Authors' response for AMT-2014-243 "MAX-DOAS observations of aerosols, formaldehyde and nitrogen dioxide in the Beijing area: comparison of two profile retrieval approaches" by T. Vlemmix et al.

The authors greatly acknowledge two anonymous referees for carefully reading the manuscript, for providing positive feedback and providing many useful suggestions for improvements.

In this document we will answer to the questions asked by the referees, and indicate how this lead to changes in the revised version of the manuscript. The original comments by the reviewers were copied to this document and printed in italics. Referee comments are given numbers as C1.2 for the second comment by referee #1. The authors' response (**AR**) is written as normal text, and indicated by **AR**. Pages and lines in the original (AMTD) manuscript are referred to as 'OM p. X, l. Y' (for instance, OM p.9675, l. 17).

Referee #1 (Oetjen)

This manuscript provides a nice and very detailed overview of the limitations and possibilities for the MAX-DOAS profile retrievals. I agree with the other reviewer that there are some sections which can be improved, but that there is otherwise no reason why this manuscript should not be accepted. There is one detail I would like to see addressed beforehand though: the manuscript lacks a few basic explanations of the MAX-DOAS technique and hence might be difficult to understand for someone outside the field:

C 1.1*: P. 9675, l. 17-20: Please add that the profile information is obtained from scanning the atmosphere at several elevation angles.*

AR 1.1: These lines have been slightly rephrased:
"MAX-DOAS instruments take spectral measurements in the ultraviolet (UV) and visible (Vis) part of the electromagnetic spectrum. Profile information is obtained from a scan which comprises measurements which are all in the same azimuthal direction but at different elevation angles. The main retrieval products are tropospheric columns, concentrations near the surface and estimates of the vertical profile shape.

C 1.2*: P. 9676, l. 23-24: 'higher uncertainties' in comparison to the tropospheric column and also in comparison to the rest of the profile. Still this lowest point of the profile has the highest sensitivity. Please add explanation.*

AR 1.2: Near-surface concentrations typically have a higher uncertainty than tropospheric columns because they are more sensitive to errors in the profile retrieval. We agree with the reviewer that this might be counterintuitive, considering the fact that the sensitivity to the lowest layers of the atmosphere is highest. The cause of the higher uncertainty for near-surface concentrations (compared to columns) is twofold. Firstly because measurements at all elevation angles of one scan contain information about the vertical column. The measurements therefore have a considerable degree of redundancy with respect to this quantity. The sampling of the lowest profile layer is less redundant and dominated by the measurement at the lowest elevation angle. Secondly because of the low vertical resolution of the measurements, deviations from the true value in one layer of the retrieved profile can be compensated by another layer just above. This may lead to oscillations in the profile and therefore to higher uncertainties for individual layers. Such oscillations generally have a much smaller impact on the total column. One way to reduce the oscillations (and increase the precision of the retrieval) is to give more weight to the a-priori (by lowering its error estimate). Because the accuracy of a-priori data is generally quite low, this approach will often not improve the accuracy of the retrievals of near-surface concentrations.

The following is added to the text (OM, P. 9676, l.24):
"(primarily because of the limited vertical resolution of MAX-DOAS measurements)"

C 1.3*: P. 9677, l. 4-7: Please give a range/estimate for the degrees of freedom for signal usually obtained with MAX-DOAS.*

AR 1.3: the following was added to the manuscript (OM, P. 9677, l.7)
": the degrees of freedom for signal typically varies from 1 to 3, see Sect. 2.5"

C 1.4: *P. 9677, l. 27-28: Isn't also another reason for not being able to compare the profiles directly the dependency of the retrieved profiles on the input parameters such as the grid used for the retrievals due to the limited degrees of freedom? Please add some discussion.*

AR 1.4: The referee refers to the following text: "Since it is not straightforward to mutually compare a wide range of possible profile shapes, we compare profiles based on their 'characteristic profile heights' H_{75}, which we define in this work as the height below which 75% of the integrated profile resides (75% of the tropospheric column)."

An argumentation for this methodology is given in the subsequent lines of the manuscript:

"The advantage of this is that H_{75} is a scalar, and together with the tropospheric column and the near-surface concentration it gives a characterization of the profile that suits the typical first order profile approximation for constituents that mainly reside in the boundary layer."

We agree with the reviewer that it is somewhat difficult to compare retrievals performed on different grids. A pragmatic solution to deal with this issue would be to interpolate both profiles on a high resolution grid, and perform a layer-by-layer comparison on this grid. The reason to follow a different approach is that in this work we are not so much interested in the accuracy of both retrieval methods for every layer of such a high-resolution profile. This would be very interesting if the measurements had a very high vertical resolution, which is not the case. Knowing that the information content is low, it is thought to make more sense to compare profiles based on just a few well-selected quantities. We believe that this is explained sufficiently in the present version of the manuscript.

This paragraph was rephrased as follows:

"The two profile retrieval methods that are compared in this study do not retrieve profiles on the same vertical grid. One way to perform a profile comparison would be to interpolate profiles retrieved with both methods to a common vertical grid. A comparison performed in this way would give results for all layers which define the common grid. Such an approach would probably be favorable if the vertical resolution of the measurements (and therefore the DOFS) was high, but this is not the case for MAX-DOAS measurements, as noted above. Because of the low DOFS (1-3), it was decided to derive from each profile three suitable quantities and to compare the two profile retrieval methods based on those quantities: tropospheric column, near-surface concentration and "characteristic profile height" H75. The latter quantity is defined as the height below which 75% of the integrated profile resides (75% of the tropospheric column). An advantage of this quantity (a scalar) is that it allows a first order description of the profile shape of pollutants which reside primarily in the atmospheric boundary layer."

C 1.5: *Throughout the manuscript, 'aerosols' is used synonymously for 'aerosol extinction' and p. 9677, l. 20 even labels aerosol extinction as a 'species'. Please correct.*

AR 1.5: This has been corrected.

Further more general comments:
C 1.6: *P. 9682, 1st paragraph: How does the uncertainty calculated in this way compare to the smoothing error that can be obtained from the optimal estimation framework?*

AR 1.6: This comment refers to the following text: "The final product that is compared to method B is a composite of the retrievals with these three a priori: for each retrieval quantity (see Sect. 2.6) the mean of the values obtained with A1–A3 is taken as the solution, and the difference between the maximum and minimum as the uncertainty. The reason to follow this approach is that the impact of the a priori is substantial and there is no external information available instead which justifies the choice for one specific a priori."

After this interesting question by the referee, a numerical comparison (based on all retrievals performed in 2011) has been performed of the a-priori based error (described above) and the smoothing error. For the aerosol retrieval as well as the trace gas retrieval it was found that the relative a-priori based error is almost in every case larger than the relative smoothing error. In 80% of all aerosol retrievals, the relative a-priori based error is up to 4 times larger than the relative smoothing error (median ratio is 1.6). In 80% of all trace gas retrievals, the relative a-priori based error is up to 12 times larger than the relative smoothing error (median ratio is 4.8), see also Fig. 1 of this document.

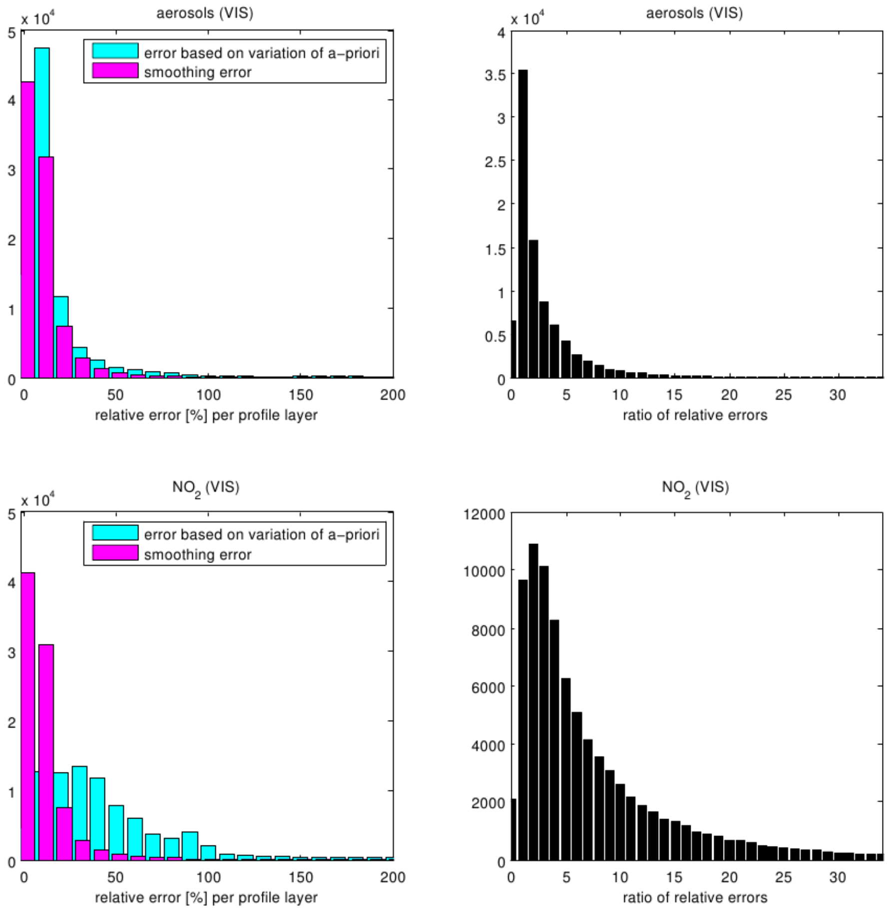

Fig. 1.: Comparison of two different error estimates: smoothing error and error based on variation of scale height of the a-priori profile. Both error estimates are expressed as relative errors. The upper row shows results (histograms) for the aerosol profile retrieval in the visible, the bottom row for the NO_2 profile retrieval. The left column shows the relative errors for both approaches, the right column shows the histogram of the relative size of the a-priori based error compared to the smoothing error.

The following has been added to the manuscript (OM, after P. 9682 l.6):

"Tests have indicated that errors estimated in this way are in general considerably larger than the smoothing error, a commonly used parameter in the optimal estimation framework to quantify the impact of the a-priori on the retrieval error (Rodgers, 2000). The relative smoothing error per layer is typically less than 20%, for both the aerosol extinction and the trace gas retrieval. The a-priori based error is about 1.6 times higher in case of the aerosol extinction retrieval and about 4.8 times higher in case of the NO_2 retrieval (both numbers are median values)."

C 1.7: *Section 3.2.1, all the numbers are stated in the table and don't need to be repeated* *here. It makes it quite difficult to follow this section. Maybe the authors should consider providing some of the information in the result sections in bullet point lists.*

AR 1.7:Where possible, numbers have been removed from this section. It was decided not to include bullet point lists, because an accessible overview of numerical results is provided by Table 3.

C 1.8: Technical comments:
MAX-DOAS, UV, Vis, VIS, BIRA, DAK, vmr, and the molecular abbreviations of the trace gases have not been introduced.
done

use 'slant column densities' instead of 'slant columns'
done

P. 9676, l. 9: 'gases' instead of 'gas'
done

P. 9676, l. 18: move bracket
done

P. 9676, l. 19: 'has been' instead of 'was'
done

P. 9677, l. 12-14: remove second 'of MAX-DOAS profiles'. Maybe use 'usability'/'usefulness' instead of 'use'.
done

P. 9679, l. 9: 'From 2010 until present, the instrument was/has been stationed. . .'
done

P. 9679, l. 22-23: 20 min for a scan? Please explain.
One scan of 9 elevations typically takes 15 minutes, see also Hendrick et al., ACP, 2014. In the design phase of the data-processing it was decided to interpolate data to a time resolution of 20 minutes, as described in the manuscript. Nothing has been changed in manuscript.

P. 9679, last 2 sentences: The causality is a bit off. Please rephrase
done

P. 9680, l. 4: Why 'forward simulations of differential slant column' (only)? What about weighting functions?
done

P. 9680, l. 8: ‘uppermost‘ instead of ‘most upper’
done

P. 9680, l. 16: ‘specific’ instead of ‘certain’
done

P. 9681, l. 14: ‘fraction’ instead of ‘part’
done

P. 9681, 2nd and 3rd paragraph: ‘decreases to zero’ instead of ‘goes to zero’
done

P. 9683, l. 25: ‘estimate’ instead of ‘estimates’
done

P. 9685, l. 22: remove ‘time’
done

P. 9686, l. 5: ‘retrievals’
not changed

P. 9686, l. 6, 7: ‘are’ instead of ‘is’
done

P. 9686, l. 13-14: Check sentence
done

P. 9688, l. 2-3: either remove ‘sometimes’ or ‘occasionally’
done

P. 9689, l. 4: ‘. . . Beijing. From. . .’
done

P. 9692, l. 12: ‘1.5 km is’: space missing
done

P. 9697, l. 4: ‘profiles grow higher than a priori’: maybe ‘extend higher’ instead. Also a bit further, ‘shrinking of profiles’, is rather an unusual expression.
done

P. 9697, l. 19: 'This study also makes clear. . .'
done

P. 9700, l. 29: 'values are higher' instead of 'above'
done

Table 3 caption: 'methods' instead of 'method'; 'abscissa' instead of 'x axis'
done

Figure 3 caption: remove first 'of'
done

Figure 7, 8 caption: 'the' instead of 'te'
done

Referee #2 (anonymous)

This is a nice and honest manuscript that I enjoyed reading. The manuscript is well written and the overall quality is good. Scientifically there are no problems and I see no reason why the manuscript should not be accepted. On the other hand I must say that in some technical matters the manuscript is still a bit clumsy. Nevertheless, the quality improved a lot after the first technical review. The clumsiness in technical matters makes the reading fatigue, which is a pity, since the manuscript is otherwise so well written. I suggest that the authors go through the manuscript once more and try to correct these little details. For example all the reported times should be double-checked. The authors should be systematic and follow the journal guidelines.

AR 2.1: Notation of reported times has been made consistent. Also it was noticed, because of this remark by the referee, that different symbols were used to denote the a priori a posteriori scale height (e.g. SH, $H_{scale}^{apost.}$, H_{scale}^{prior}). This notation has been simplified (to H_s^{prior} and H_s^{post}) and made consistent throughout the revised manuscript. Also the various ways to refer to the visible part of the electromagnetic spectrum (e.g. VIS / visible / Vis) have been made consistent (now only Vis). Altogether we think the manuscript has made another step forward in terms of 'technical matters', also because of the changes made after comments by referee #1.