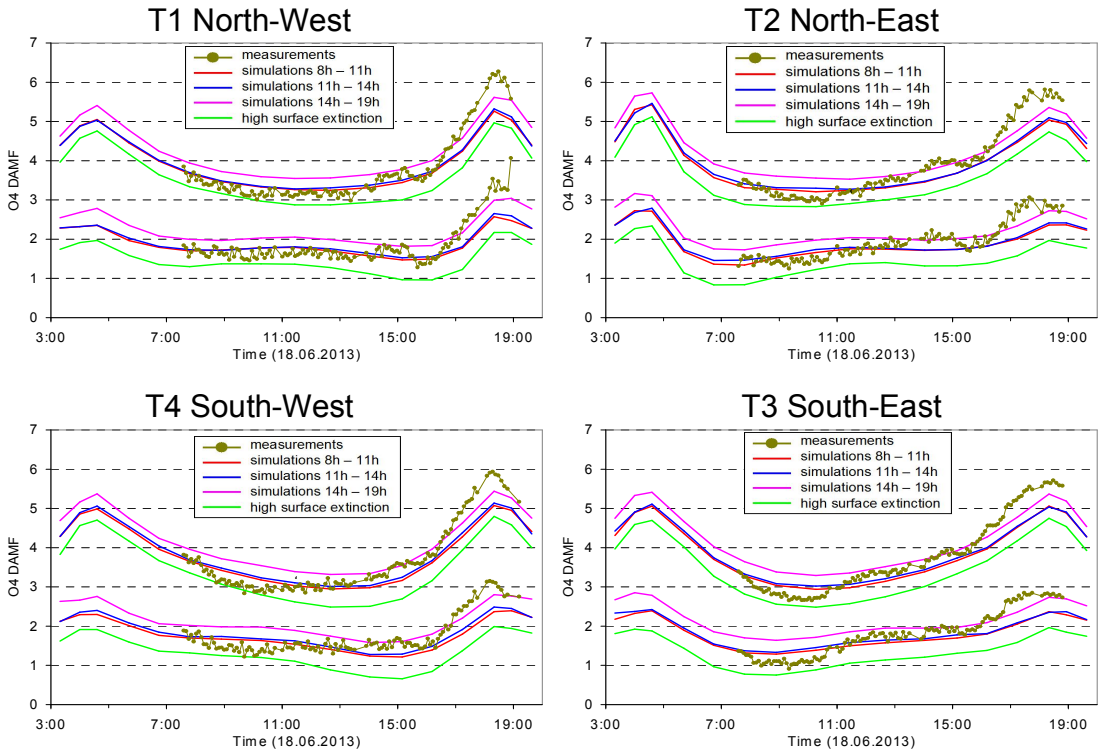


Fig. A4c Comparison results for 6° elevation angles on 18 June 2013 including the RTM results for the modified aerosol extinction profile (green line).

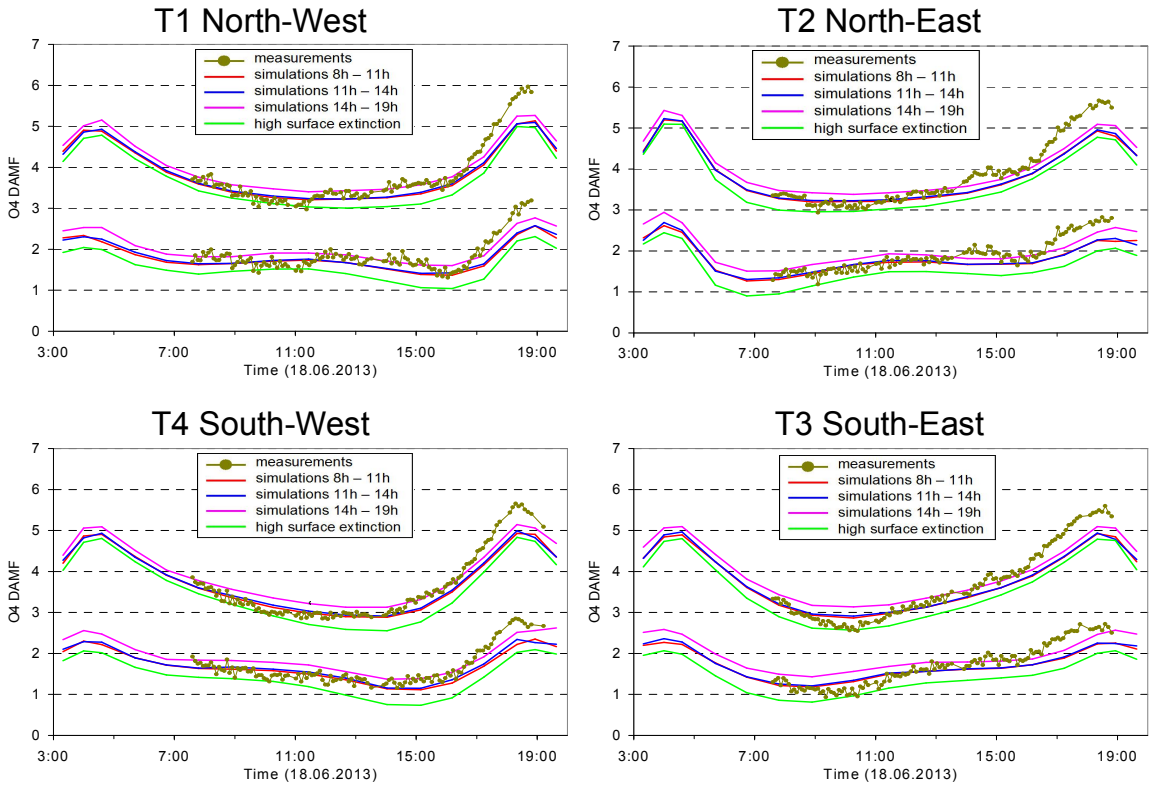


Fig. A4d Comparison results for 10° elevation angles on 18 June 2013 including the RTM results for the modified aerosol extinction profile (green line).

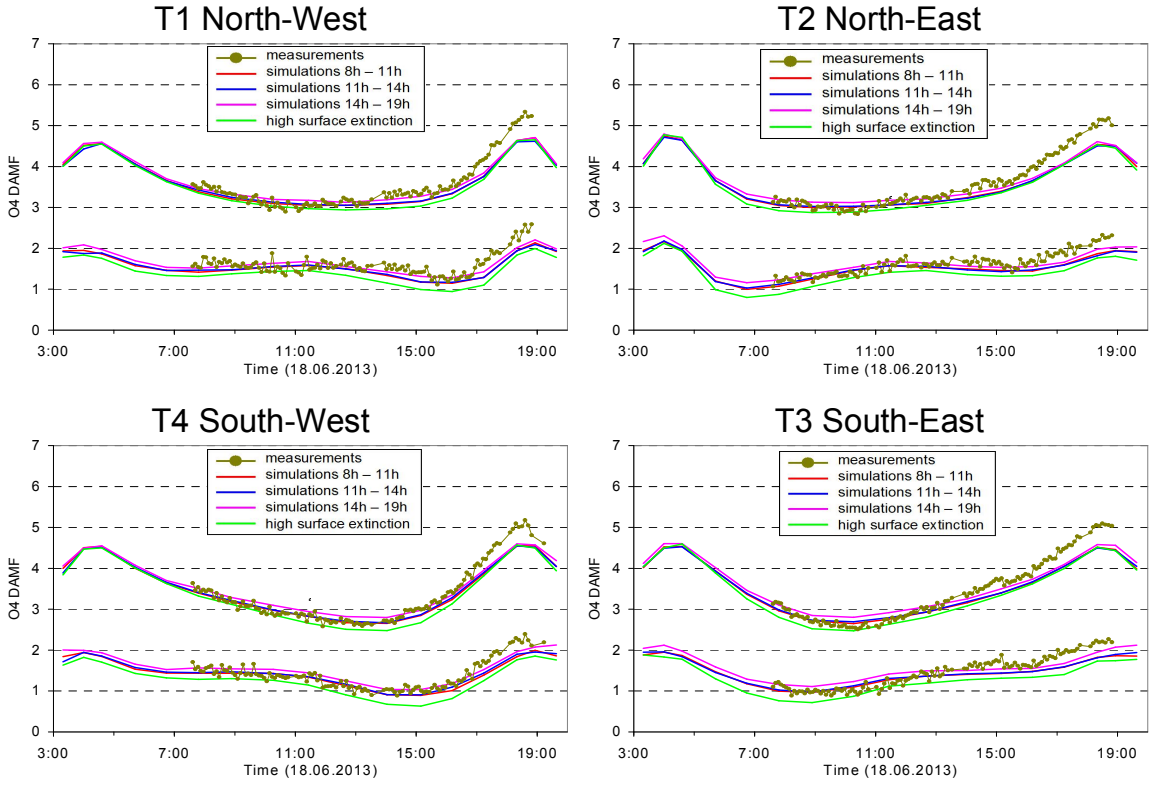


Fig. A4e Comparison results for 15° elevation angles on 18 June 2013 including the RTM results for the modified aerosol extinction profile (green line)..

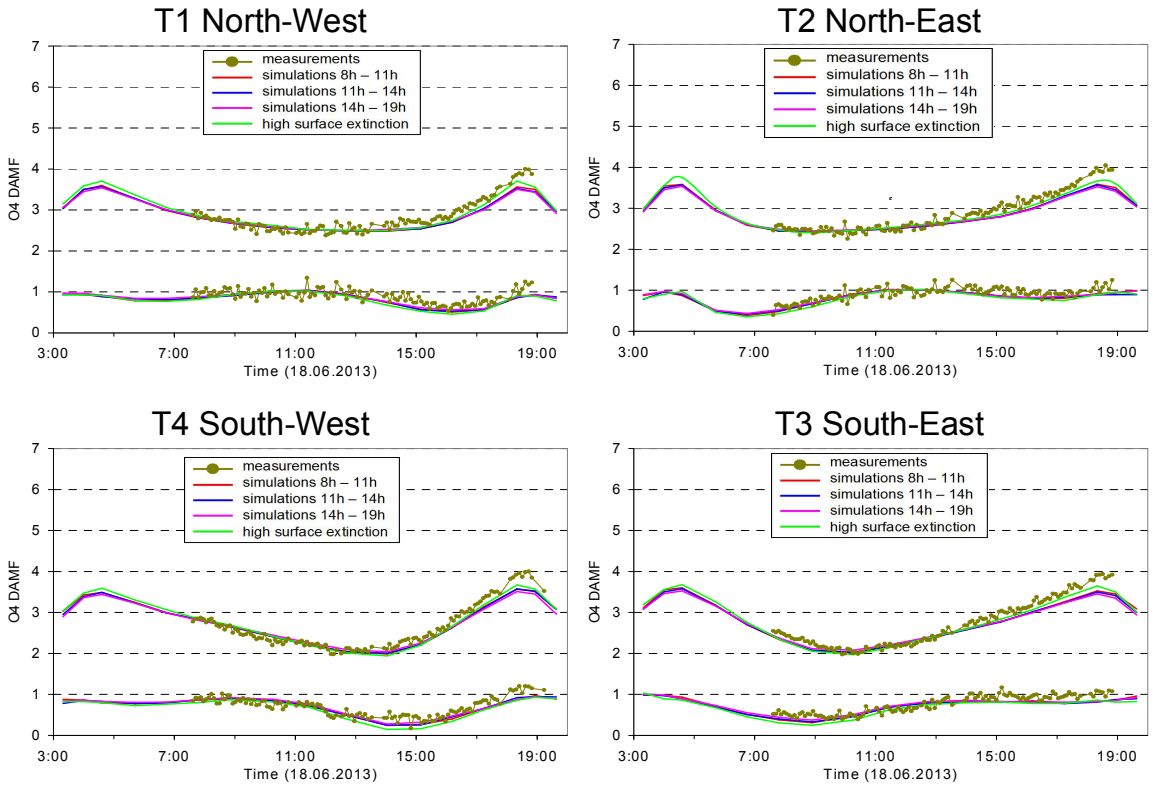


Fig. A4f Comparison results for 30° elevation angles on 18 June 2013 including the RTM results for the modified aerosol extinction profile (green line)..

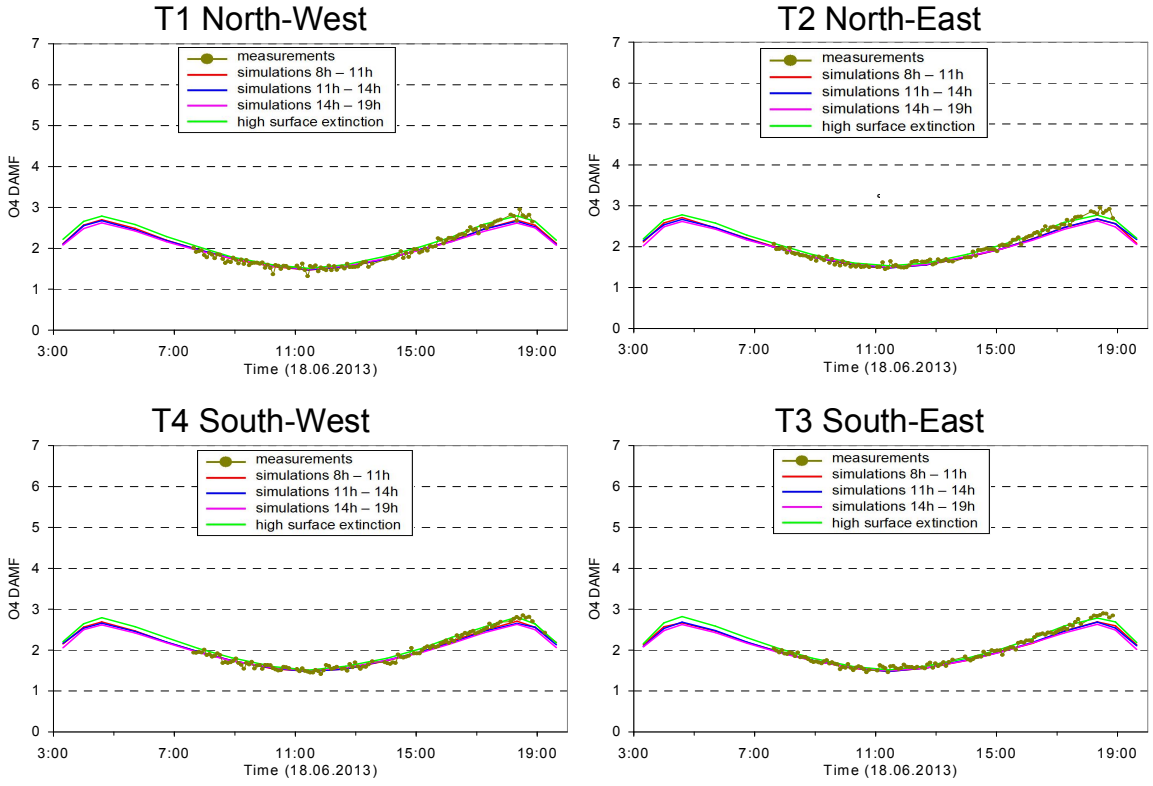


Fig. A4g Comparison results (only O_4 AMFs) for 90° elevation angles on 18 June 2013 including the RTM results for the modified aerosol extinction profile (green line).

Appendix A3 Comparison of the different procedures to extracted height profiles of temperature, pressure and O_4 concentration

Extraction of temperature and pressure profiles

For the two selected days during the MAD-CAT campaign two data sets of temperature and pressure are available: surface measurements close to the measurement site and vertical profiles from ECMWF ERA-Interim re-analysis data (see Table 5). Both data sets are used to derive the O_4 concentration profiles for the three selected periods on both days. The general procedure is that first the temperature profiles are determined. In a second step, the pressure profiles are derived from the temperature profiles and the measured surface pressure. For the temperature profile extraction, three height layers are treated differently:

-below 1 km

Between the surface (~150 m above sea level) and 1 km, the temperature is linearly interpolated between the average of the in situ measurements of the respective period and the ECMWF data at 1 km (see next paragraph). This procedure is used to account for the diurnal variation of the temperature close to the surface. Here it is important to note that for this surface-near layer the highest accuracy is required, because a) the maximum O_4 concentration is located near the surface, and b) the MAX-DOAS measurements are most sensitive close to the surface.

-1 km to 20 km

In this altitude range, the diurnal variation of the temperature becomes very small. Thus the average of the four ECMWF profiles of each day is used (for simplicity, a 6^{th} order polynomial is fitted to the ECMWF data).

-Above 20 km

In this altitude range the accuracy of the temperature profile is not critical and thus the ECMWF temperature profile for 00:00 UTC of the respective day is used for simplicity.

The temperature profiles for 8 July 2013 extracted in this way are shown in Fig. 4 (left). Close to the surface the temperature variation during the day is about 10 K.

In the next step, the pressure profiles are determined from the surface pressure (obtained from the in situ measurements) and the extracted temperature profiles according to the ideal gas law. In principle the effect of atmospheric humidity could also be taken into account, but the effect is very small for surface-near layers and is thus ignored here. The derived pressure profiles for 8 July 2013 are shown in Fig. 4 (right). Excellent agreement with the corresponding ECMWF pressure profiles is found.

Here it should be noted that in principle also the ECMWF pressure profiles could be used. However, we chose to determine the pressure profiles from the surface pressure and the extracted temperature profiles, because this procedure can also be applied if no ECMWF data (or other information on temperature and pressure profiles) is available.

If no profile data (e.g. from ECMWF) are available, temperature and pressure profiles can also be extrapolated from surface measurements e.g. by assuming a constant lapse rate of -0.65 K / 100 m for the altitude range between the surface and 12 km, and a constant temperature above 12 km (as stated above, uncertainties at this altitude range have only a negligible effect on the O_4 VCD). If no measurements or model data are available at all, a fixed temperature and pressure profile can be used, e.g. the US standard atmosphere (United States Committee on Extension to the Standard Atmosphere, 1976).

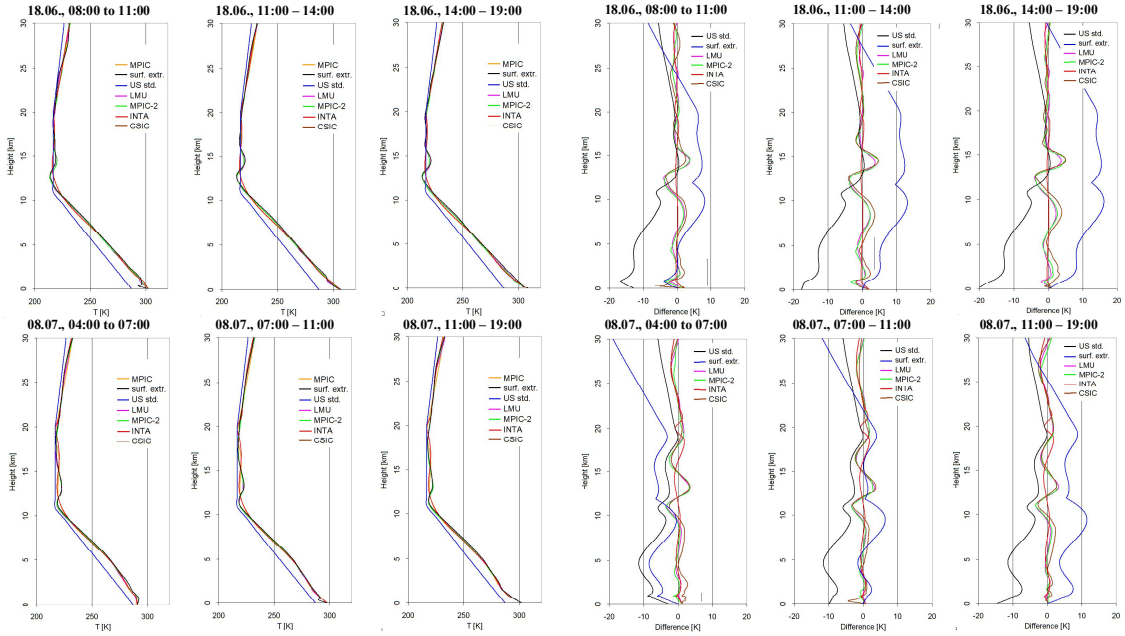


Fig. A5a Left: Comparison of temperature profiles extracted by the different groups (also shown are the profiles from the US standard atmosphere and the profiles extrapolated from the surface measurements). Right: Differences of these profiles compared to the MPIC standard extraction.

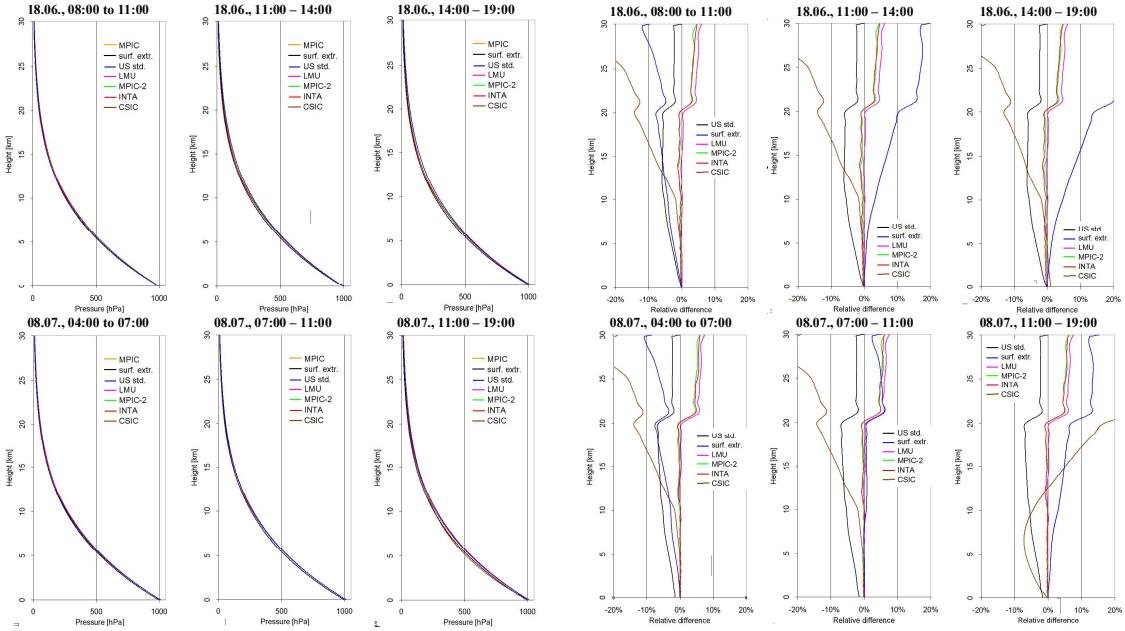


Fig. A5b Left: Comparison of pressure profiles extracted by the different groups (also shown are the profiles from the US standard atmosphere and the profiles extrapolated from the surface measurements). Right: Differences of these profiles compared to the MPIC standard extraction.

Determination of the uncertainties of the O_4 profiles and O_4 VCDs caused by uncertainties of the input parameters

The uncertainties of the O_4 profiles and O_4 VCDs are derived by varying the input parameters according to their uncertainties. The following results are obtained:

-The variation of the temperature (whole profile) by about 2K leads to variations of the O_4 concentration (or O_4 VCD) by about 0.8%.

-The variation of the surface pressure by about 3 hPa leads to variations of the O_4 concentration (or O_4 VCD) by about 0.7%.

-The effect of uncertainties of the relative humidity depends strongly on temperature: For surface temperatures of 0°C, 10°C, 20°C, 30°C, and 35°C a variation of the relative humidity of 30% leads to variations of the O_4 concentration (or O_4 VCDs) of about 0.15%, 0.3%, 0.6%, 1.2%, and 1.6%, respectively. If the effect of atmospheric humidity is completely ignored (dry air is assumed), the resulting O_4 concentrations (or O_4 VCDs) are systematically overestimated by about 0.3%, 0.7%, 1.3%, 2.5%, and 4% for surface temperatures of 0°C, 10°C, 20°C, 30°C, and 35°C, respectively (assuming a relative humidity of 70%). In this study we used the relative humidity measured by the in situ sensors. We took these values not only for the surface layers, but also for the whole troposphere. Here it should be noted that the related uncertainties of the absolute humidity decrease quickly with altitude because the absolute humidity itself decreases quickly with altitude. Since both selected days were warm or even hot summer days, we estimate the uncertainty of the O_4 concentration and O_4 VCDs due to uncertainties of the relative humidity to 1% and 0.4% on 18 June and 8 July, respectively.

Assuming that the uncertainties of the three input parameters are independent, the total uncertainty related to these parameters is estimated to be about 1.5%.

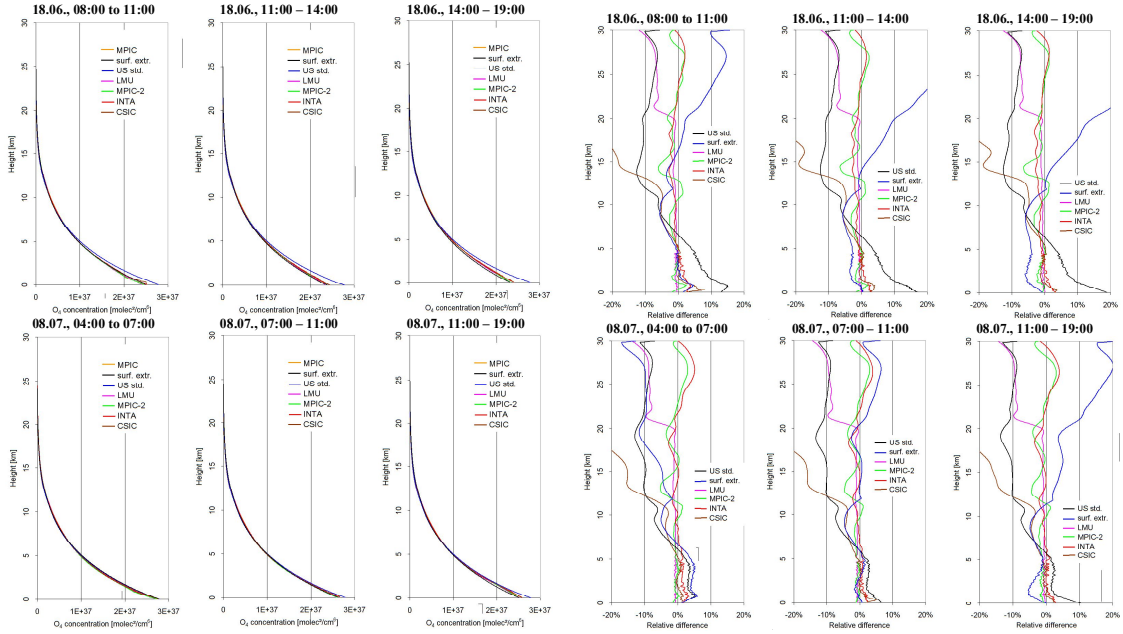


Fig. A5c Left: Comparison of O_4 concentration profiles extracted by the different groups (also shown are the profiles from the US standard atmosphere and the profiles extrapolated from the surface measurements). Right: Differences of these profiles compared to the MPIC standard extraction.

Appendix A4 Results of the sensitivity studies of simulated and measured O_4 (d)MFs

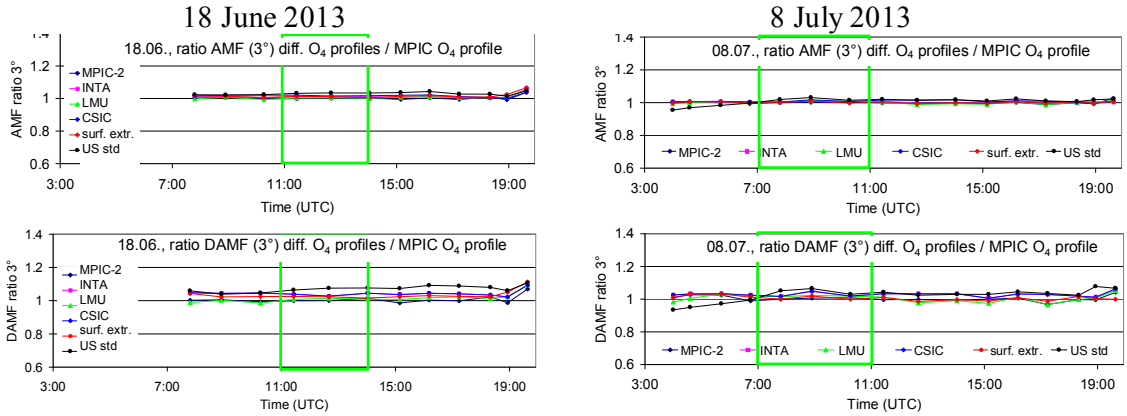


Fig. A6 Ratio of the O_4 AMFs (top) and O_4 dAMFs (bottom) derived for different O_4 profiles versus the standard O_4 profile (MPIC) for both selected days. Besides the O_4 profiles extracted by the different groups, also the O_4 profiles derived from the US standard atmosphere and for the extrapolation of the surface values are included.

Table A4 Average ratios of O_4 (d)AMFs simulated for different O_4 profiles versus the results for the standard settings (using the MPIC O_4 profiles) for the two middle periods on both selected days.

	AMF ratios			dAMF ratios	
O_4 profile extraction	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
MPIC-2	1.00	1.00		1.00	1.00
INTA	1.01	1.01		1.02	1.01
LMU	1.00	1.00		1.01	1.02
CSIC	1.02	1.01		1.04	1.02
Lapse rate	1.01	1.00		1.02	1.01
US std. atm.	1.03	1.02		1.07	1.04

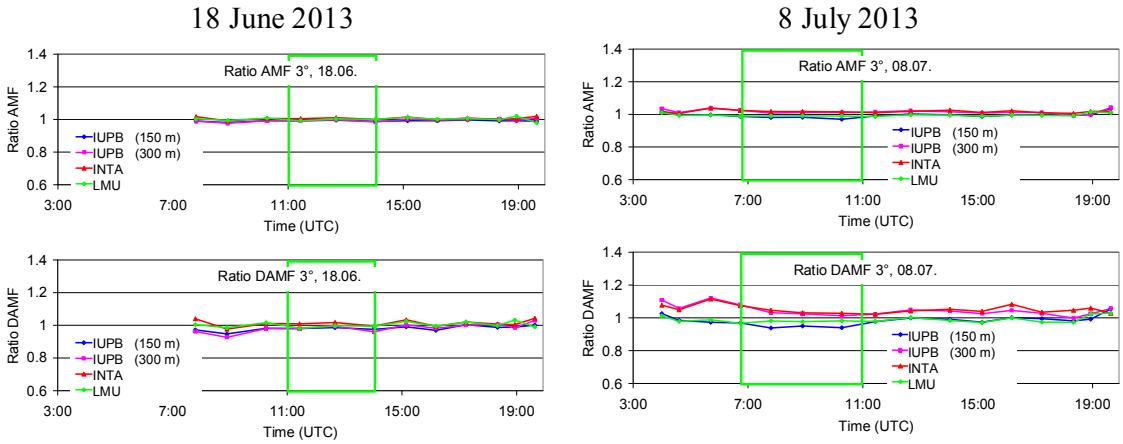


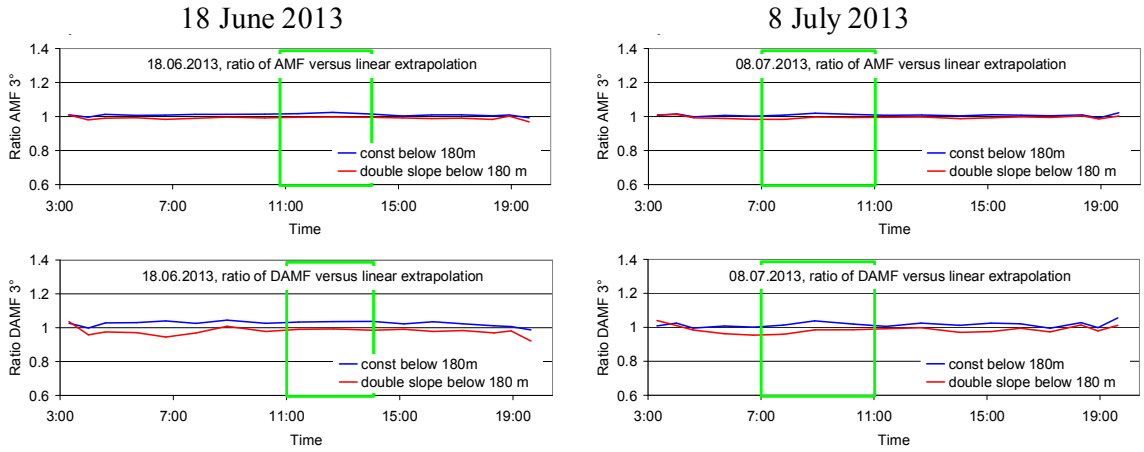
Fig. A7 Ratio of the O_4 AMFs (top) and O_4 dAMFs (bottom) derived for aerosol extinction profiles extracted by different groups versus the standard aerosol extinction profiles (MPIC) for both selected days.

Table A5 Average ratios of O_4 (d)AMFs simulated for different aerosol extinction profiles versus the results for the standard settings (using the MPIC aerosol extinction profiles) for the two middle periods on both selected days.

	AMF ratios			dAMF ratios	
Aerosol profile extraction	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
INTA	1.01	1.02		1.01	1.04

IUP-B 150 m	0.99	0.98		0.98	0.96
IUP-B 300 m	0.99	1.01		0.98	1.03
LMU	1.00	0.99		0.99	0.98

4620
4621



4622
4623 Fig. A8 Ratio of the O_4 AMFs (top) and O_4 dAMFs (bottom) derived for different
4624 extrapolations of the aerosol extinction profiles below 180 m versus those for the standard
4625 settings (linearly extrapolated profiles) for both selected days.
4626
4627
4628
4629 Table A6 Average ratios of O_4 (d)AMFs simulated for aerosol extinction profiles with
4630 different extrapolations below 180 m versus the results for the standard settings (linear
4631 extrapolation) for the two middle periods on both selected days.

	AMF ratios			**dAMF ratios**	
Extrapolation below 180 m	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Constant extinction	1.02	1.01		1.04	1.02
Double slope	1.00	0.99		0.99	0.98

4632
4633
4634

18 June 2013 8 July 2013

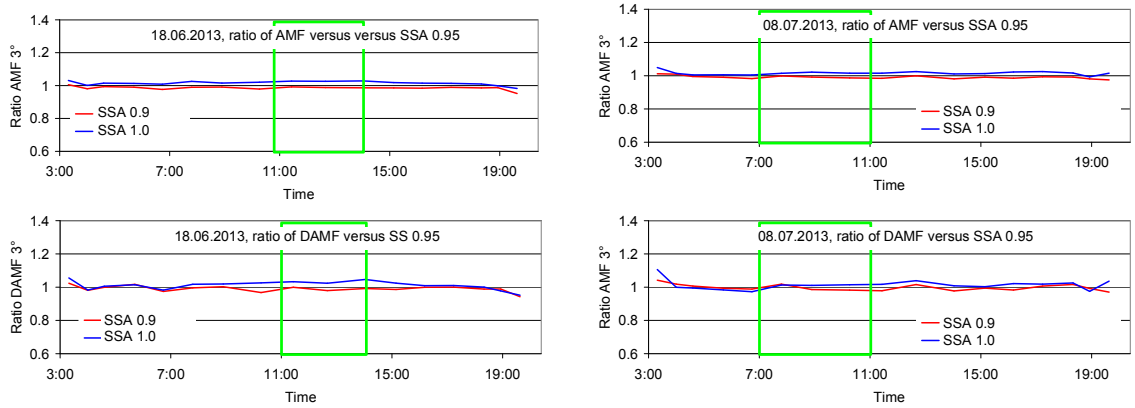


Fig. A9 Ratio of the O_4 AMFs (top) and O_4 dAMFs (bottom) derived for different aerosol single scattering albedos versus those for the standard settings (single scattering albedo of 0.95) for both selected days.

Table A7 Average ratios of O_4 (d)AMFs simulated for different aerosol single scattering albedos (SSA) versus the results for the standard settings (single scattering albedo of 0.95) for the two middle periods on both selected days.

	AMF ratios			**dAMF ratios**	
Single scattering albedo	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
0.9	0.99	0.99		0.99	0.99
1.0	1.03	1.01		1.03	1.01

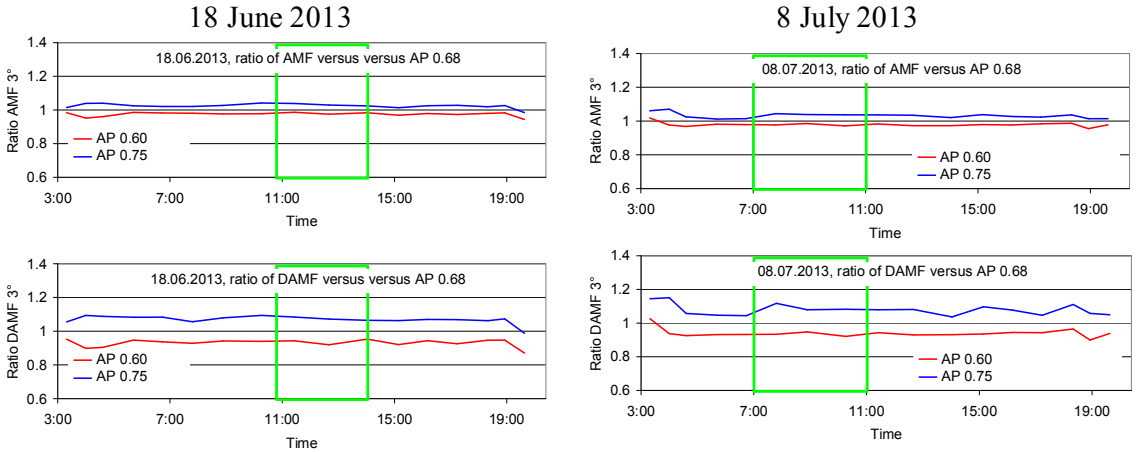


Fig. A10 Ratio of the O_4 AMFs (top) and O_4 dAMFs (bottom) derived for different aerosol phase functions (HG-parameterisation with different asymmetry parameters) versus those for the standard settings (asymmetry parameter of 0.68) for both selected days.

Table A8 Average ratios of O_4 (d)AMFs simulated for different aerosol phase functions (HG-parameterisation with different asymmetry parameters (AP) versus the results for the standard settings (asymmetry parameter of 0.68) for the two middle periods on both selected days.

	AMF ratios			**dAMF ratios**	
Asymmetry parameter	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
0.6	0.98	0.98		0.94	0.94
0.75	1.03	1.03		1.08	1.07

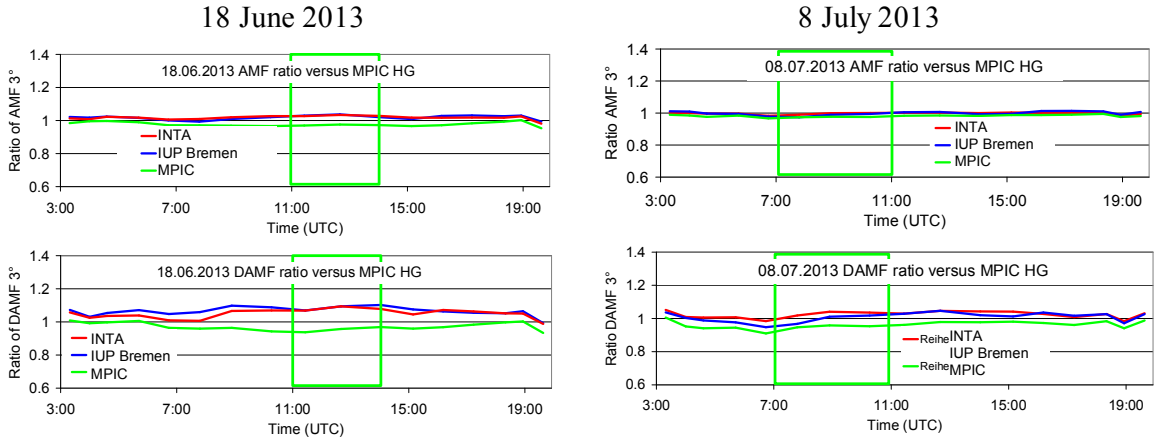


Fig. A11 Ratio of the O_4 AMFs (top) and O_4 dAMFs (bottom) simulated by INTA and IUP-Bremen and MPIC (SCIATRAN) for phase functions derived from the sun photometer measurements versus those simulated by MPIC using the Henyey Greenstein phase function for asymmetry parameter of 0.68 for both selected days.

Table A9 Average ratios of O_4 (d)AMFs simulated by INTA and IUP-Bremen and MPIC (SCIATRAN) for phase functions derived from the sun photometer measurements versus those simulated by MPIC using the Henyey Greenstein phase function for asymmetry parameter of 0.68 for the two middle periods on both selected days.

	AMF ratios			**dAMF ratios**	
Group (RTM)	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
INTA (LIDORT)	1.03	1.00		1.09	1.02
IUP-Bremen (SCIATRAN)	1.03	0.99		1.08	0.99
MPIC	0.97	0.98		0.95	0.95

(SCIATRAN)					

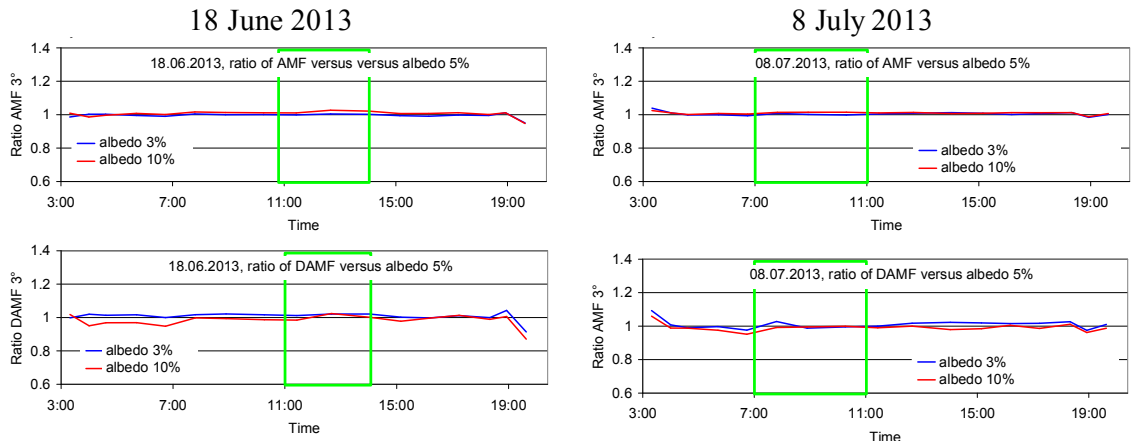


Fig. A12 Ratio of the O_4 AMFs (top) and O_4 dAMFs (bottom) for different surface albedos versus those for an albedo of 5 % for both selected days.

Table ~~A12~~ A10 Average ratios of O_4 (d)AMFs for different surface albedos versus those for an albedo of 5 % for the two middle periods on both selected days.

	AMF ratios			**dAMF ratios**	
Surface albedo	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
3 %	1.00	1.00		1.02	1.00
10 %	1.02	1.01		1.00	0.99

18 June 2013 8 July 2013

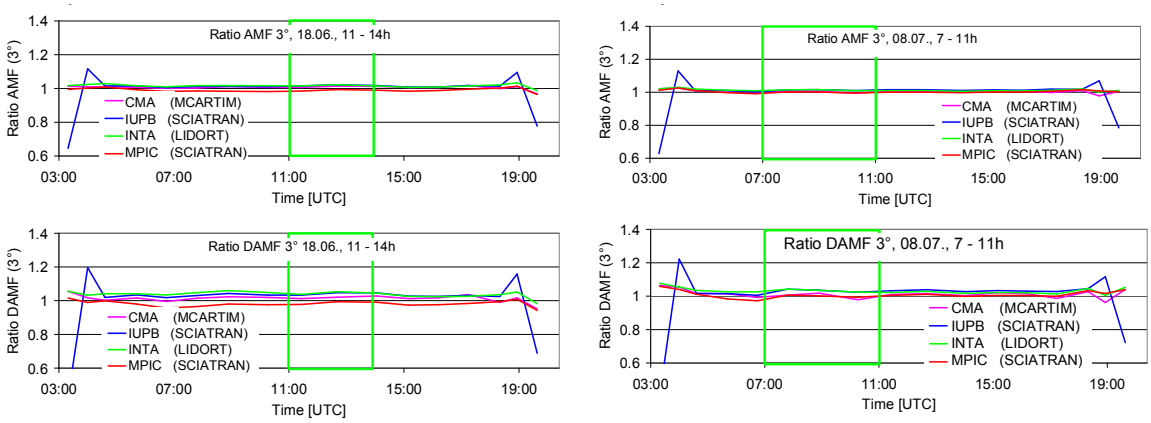


Fig. A13 Ratio of the O_4 AMFs (top) and O_4 dAMFs (bottom) simulated by different groups using different radiative transfer models versus those for the MPIC simulations using MCARTIM for both selected days.

Table A11 Average ratios of O_4 (d)AMFs simulated by different groups using different radiative transfer models versus those for the MPIC simulations using MCARTIM for the two middle periods on both selected days.

	AMF ratios			**dAMF ratios**	
Group (RTM)	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
CMA (M~~A~~CARTIM)	1.01	1.00		1.02	1.00
IUP-Bremen (SCIATRAN)	1.02	1.01		1.04	1.03
INTA (LIDORT)	1.02	1.01		1.05	1.03
MPIC (SCIATRAN)	0.99	1.00		0.99	1.00

18 June 2013 8 July 2013

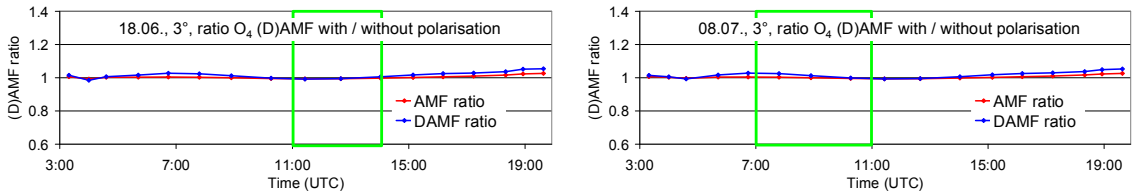


Fig. A14 Ratio of the O_4 (d)AMFs considering polarisation versus those without considering polarisation for both selected days.

Table A12 Average ratios of O_4 (d)AMFs considering polarisation versus those without considering polarisation for the two middle periods on both selected days.

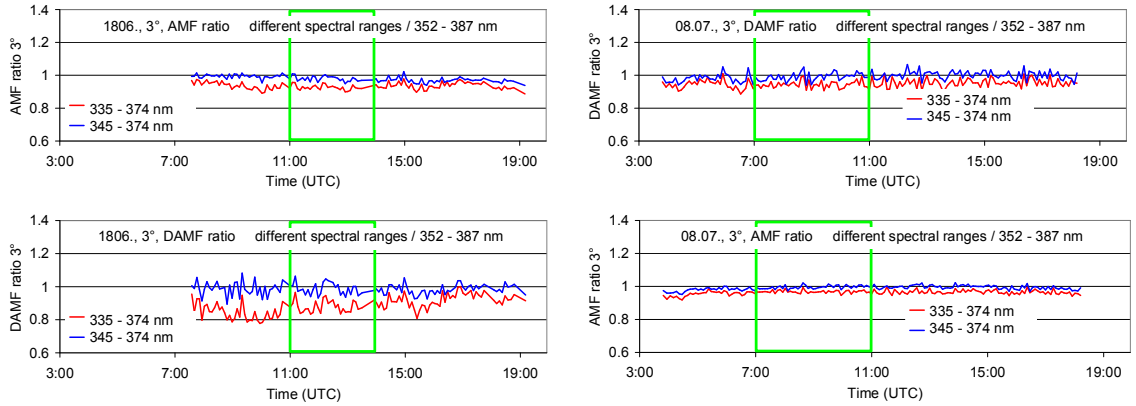
	AMF ratios			**dAMF ratios**	
	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Considering polarisation	1.00	1.00		1.00	1.01

Table A13 Average ratios of O_4 (d)AMFs derived from synthetic spectra versus those obtained from radiative transfer simulations at 360 nm for the two middle periods on both selected days.

	AMF ratios			**dAMF ratios**	
Temperature dependence / noise	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
T dep. considered / no noise	1.01	1.02		1.01	1.00
no T dep. considered / no noise	1.00	1.01		1.00	1.00
no T dep. considered / noise	0.99	1.00		1.00	1.01

18 June 2013 8 July 2013

a) measured spectra



b) synthetic spectra

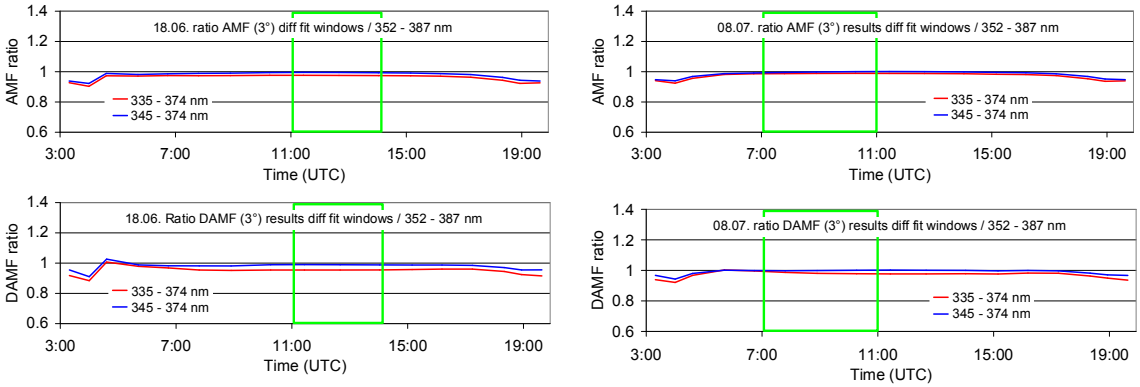


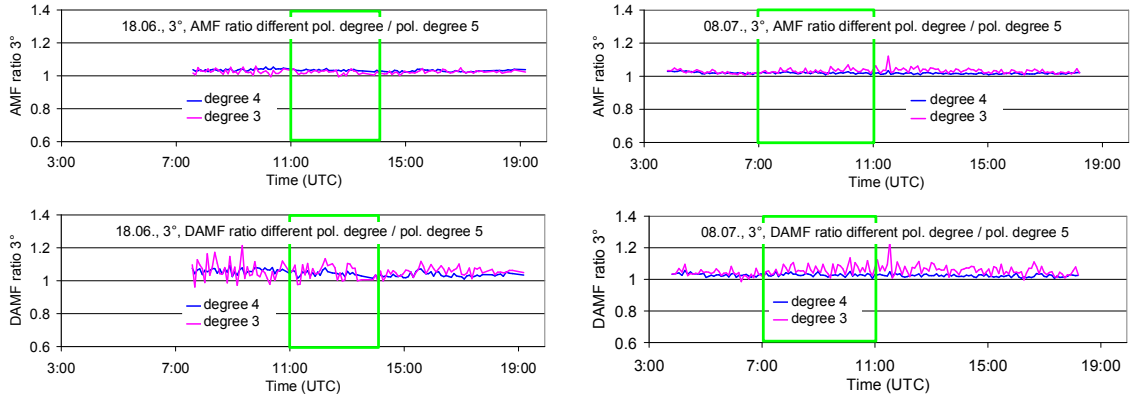
Fig. A15 Ratio of the O_4 (d)AMFs derived for different fit windows versus those for the standard fit window (352 – 387 nm) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

Table A14 Average ratios of O_4 (d)AMFs derived for different fit windows versus those for the standard fit window (352 – 387 nm) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			**dAMF ratios**	
Spectral range	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured Spectra					
335 – 374 nm	0.93	0.97		0.88	0.94
345 – 374 nm	0.98	1.00		0.99	0.99
Synthetic Spectra					
335 – 374 nm	0.98	0.99		0.95	0.98
345 – 374 nm	0.99	1.00		0.99	1.00

18 June 2013 8 July 2013

a) measured spectra



b) synthetic spectra

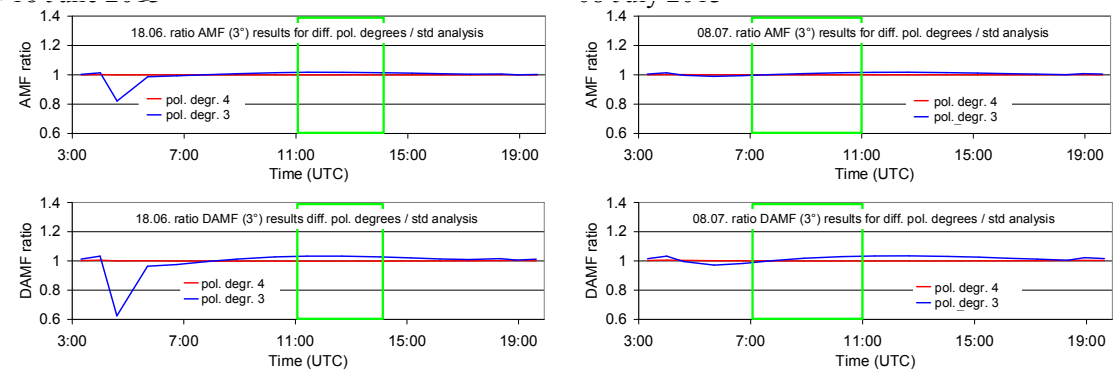


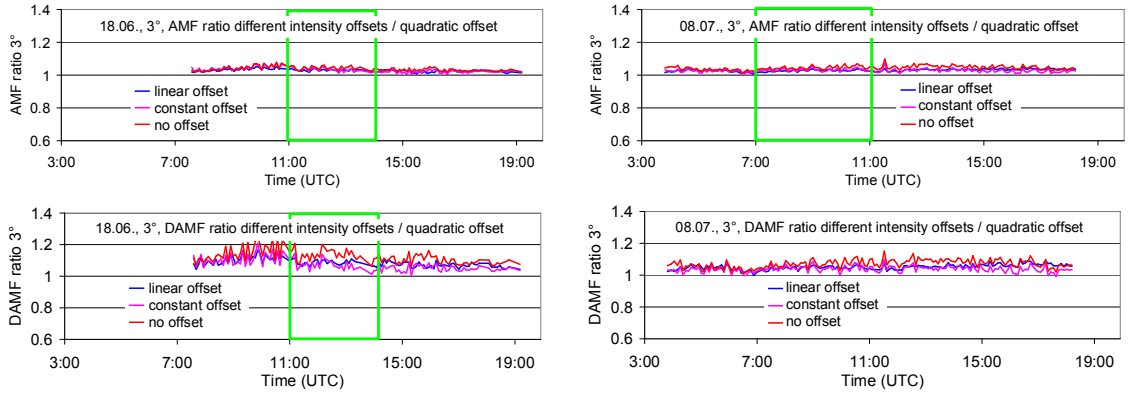
Fig. A16 Ratio of the O_4 (d)AMFs derived for different polynomials versus those for the standard analysis (polynomial degree 5) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

Table A15 Average ratios of O_4 (d)AMFs derived for different polynomials versus those for the standard analysis (polynomial degree 5) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			**dAMF ratios**	
Degree of polynomial	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured Spectra					
4	1.04	1.02		1.06	1.03
3	1.03	1.03		1.06	1.06
Synthetic Spectra					
4	1.00	1.00		1.00	1.00
3	1.02	1.01		1.03	1.01

18 June 2013 8 July 2013

a) measured spectra



b) synthetic spectra

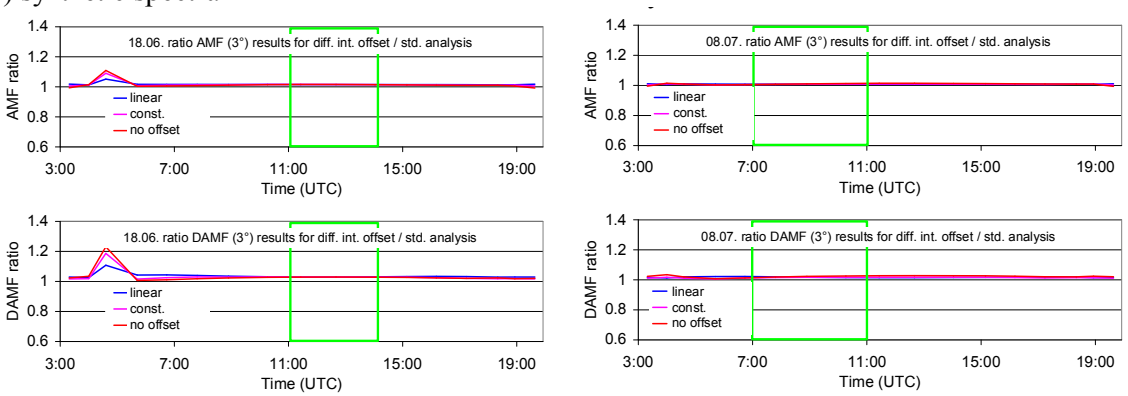


Fig. A17 Ratio of the O_4 (d)AMFs derived for different intensity offsets versus those for the standard analysis (intensity offset of degree 2) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

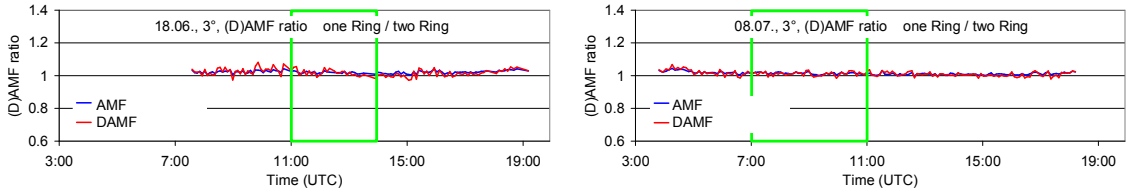
Table A16 Average ratios of O_4 (d)AMFs derived for different intensity offsets versus those for the standard analysis (intensity offset of degree 2) for the two middle periods on both

selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			dAMF ratios	
Intensity offset	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured Spectra					
Linear	1.04	1.03		1.11	1.05
Constant	1.05	1.03		1.11	1.04
No offset	1.05	1.05		1.16	1.07
Synthetic Spectra					
Linear	1.01	1.01		1.03	1.02
Constant	1.02	1.01		1.03	1.02
No offset	1.02	1.01		1.03	1.02

18 June 2013 8 July 2013

a) measured spectra



b) synthetic spectra

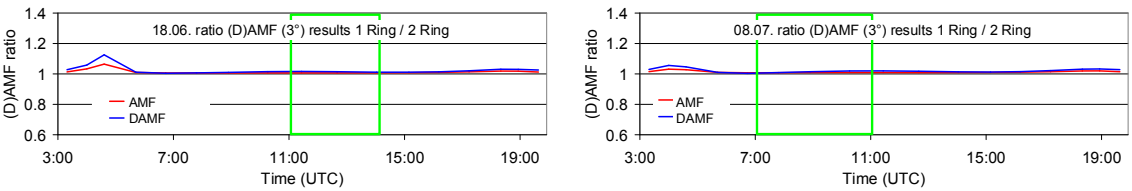


Fig. A18 Ratio of the O_4 (d)AMFs derived for the analysis with only one Ring spectrum versus those for the standard analysis (using two Ring spectra) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

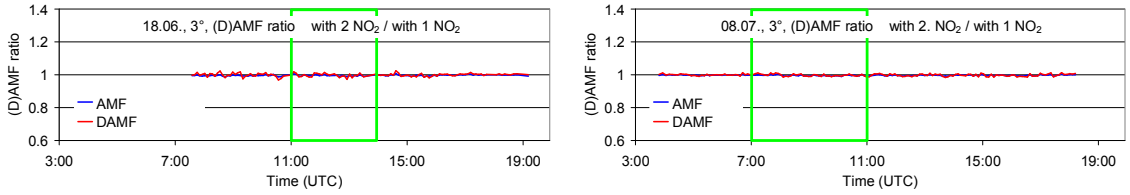
Table A17 Average ratios of O_4 (d)AMFs derived for the analysis with only one Ring spectrum versus those for the standard analysis (using two Ring spectra) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument;

bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			dAMF ratios	
Ring correction	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured Spectra					
Only one Ring spectrum	1.02	0.99		1.01	0.99
Synthetic Spectra					
Only one Ring spectrum	1.01	1.01		1.01	1.01

18 June 2013 8 July 2013

a) measured spectra



b) synthetic spectra

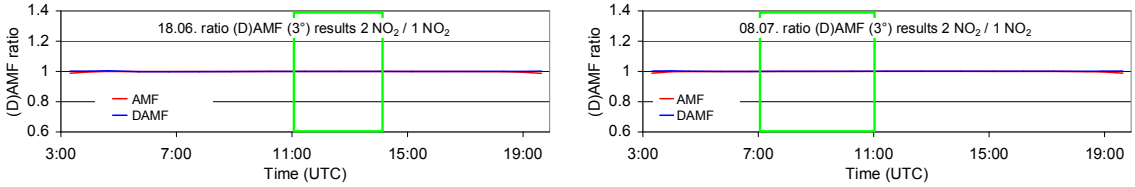


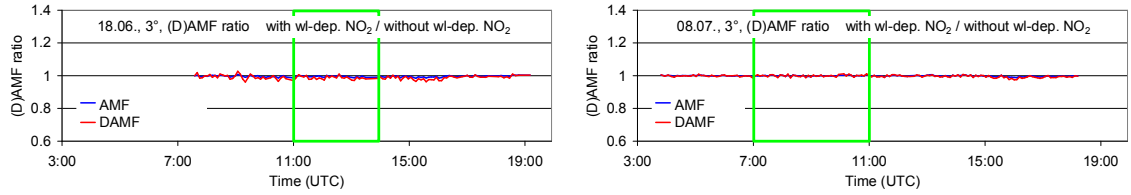
Fig. A19 Ratio of the O_4 (d)AMFs derived for the analysis with a second NO_2 cross section (for 220 K) versus those for the standard analysis (only NO_2 cross section for 294 K) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

Table A18 Average ratios of O_4 (d)AMFs derived for the analysis with a second NO_2 cross section (for 220 K) versus those for the standard analysis (only NO_2 cross section for 294 K) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			**dAMF ratios**	
NO_2 cross sections	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured Spectra					
294 & 220 K	1.00	1.00		1.00	1.00
Synthetic Spectra					
294 & 220 K	1.00	1.00		1.00	1.00

18 June 2013 8 July 2013

a) measured spectra



b) synthetic spectra

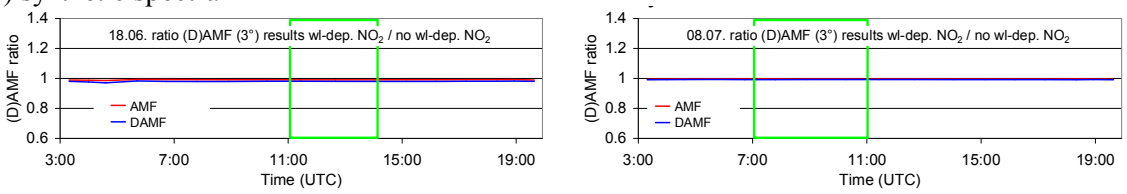


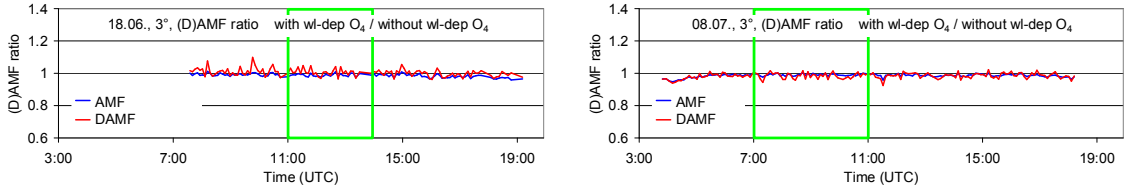
Fig. A20 Ratio of the O_4 (d)AMFs derived for the analysis with a second NO_2 cross section (cross section times wavelength) versus those for the standard analysis (only one NO_2 cross section) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

Table A19 Average ratios of O_4 (d)AMFs derived for the analysis with a second NO_2 cross section (cross section times wavelength) versus those for the standard analysis (only one NO_2 cross section) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			**dAMF ratios**	
NO_2 wavelength dependence	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured Spectra					
additional cross for wavelength dependence	1.00	1.00		0.99	1.00
Synthetic Spectra					
additional cross for wavelength dependence	0.99	1.00		0.98	0.99

18 June 2013 8 July 2013

a) measured spectra



b) synthetic spectra

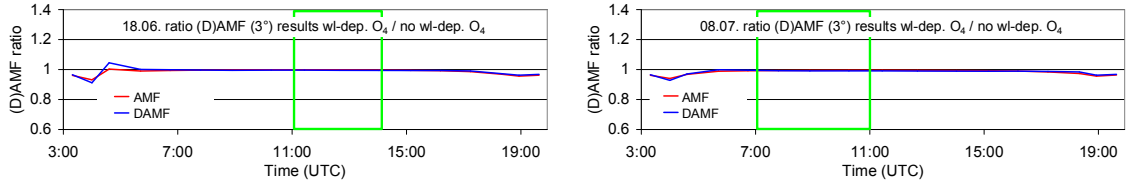


Fig. A21 Ratio of the O_4 (d)AMFs derived for the analysis with a second O_4 cross section (accounting for the wavelength dependence) versus those for the standard analysis (only one O_4 cross section) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

Table A20 Average ratios of O_4 (d)AMFs derived for the analysis with a second O_4 cross section (accounting for the wavelength dependence) versus those for the standard analysis (only one O_4 cross section) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			**dAMF ratios**	
O_4 wavelength dependence	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured Spectra					
additional cross for wavelength dependence	0.99	0.99		1.01	0.99
Synthetic Spectra					
additional cross for wavelength dependence	1.00	0.99		1.00	0.99

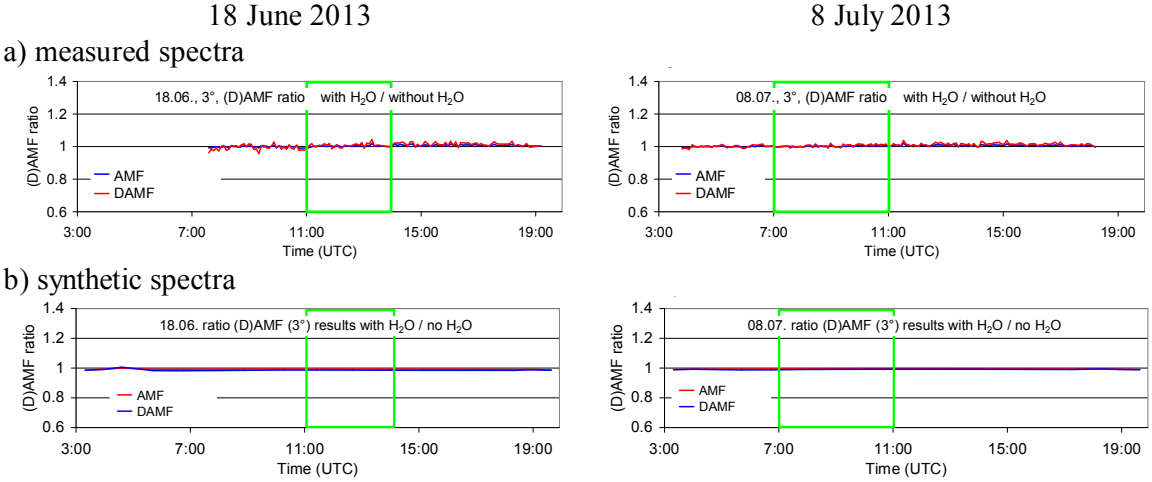


Fig. A22 Ratio of the O_4 (d)AMFs derived for the analysis including a H_2O cross section versus those for the standard analysis (no H_2O cross section) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

Table A21 Average ratios of O_4 (d)AMFs derived for the analysis including a H_2O cross section versus those for the standard analysis (no H_2O cross section) for the standard analysis (only one O_4 cross section) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			**dAMF ratios**	
H_2O cross section	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured spectra					
H_2O cross section included	1.00	1.00		1.01	1.01
Synthetic Spectra					
H_2O cross section included	0.99	1.00		0.99	0.99

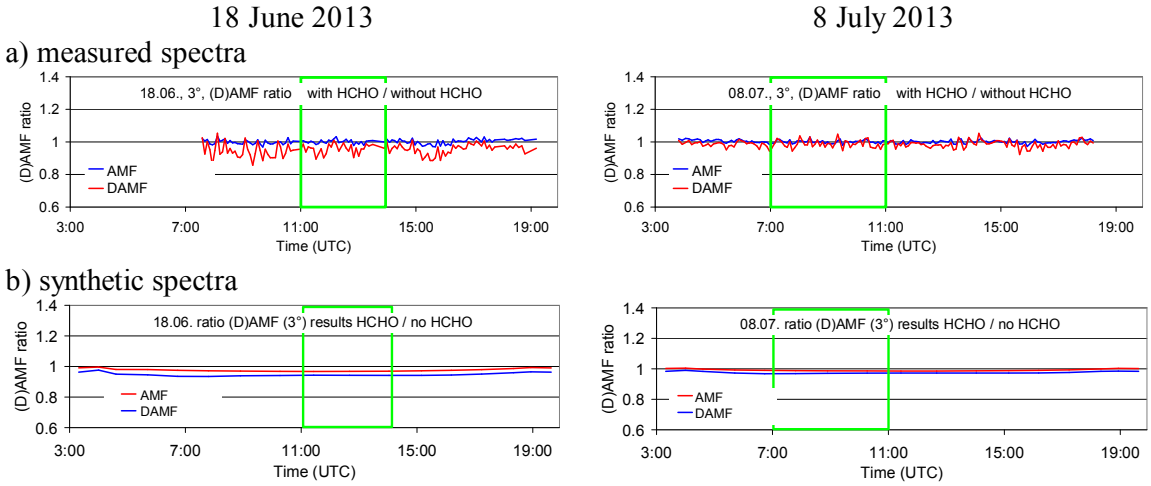


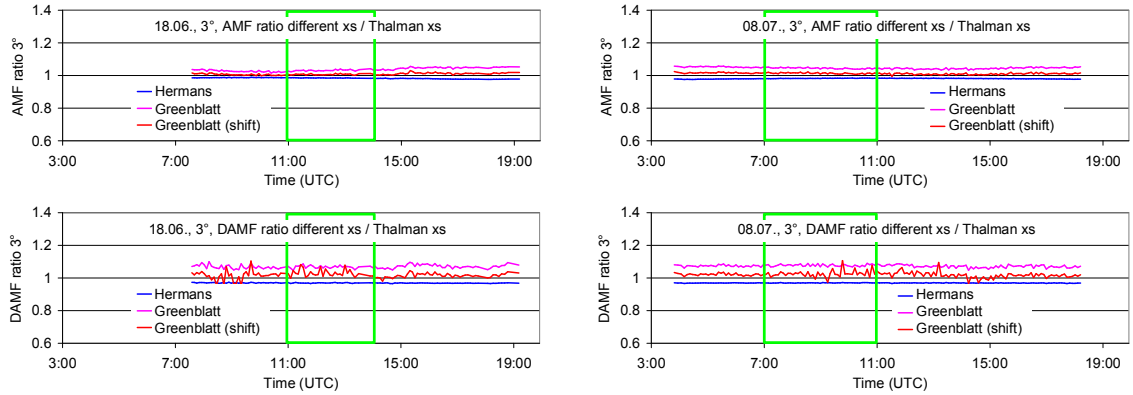
Fig. A23 Ratio of the O_4 (d)AMFs derived for the analysis including a HCHO cross section versus those for the standard analysis (no HCHO cross section) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

Table A22 Average ratios of O_4 (d)AMFs derived for the analysis including a HCHO cross section versus those for the standard analysis (no HCHO cross section) for the standard analysis (only one O_4 cross section) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			**dAMF ratios**	
HCHO cross section	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured Spectra					
HCHO cross section included	1.00	1.00		0.96	0.98
Synthetic Spectra					
HCHO cross section included	0.97	0.99		0.94	0.97

18 June 2013 8 July 2013

a) measured spectra



b) synthetic spectra

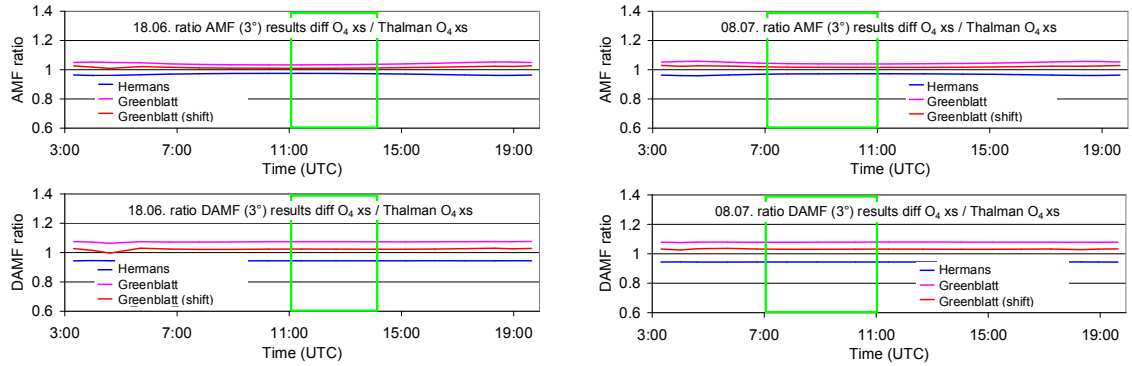


Fig. A24 Ratio of the O_4 (d)AMFs derived for the analyses using different O_4 cross sections versus those for the standard analysis (using the Thalman and Volkamer (2013) cross section) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

Table A23 Average ratios of O_4 (d)AMFs derived for the analyses using different O_4 cross section versus those for the standard analysis (using the Thalman ~~et al.~~and Volkamer cross section) for the standard analysis (only one O_4 cross section) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

	AMF ratios			**dAMF ratios**	
O_4 cross section	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured spectra					
Hermans	0.98	0.98		0.97	0.97
Greenblatt	1.03	1.04		1.07	1.08
Greenblatt shifted	1.01	1.01		1.03	1.03
Synthetic Spectra					
Hermans	0.97	0.97		0.94	0.94
Greenblatt	1.03	1.04		1.07	1.08
Greenblatt shifted	1.01	1.02		1.02	1.03

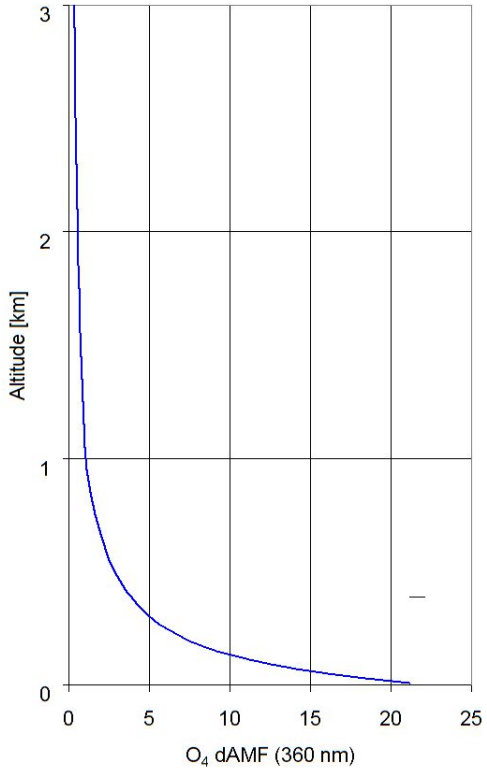


Fig. 25 O_4 differential box-AMFs (with 20m vertical resolution) used for the simulation of the temperature-dependent O_4 absorption spectra. They are averages of radiative transfer simulations for several scenarios. Simulations are performed for a surface albedo of 6 %, aerosol profiles with constant extinction between 0 and 1000m and different AOD (0.1, 0.3, 0.7) and for all combinations of SZA (40, 60°), relative azimuth angles (0, 90, 180°) and elevation angles (2° and 3°).

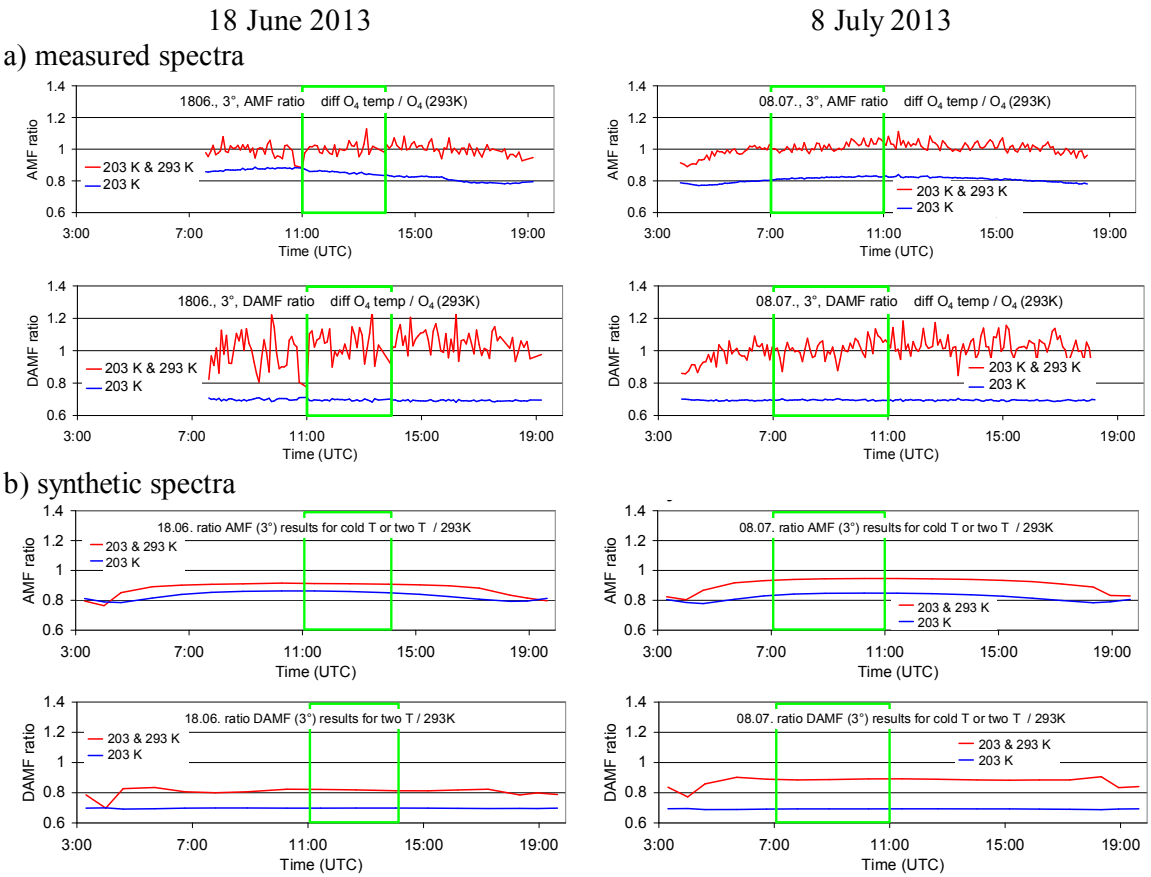


Fig. A26 Ratio of the O_4 (d)AMFs derived for O_4 cross sections at different temperatures (either 203 K or both 203 and 293 K) versus those for the standard analysis (using the O_4 cross section for 293 K) for both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section).

Table A24 Average ratios of O_4 (d)AMFs derived O_4 cross sections at different temperatures (either 203 K or both 203 and 293 K) versus those for the standard analysis (using the O_4 cross section for 293 K) for the two middle periods on both selected days (top: results for spectra measured by the MPIC instrument; bottom: results for synthetic spectra taking into account the temperature dependence of the O_4 cross section). For the simultaneous fit of both temperatures also the results for the spectral range 345 – 374 nm (one O_4 absorption band) are included.

	AMF ratios			**dAMF ratios**	
O_4 cross sections	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
Measured Spectra					
203 K	0.85	0.82		0.70	0.70
203 & 293 K	1.00	1.02		1.04	1.01
203 & 293 K (345 – 374 nm)	0.91	1.04		0.95	1.02
Synthetic Spectra					
203 K	0.86	0.84		0.70	0.69
203 & 293 K	0.91	0.94		0.82	0.89
203 & 293 K (345 – 374 nm)	0.99	1.00		0.99	1.00

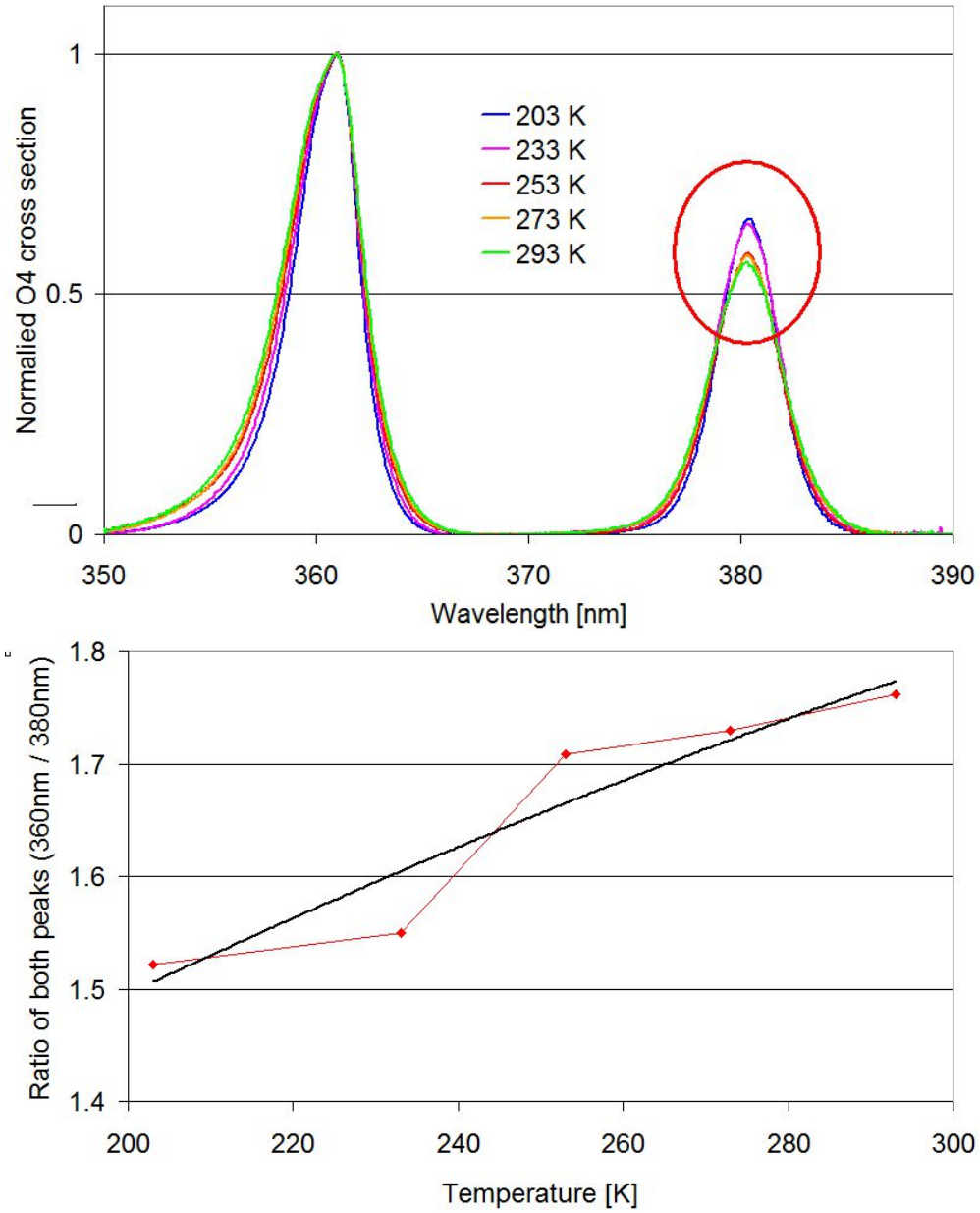


Fig. A27 Top: Comparison of the O_4 cross sections from Thalman and Volkamer (2013) for different temperatures. The cross sections are divided by the maximum values at 360 nm. After this normalisation, the resulting values at 380 nm fall into two groups (high values for 203 & 233K, low values for 253, 273, 293K). Bottom: Ratio of the peaks of the O_4 cross section at 360 nm and 380 nm as function of temperature (red points). The black curve is a fitted low order polynomial.

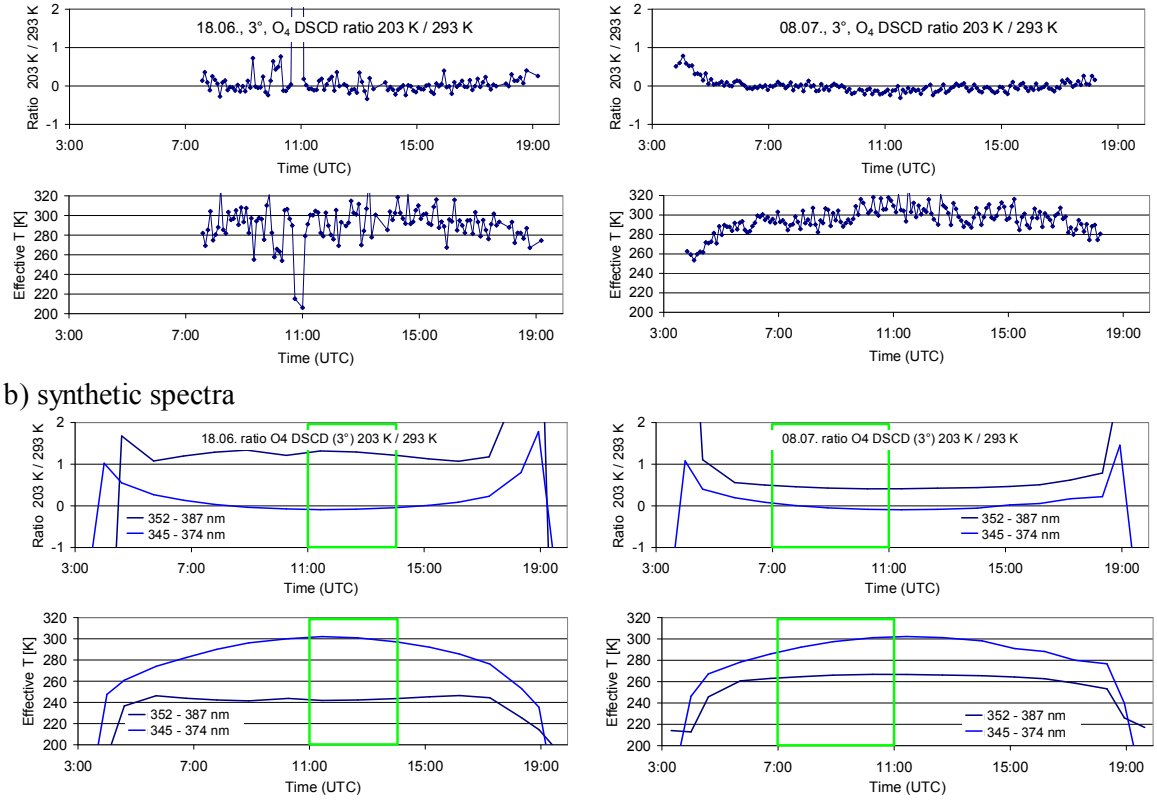


Fig. A28 Ratio of the derived O_4 dSCDs for 203 K and 293 K as well as the derived effective temperatures for the analyses with both cross sections included.

Table A25 a) Average ratios of O_4 (d)AMFs derived from the analysis of MPIC spectra by different groups versus the analysis of MPIC spectra by MPIC (standard analysis). b) Average ratios of O_4 (d)AMFs derived from spectra of other groups analysed by MPIC versus the analysis of MPIC spectra by MPIC (using the same analysis settings and spectral range: 335 – 374 nm). c) Average ratios of O_4 (d)AMFs derived from spectra of other groups analysed by the same groups using individual analysis settings versus the analysis of MPIC spectra by MPIC (standard analysis).

	AMF ratios			**dAMF ratios**	
Measurements / Analysis	18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00		18 June 2013, 11:00 – 14:00	8 July 2013, 7:00 – 11:00
a) MPIC spectra analysed by other groups					
BIRA	0.96	0.98		0.95	0.95
IUP-B	1.03	0.98		1.05	0.99
INTA	1.02	0.97		1.05	0.94
CMA	0.97	0.98		0.98	0.95
CSIC	0.94	0.94		0.95	0.94
b) Other spectra analysed by MPIC (335 – 374 nm)					
BIRA	0.98	0.99		0.89	0.95
IUP-B	1.05			1.07	
IUP-HD	0.97			1.00	
c) Other spectra analysed by the same groups					
BIRA	0.94	0.94		0.91	0.92
IUP-B	0.95			0.88	
IUP-HD	1.01			1.04	

Appendix A5 Extraction of aerosol extinction profiles

In this section, the procedure for the extraction of aerosol extinction profiles is described. The aerosol profiles are derived from the ceilometer measurements (yielding the profile information) in combination with the sun photometer measurements (yielding the vertically integrated aerosol extinction, the aerosol optical depth AOD).

The ceilometer raw data consist of range-corrected backscatter profiles averaged over 15 minutes. The profiles range from the surface to an altitude of 15360m with a height resolution of 15m. Here it is important to note that due to limited overlap of the outgoing Laser beam~~n~~ and the field of view of the telescope, no profile data is available below 180 m. The ceilometer profiles (hourly averages) are shown in Fig. A29 for both selected days.

The AERONET sun photometer data provide the AOD at different wavelengths (340, 360, 440, 500, 675, 870, and 1020 nm) in time intervals of 2 – 25 min if the direct sun is visible.

To determine profiles of aerosol extinction from the ceilometer backscatter data, several processing steps have to be performed. They are described in the sub-sections below. Note that in this section the individual steps are described according to the MPIC procedure. The extracted profiles from other groups differ slightly compared to the results of the MPIC procedure, especially with respect to the altitude above which the extinction was set to zero (see Fig. 9).

A) Smoothing and extrapolating of the ceilometer backscatter profiles

First, the ceilometer data are averaged over several hours to reduce the scatter. For that purpose on both days three time periods are identified, for which the backscatter profile show relatively small variations. The profiles for these periods are shown in Fig. A29. In addition to the temporal averaging, the profiles are also vertically smoothed above 2 km. Above altitudes between 5 to 6 km (depending on the period) the (smoothed) ceilometer backscatter profiles become zero. Thus the aerosol extinction profiles above these altitudes are set to zero. Below 180 m above the surface the ceilometer becomes 'blind' for the aerosol extinction because of the insufficient overlap between the outgoing laser beam and the field of view of the telescope. Thus the profiles have to be extrapolated down to the surface. This extrapolation constitutes an important source of uncertainty. To estimate the associated ~~errors~~uncertainties, the extrapolation is performed in three different ways:

1) The value below 180 m are set to the value measured at 180m.

2) The values below 180m are linearly extrapolated assuming the same slope below 180 m as between 180m and 240m.

3) The values below 180m are linearly extrapolated by twice the ~~double~~ slope between 180m and 240m.

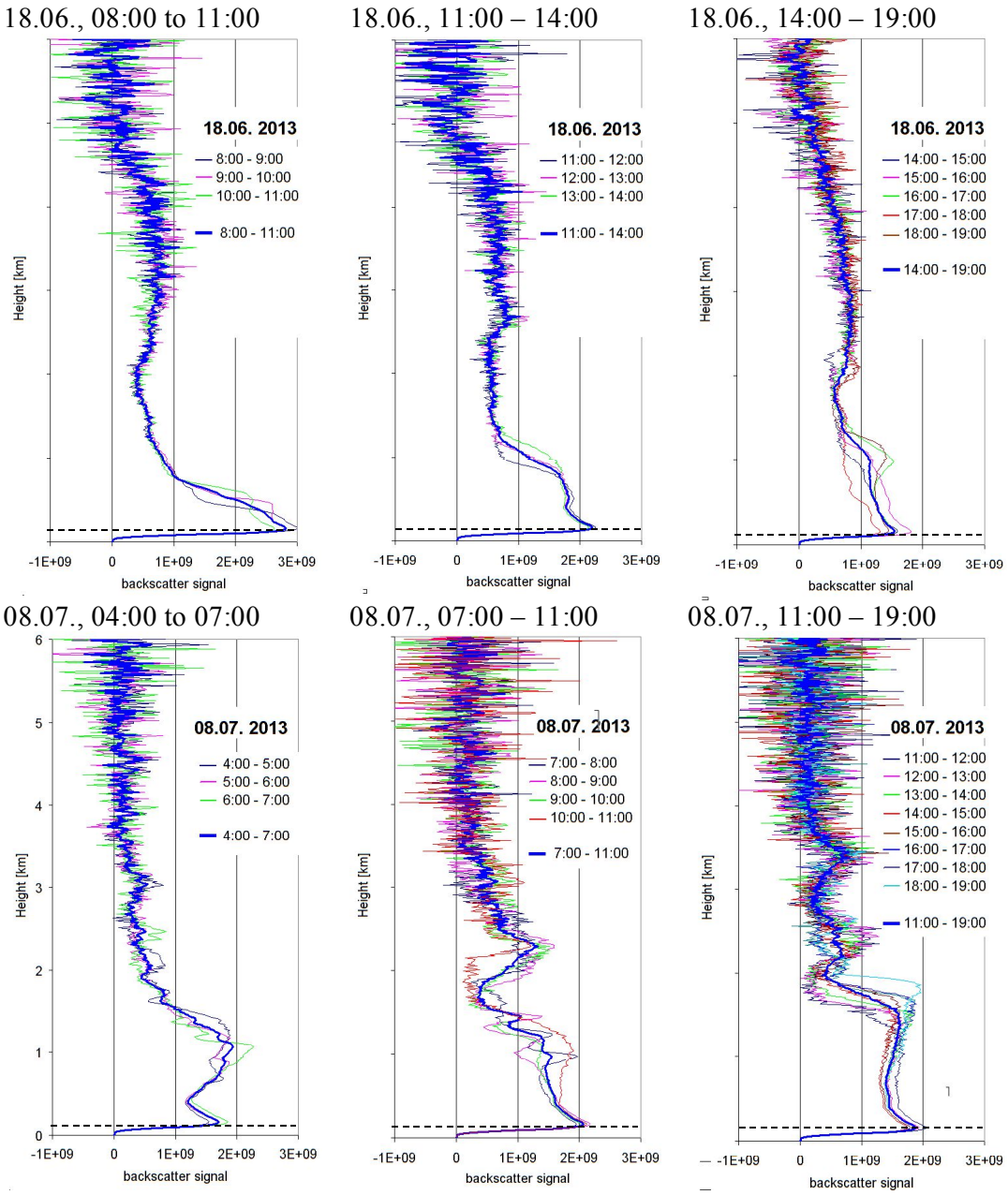


Fig. A29 Range-corrected backscatter profiles (hourly averages) for the three selected periods on both days. Also the averages over the ~~the~~ whole periods are shown (thick lines). Note that the backscatter signal below 180 m (below the dashed horizontal line) is invalid due to the limited overlap of the ceilometer instrument.

B) Scaling of the Ceilometer profiles by sun photometer AOD at 1020 nm

The scaling of the ceilometer backscatter profiles by the AOD at 1020 nm is an intermediate step, which is necessary for the correction of the aerosol self-extinction. The average AOD at

1020 nm for the different selected time periods on both days is shown in Table A26. In that table also the average values at 380 nm are shown, which are used for a second scaling (see below).

The backscatter profiles are vertically integrated and then the whole profiles are scaled by the ratio:

$$AOD_{1020nm} / B_{int} \quad (A1)$$

Here B_{int} indicates the integrated backscatter profile.

Note that the wavelength of the ceilometer measurements (1064 nm) is slightly different from the sun photometer measurements (1020 nm), but the difference of the AOD is negligible (typically < 4%).

Table A26 Average AOD at 1020 and 360 nm derived from the sun photometer.

Time	AOD 1020 nm	AOD 360 nm*
18.06.2013, 08:00 - 11:00	0.124	0.379
18.06.2013, 11:00 - 14:00	0.122	0.367
18.06.2013, 14:00 - 19:00	0.118	0.296
08.07.2013, 04:00 - 07:00	0.045	0.295
08.07.2013, 07:00 - 14:00	0.053	0.333
08.07.2013, 11:00 - 19:00	0.055	0.348

*Average of AOD at 340 nm and 380 nm.

C) Correction of the aerosol extinction

The photons received by the ceilometer have undergone atmospheric extinction. Here, Rayleigh scattering can be ignored because of the long wavelength of the ceilometer (optical depth below 2 km is < 0.001). However, while the extinction due to aerosol scattering is also small at these long wavelengths it systematically affects the ceilometer signal and has to be corrected. The extinction correction is performed according to the following formula:

$$\alpha_{i,corr} = \frac{\alpha_i}{\exp\left(-2 \cdot \sum_{j=0}^{i-1} \alpha_{j,corr} \cdot \left(z_j - z_{j-1}\right)\right)} \quad (A2)$$

Here α_i represent the uncorrected extinction and $\alpha_{i,corr}$ represents the corrected extinction at height layer i (with z_i is the lower boundary of that height layer). Equation C1 has to be subsequently applied to all height layers starting from the surface (z_0). Note that the factor of two accounts for the extinction along both paths between the instrument and the scattering altitude (~~way~~ upward and downward~~up and down~~). The extinction correction is performed at a vertical resolution of 15m.

After the extinction correction, the profiles are scaled by the corresponding AOD ~~a~~at 360 nm (see table A26). In Fig. A30 the profiles with and without extinction correction are shown. The extinction correction slightly increases the values at higher altitudes and decreases the values close to the surface. The effect of the extinction correction is larger on 18 June 2013 (up to 12 %).

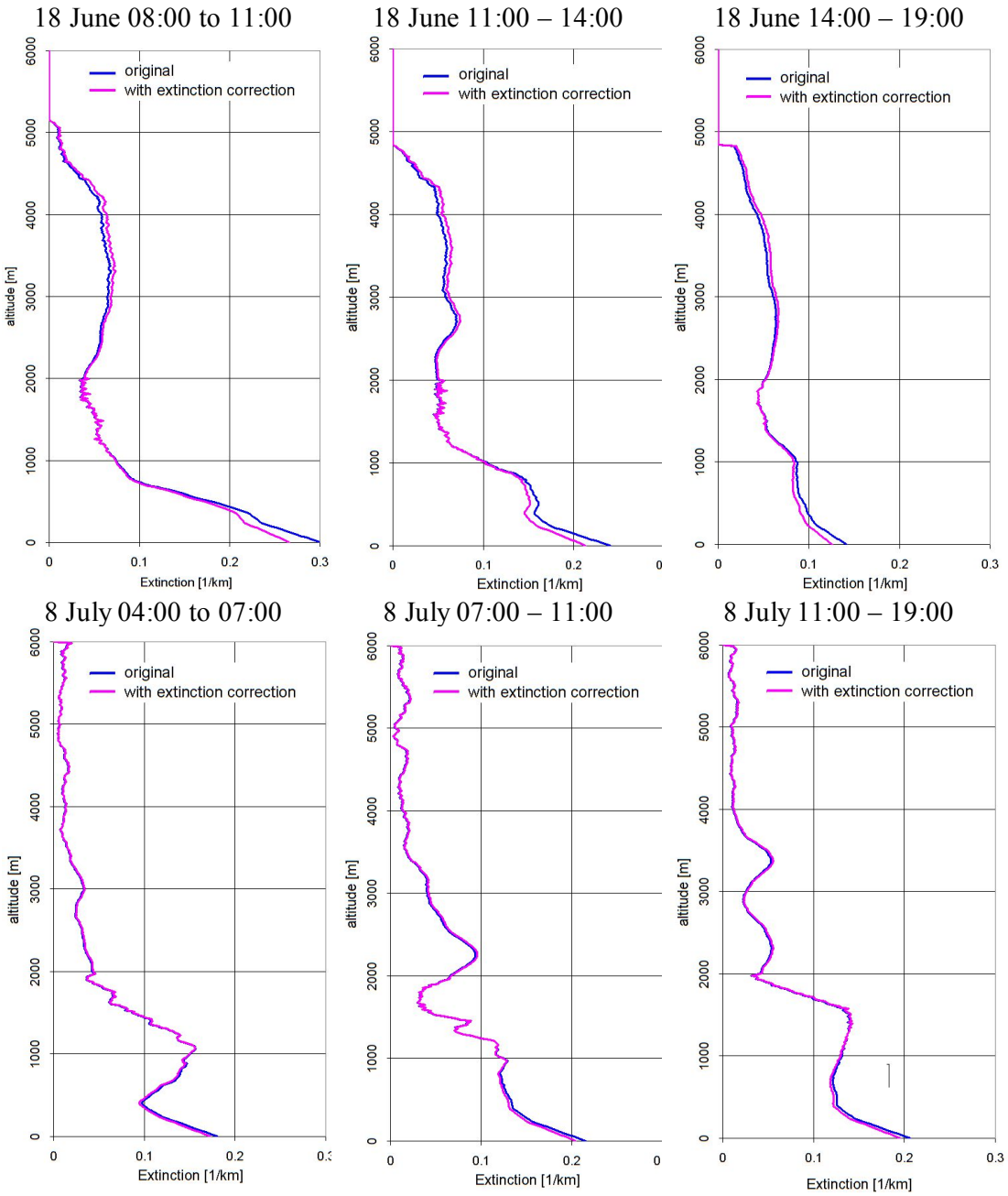


Fig. A30 Comparison of profiles (linear extrapolation below 180 m) without (blue) and with (magenta) extinction correction. Both profiles are scaled to the same total AOD (at 360 nm) determined from the sun photometer.

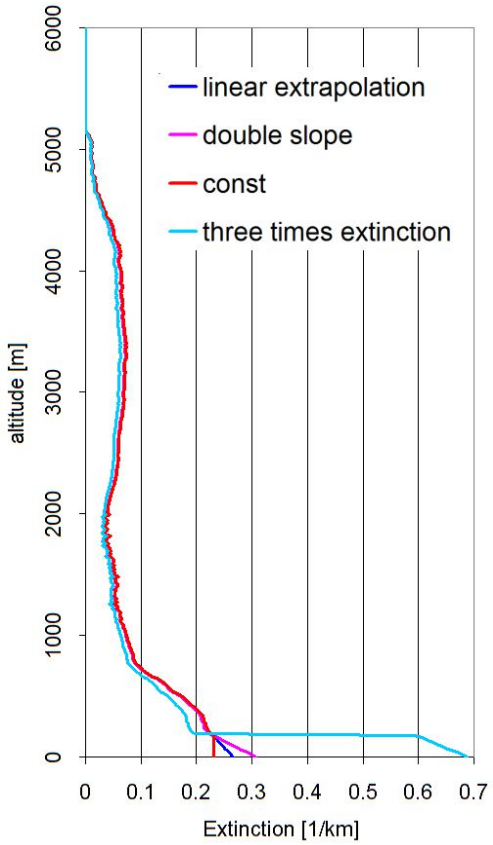
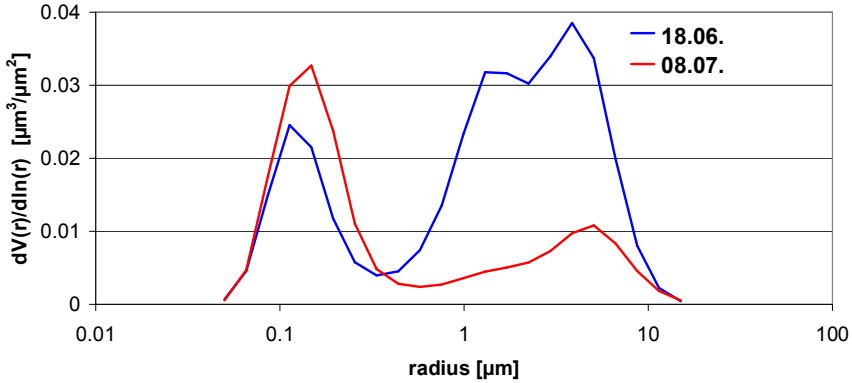


Fig. A31 Aerosol profile (light blue) with extreme extinction close to the surface (below 180 m, the altitude for which the ceilometer is sensitive) extracted for the first period (8:00 – 11:00) on 18 June 2013. Also shown are the profiles extrapolated below 180 as described above.

D) Influence of a changing LIDAR ratio with altitude

For the extraction of the aerosol profiles described above, a fixed LIDAR ratio was assumed, which implies that the aerosol properties are independent from altitude. However, this is a rather strong assumption, because it can be expected that the aerosol properties (e.g. the size) change with altitude. With the available limited information, it is impossible to derive detailed information about the altitude dependence of the aerosol properties, but it can be quantified how representative the ceilometer measurements at 1064 nm are for the aerosol extinction profiles at 360 nm. For these investigations we again focus on the middle periods of both selected days. From the AERONET Almucantar observations information on the size distribution for these periods is available (see Fig. A32). On both days two pronounced modes (fine and coarse mode) are found with a much larger coarse mode fraction on 18 June compared to 8 July (on 18 June also the coarse mode is broader and shows two distinct maxima). From the AERONET observations, also separate phase functions for the fine and coarse mode as well as the relative contributions of both modes to the total aerosol optical depth at 500 nm are available. On 18 June and 8 July the relative contributions of the coarse mode fraction to the total AOD at 500 nm are about 39 % and 5 %, respectively (see table A27). Assuming that the AOD of the coarse mode fraction is independent of wavelength, the relative contributions of the coarse mode at 360 nm and 1064 nm can be derived (see Table A27).



Fg. A32 Size distributions derived from AERONET Almucantar observations on 18 June (07:24 & 15:34) and 08 July (07:32 & 15:38).

Table A27 Contributions of the coarse mode to the total AOD at different wavelengths derived from AERONET observations. The relative contributions are calculated assuming that the AOD of the coarse mode at 500 nm (0.093 and 0.010 on 18 June and 8 July, respectively) does not depend on wavelength.

Date	Total AOD 360 nm	Total AOD 500 nm	Total AOD 1064 nm*	Relative contribution of coarse mode 360 nm	Relative contribution of coarse mode 500 nm	Relative contribution of coarse mode 1064 nm
18 June, 11:00 – 14:00	0.37	0.242	0.119	24.9%	38.7%	77.7%
08 July, 07:00 – 11:00	0.33	0.207	0.0535	3.0%	4.8%	18.7%

*extrapolated from the measurements at 675 nm and 1020 nm)

It is found that on 18 June the coarse mode clearly dominates the AOD at 1064 nm, whereas on 8 July it only contributes about 20 % to the total AOD. As expected the relative contributions of the coarse mode to the AOD at 360 nm are much smaller (25 % and 3%).
In the last step the probability of aerosol scattering in backward direction is considered, because the ceilometer receives scattered light from that direction. For that purpose the ratios of the optical depths are multiplied by the corresponding values of the normalised phase functions at 180° and in this way the relative contributions to the backscattered signals from the coarse mode for both wavelengths and both days are calculated (Table A28). Interestingly, on 8 July the contributions of the coarse mode to the backscattered signal at both wavelengths differs by only about 10%. In contrast, on 18 June the difference is much larger.

Table A28 Ratio of phase functions (coarse / fine) in backward direction and relative contribution of coarse mode to the backscattered signal at both wavelengths

Date	Ratio phase function at 360 nm	Ratio phase function at 1064 nm	Relative contribution of coarse mode at 360 nm	Relative contribution of coarse mode at 1064 nm
18 June, 11:00 – 14:00	1.13	0.61	27.3%	68.0%
08 July, 07:00 – 11:00	2.7	0.99	7.8%	18.3%

For 8 July, the results can be interpreted in the following way: at 360 nm the aerosol profiles extracted as described above overestimate the contribution from the coarse mode by about 10%. To estimate the effect of this overestimation we construct modified aerosol extinction profiles, in which 10% of the total AOD is relocated. Since we expect that the coarse mode aerosols are usually located at low altitude, we construct 4 different modified profiles (see Fig. A33) with different altitudes (1.5 km, 1 km, 0.75 km, or 0.5 km), below which 10% of the aerosol extinction is relocated to altitudes above (assuming that the coarse mode aerosol is only located below these altitudes). Of course, such a sharp boundary is not very realistic, but it allows to quantify the overall effect of the relocation. We selected the aerosol profile for 8 July extracted by INTA, which reached up to 7 km (see Fig. 9). It should be noted that if 10 % of the total AOD is relocated from the lowest layer to only the upper most layer no further enhancement of the O_4 dAMF is found (see appendix A6).

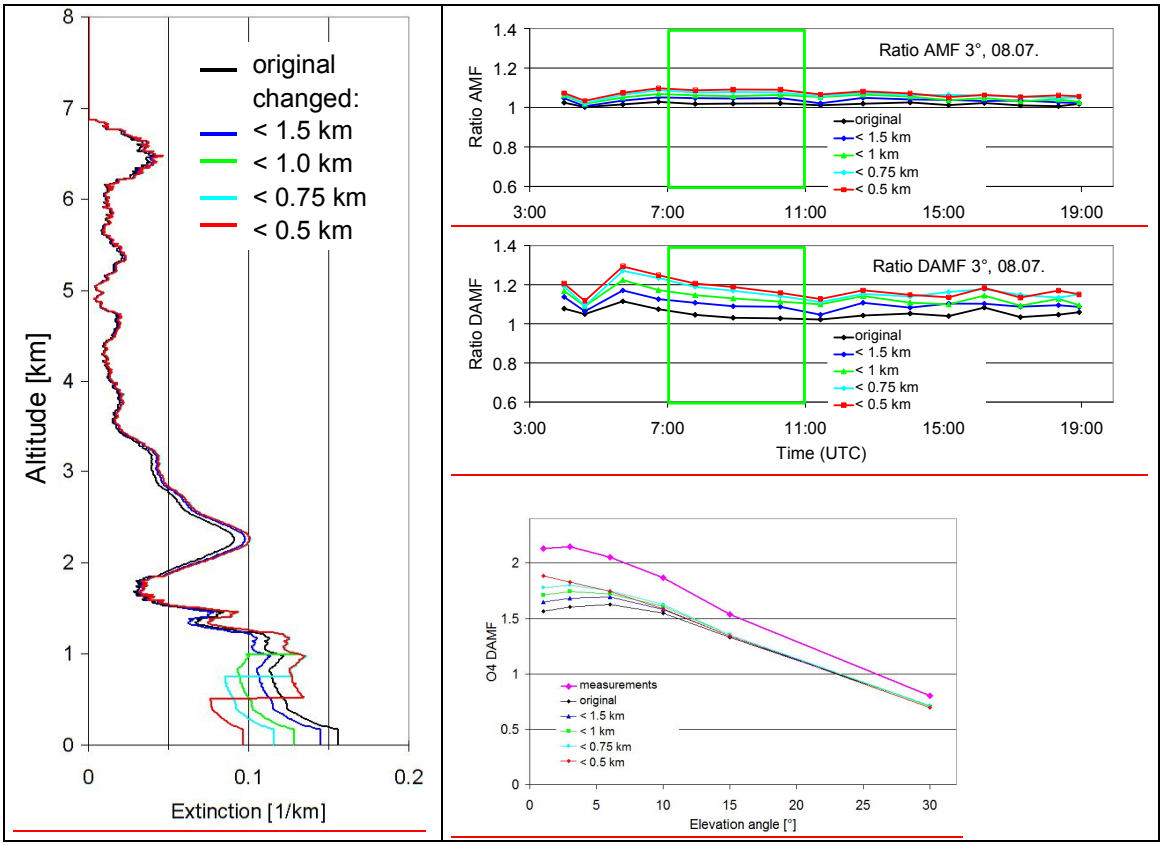


Fig. A33 Left: Modified aerosol profiles for 08 July assuming that the coarse mode aerosol is only located in the lowest part of the atmosphere. Top right: ratios of the (d)AMFs calculated for the modified profiles compared to the dAMFs for the standard settings. With decreasing layer height the (d)AMFs increase systematically, because the aerosol extinction close to the surface decreases. Right bottom: comparison of the measured elevation dependence of the O_4 dAMFs for the period 7:00 – 11:00 on 8 July and simulation results for the different profiles.

Table A29 Ratio of the (d)AMFs for the modified profiles versus those of the standard settings

	original INTA	coarse mode below 1.5 km	coarse mode below 1 km	coarse mode below 0.75 km	coarse mode below 0.5 km
AMF	1.02	1.04	1.05	1.06	1.08
dAMF	1.04	1.09	1.13	1.17	1.18

For all modified profiles, a systematic increase of the O_4 (d)AMFs compared to those for the standard settings is found. For the O_4 dAMFs this increase can be up to 18 % (see Table A29. From the comparison of the elevation dependence of the measured and simulated O_4 dAMFs (see Fig. A33), we conclude that the aerosol profile with the coarse mode aerosol below 0.75 km is probably the most realistic one. The main conclusion from this section is that the dAMF for 8 July derived from the standard settings probably underestimates the true dAMF by about 17±5 %.

For 18 June we did not perform similarly detailed calculations, because on that day the uncertainties of the aerosol extinction profile caused by the missing sensitivity of the ceilometer below 180 m are much larger than on 8 July. On 18 June also the magnitude of the relocation of the aerosol extinction between different altitudes would be much larger than on 8 July.

Appendix A6 Influence of elevated aerosol layers on the O_4 (d)AMF

Ortega et al. (2016) showed that for their measurements the consideration of elevated aerosol layers (between about 3 and 5 km) is essential to bring measured and simulated O_4 (d)AMFs into agreement. They also used LIDAR measurements at similar wavelengths as the MAX-DOAS observations. In our study, we consider aerosol layers over an even larger altitude range (up to 7 km). Nevertheless, it is interesting to see how the simulated O_4 (d)AMFs change if the extinctions at various altitude ranges are changed systematically. Here we chose the aerosol extinction profile extracted by INTA for the period 7:00 to 11:00 on 8 July, because it contains substantial amounts of aerosols in elevated layers (see Fig. 9). During that period three distinct aerosol layers can be identified (see Table A30).

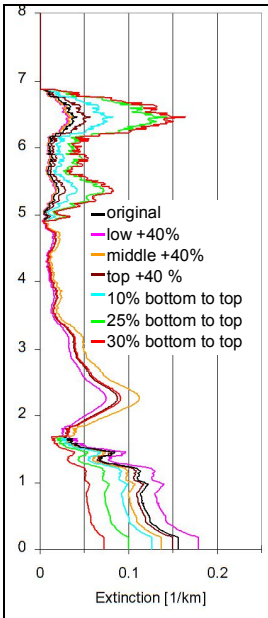
Table A30 Selection of different aerosol layers on 08 July (07:00 – 11:00)

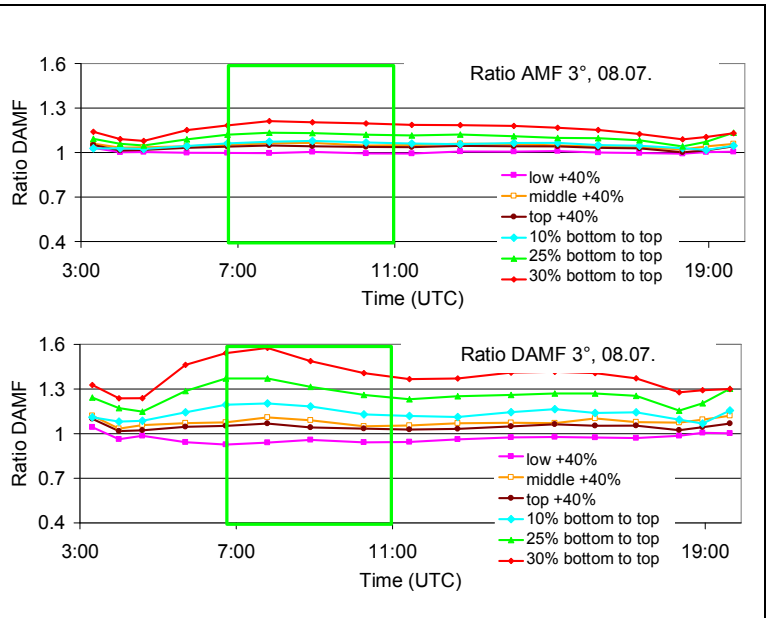
layer	AOD	Relative contribution to total AOD
0 – 1.68 km	0.186	55.4 %
1.68 – 4.9 km	0.116	34.5 %
4.9 – 7 km	0.035	10.4 %

Then, the extinction of the individual aerosol layers were increased by 40 % compared to the original profile. After that modification the whole profiles are scaled with a constant factor to match the AOD of the sun photometer observations. The modified profiles are then used for the simulation of O_4 (d)AMFs. A second set of profiles was created to investigate the effect of extreme relocations: here certain fractions (10%, 25% or 30%) of the total AOD were relocated from the bottom layer to the top layer.

The modified profiles and the ratios of the corresponding O_4 (d)AMFs versus the O_4 dAMFs of the original profile are shown in Fig. A34. For the O_4 AMFs the relocations of the extinction profiles lead to a general increase of the O_4 AMFs of up to 20%. For the O_4 dAMFs for most modified profiles a strong increase compared to the original profile is found. Only for the profile with an increase of the extinction in the lowest layer a slight decrease is observed. For the profiles with the extreme relocations the increase of the O_4 dAMFs reaches almost 50%.

From these results it can be concluded that for a relocation of about 27% almost perfect agreement with the measurements is found (see Fig. A34). For such an aerosol profile simulations and measurements could be brought into agreement without a scaling factor. However, such a large redistribution is not supported by the AERONET inversion products (see appendix A5). It should also be noted that for such a profile, about 73% of the total AOD would be located above about 1.7km. Moreover, for such an aerosol profile it is found that the simulated O_4 AMFs for 90° elevation systematically underestimate the measured O_4 AMFs at high SZA by about 15% (see Fig. A34), whereas much better agreement is found for the standard settings. The underestimation of the O_4 AMFs for 90° elevation is caused by the high aerosol amount at high altitudes, which increases the scattering altitude of the solar photons observed at 90° elevation. A similar effect could be caused by cirrus clouds, but on the selected days there are no indications for such clouds in the ceilometer data.





Diurnal variation of the O_4 AMF for 90° elevation	Elevation dependence of the O_4 dAMF

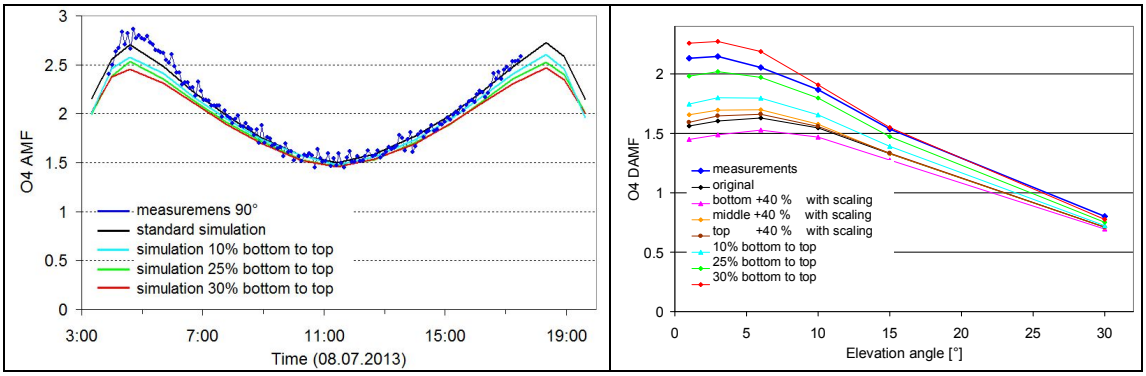


Fig. A34 Top left: Aerosol profiles used for the simulations (see text). Top right: Ratios of the O_4 (d)AMFs simulated for the modified profiles versus those of the original profile. Bottom: comparison of the measured diurnal variation (SZA dependence) for 90° elevation (left), and the elevation dependence of the O_4 dAMFs for the period 7:00 – 11:00 on 8 July (right).

Table A31 Ratios of (d)AMFs for 8 July 2013 for the modified profiles with respect to the original profile

	low +40 %	middle +40 %	top +40 %	10% bottom to top	25% bottom to top	30% bottom to top
AMF	1.00	1.06	1.04	1.07	1.12	1.20
dAMF	0.94	1.08	1.04	1.17	1.31	1.48