Reply to comments from Referee #1

This paper by Donner et al. describes and evaluates four different methods for elevation angle calibration of MAX-DOAS instruments in the field. It evaluates the four methods using multiple MAX-DOAS instruments during the CINDI field campaign and concludes that all four methods are suitable for field calibration. Since there is a lack of papers documenting calibration of MAX-DOAS elevation angle, and this paper would be a first step towards standardizing elevation angle calibration for MAX-DOAS instruments, I suggest the paper be accepted with minor changes.

We would like to thank the referee for this positive assessment and several helpful comments.

Major comments:

The paper lacks description of laboratory calibration methods used before the field campaigns. It is mentioned that the instruments are calibrated in the lab before deployment on the field but no descriptions are provided. I think it would be very helpful to have some description of laboratory calibration methods and compare the pros and cons of the laboratory vs field calibration methods. What are the challenges of reproducing laboratory calibration methods? Is laboratory calibration better than field calibration? What is the accuracy and precision of laboratory calibrations?

Thanks for this helpful comment. We have asked all groups how they calibrated their instruments in the laboratory. 7 groups responded and a short overview on their results was added to section 3.1 “General approach”. Additionally, the following statement (P.2 L.32-P.3 L.2) in the introduction part of the paper:

“In the past, these calibrations were mainly done in laboratories where fixed target points were used as references and the elevations were calibrated accordingly. However, when the instruments were brought to the field, only rarely (if at all) the accuracy of the a-priori elevation angle calibration was checked under real measurement conditions.”

was changed to:

“In principle, these calibrations can be best done in the laboratory under stable and controlled conditions, where fixed target points are used as references and the corresponding elevations can be calibrated accordingly. In particular, the FOV should be determined already in the laboratory. Nevertheless, elevation calibration in the field is indispensable, because during transport from the laboratory to the field and during installation on the measurement site, it is likely that the instrument characteristics might change. In the past, however, when the instruments were brought to the field, only rarely (if at all) the accuracy of the a-priori elevation angle calibration was checked under real measurement conditions.

The paper concludes that all four methods are suitable for field calibration of MAXDOAS instrument even though the horizon scans could result in an offset of 0.3-0.6 degrees. In the introduction, it is mentioned that a 1 degree error in elevation angle at 0 degree elevation angle could result in ~20% error in NO_2 dSCD. Thus, recommending a method which could result in similar bias that the authors are trying to minimize does not seem right. I suggest the authors modify their conclusions to reflect this information provided in the introduction. I also suggest recommending one or two methods based on uncertainty of the method and ease of implementation.

Many thanks for this hint which indicates that our wording was not clear enough. The large spread (0.9°) of the results of the horizon scans is related to differences and errors of the initial laboratory

calibration (or changes during transport to the field). Therefore, they don't represent the accuracy of the horizon scan method. Note that the spread found for the horizon scans is consistent with the one obtained from the far lamp measurements. Moreover, the systematic difference (0.3° - 0.6°) between the horizon scans and the far lamp scans is easily explainable by the height difference of the lamp and the visible horizon. In summary, we conclude that also the horizon scan method is in principle quite accurate as long as favorable measurements conditions, like high visibility and non-rapidly changing clouds are selected.
We made these points clearer in the paper, especially in the conclusions.

The far lamp, near lamp and white stripe test all relies on knowing the optical axis of the instrument for accurate calibration. If the optical axis of the instrument is well known, could you calibrate the elevation angle using a bubble or digital level? Did any group level their instruments using such a level before the field campaign? I think this is a very important information as bubble or digital level is the simplest way to calibrate elevation angle. So, how does this method compare with the methods presented in the paper?

The reviewer is completely right that, if the optical axis is precisely known, a water level would be completely sufficient to calibrate the elevation angles. But this assumption is not always fulfilled, e.g. because the fibre bundle is not perfectly centered etc. Because of these uncertainties the in-field calibration of the elevation angles as described in this paper are indispensable. This information was added and more stressed in the "General approach" section (3.1) and conclusions section of the paper.

The paper simply glossed over the backlash issue. This is especially important for the 2D-MAX-DOAS instruments which are capable of doing elevation angle scans at any azimuth angles. The authors found a 0.4 degree difference between scanning from the bottom vs top and decided to just scan from the bottom. I think 0.4 degree is quite significant. How many of the instruments suffer from such a backlash issue? I suggest the authors include some comments/best practices to avoid such issues especially for a 2-D MAX-DOAS?

The reviewer is right when stating that 0.4° difference between scanning from below vs. from above is significant for the elevation of the telescope. For that reason we mentioned this issue in the paper. Nevertheless, we have asked all groups for their handling of this issue. The outcome was that depending on the kind of stepper/motor not all instruments suffered from such backlash issues, others actively corrected for this using inclinometers (e.g. LMU) or active sun trackers (e.g. BIRA). Most of the instruments which experienced backlash issues solved this issue by simply scanning always from the same direction (in elevation and azimuth direction) as mentioned in the paper. The effect of backlash (maximum difference between scanning directions) ranges from fractions of a degree to roughly 1°. While the effect is very important for the elevation pointing of the instruments, the effect has a smaller resulting influence on the measurements in the azimuth direction. The information provided here was also added to the paper in section 3.2.2.

Minor Comments:

P1, Line 5: This method was applied to more than 12 instruments as can be seen from Figure 19 - 21. Why did you not include all the instruments in the paper? I suggest make the number of instruments in the paper consistent.

Thanks for pointing this lack of clarity out. The horizon scans were performed by all instruments during the campaign which followed the standardised measurement protocol and reported them to the referee of the campaign (28 instruments). However, only 12 instruments (from 11 groups) participated in the far lamp measurements. We removed the number of instruments from the abstract (P.1 L.5). Further, this information was added to the paper in section 2.2.

P4, Line 27: How accurate are these inclinometers? I would think the motor steps are more accurate than the inclinometers.

According to the instrument's manufacturer the accuracy of the inclinometers used for the EnviMes instruments is 0.1° and the precision is 0.03°. Depending on the actual properties of the motors the inclinometers are at least as accurate as the motors if not better.

P5, Line 20: What was the resolution used for the initial calibration?

Thanks for pointing this lack of clarity out. The pre-calibration was done using a water level during the setup of the instrument. Then the finer adjustment was done using the results of the far lamp scans from 7th (in this night the lamp measurements were tested by our group with an scan resolution of 0.1° but the scanning was done manually), 8th and 10th September. The other two nights can then serve as tests of reproducibility. All values in this paper are given relative to the elevation calibration which was obtained by these finer adjustments and which was finally used for the campaign. This information was added to the text.

P7, Line 10: How is the light source aligned with the optical axis of the instrument using the laser level?

As sketched in Figure 7, the laser level illuminates both the instrument and the position of the lamp. The telescope and the lamp are centered around the position of the laser beam. The associated uncertainties of the relative vertical positions between instrument and lamp are estimated to about 0.1° (leveling accuracy of the laser level) and roughly 0.04° (thickness of the laser beam). This information was added to the revised text.

P7: Near lamp measurements: What are the pros and cons of near lamp measurements? What is the expected accuracy of near lamp measurements? I think the near lamp measurements need to be described in detail as it is likely easier to set up.

Thanks for this hint. We have revised the description of the near lamp measurements and added some more details and an improved error assessment (see also previous comment). Pros and cons are discussed in the conclusions section of the paper. Further, also the expected accuracy is given there.

P9, Line 5: How does the calculated FOV compare to the reported FOV in Table 2 for different instruments?

Thanks for this interesting question. Despite the fact that the determination of the FOV is not the main aim of the paper we added a new section "4.4 FOV determination" and a new figure (Fig. 21) to the paper, where a comparison between the retrieved FOVs (from horizon and lamp scans) and the reference FOV is provided. In general reasonable agreement is found. However, systematically slightly larger FOVs are found for the horizon scans than for the lamp scans. This can - at least for the 1D instruments - be explained by the fact that the far lamp was not always in the center of the azimuth dimension of the FOV. Also, the determination of the FOV seems to be less stable as compared to the determination of the target positions.

Section 4.1: I think the main message of this is lost amongst different type of instruments and different scanning modes. Please consider making this section concise. The main message of this *section is (1) independent laboratory calibration between different groups agree within 0.9 degrees, and (2) far lamp calibration method is stable. I don't think all the figures are needed to convey these points.*

We agree that this section has a lot of figures and descriptions associated with it. However, most of the figures are needed to illustrate the variety of different instruments and scanning schemes. In order to make this section clearer, we have combined Figures 12-15 and added subsections to give more structure to this section.

P14, Line 4-6: It seems that 0.9 degree spread is related to the initial laboratory calibration and not FOV? Why do you think this is related to the FOV?

Thanks for pointing this lack of clarity out. This is a misunderstanding. We did not mean that there is a relationship between the deviation of the elevation calibration and the FOV. We just wanted to express that the spread is of the same magnitude as a typical FOV and does actually matter. To avoid this misunderstanding, we removed this statement from the text.

P14, Line 28 – P15, Line 17: A lot of text to say we don't know what is going on. And it does feel like the author is rambling at times. Please be concise. May be it is better left for a separate paper.

We agree that this part is quite detailed and might be not completely suited for this paper. However, we want to keep the main message of this part. Therefore, we have removed Figure 23 from the paper. Further, the text was shortened in the revised version of the paper.

P15, Line 18: Change section title to "comparison between far lamp and horizon elevations" as there is already a comparison between methods section.

Thanks for this hint. We changed the section title accordingly.

P17, Line 13: How do you come up with 0.1 degree uncertainty? Based on the far lamp and near lamp measurements results for IUP-Hd, there is a bias of -0.3 degrees for the near lamp measurements?

The uncertainty of +-0.1° was estimated as described in the paper and is dominated by the errors introduced by the determination of the lamp position as described in section 3.2 and in the conclusions. This value is an estimate for the uncertainty of this method and not a difference/bias to another method. The near lamp scans in the UV and the VIS spectral range agree within 0.03°. The difference between the far lamp measurements (retrieved lamp position is roughly -0.4°) and the near lamp measurements is roughly 0.3° as mentioned by the reviewer. This, however, fits to the expectations, since the IUP-HD instrument should see the lamp below 0° at roughly. -0.2° as described in the paper.

P18, Line 2: How do you come up with 0.1 degree uncertainty? Is this an estimate of reproducibility error? What is the uncertainty of the Gaussian fit in figure 11?

As for the previous comment, the uncertainty of +-0.1° was estimated as described in the paper and is dominated by the errors introduced by the determination of the stripe position as described in section 3.5 and in the conclusions. This value was obtained by combining the errors of the Gaussian fit and the error of the determination of the stripe position. It is an estimate for the systematic uncertainty of this method. The reproducibility error, however, is a statistical error which is dominated by the motor precision as described in the new Table A2 and the conclusion section of the paper. Fit errors were added to all figures showing Gaussian fits.

Table 1: Please remove instrument ID. While it saves the authors from remaking Figure 21 and 22, it is very confusing to the reader. There are already too many acronyms and having a suffix that is not needed is not helpful.

We have removed the instrument ID from the table. Further, Tables 1 and 2 were merged to a new table 1. Additionally, the corresponding Figures were adjusted.

Table 3: Was the instrument calibrated using the far lamp before 08/09? Are these just test of reproducibility?

See answer above for minor comment P5, Line 20. All values in this Table are given relative to the elevation calibration which was finally used for the campaign. This information was added to the text in section 3.2.2.

Figure 3: Based on the lower part, the elevation angle of the lamp should be negative?

Yes, this is correct. The estimation of the lamp position (lamp elevation) relative to the instruments is mentioned and explained in the text serval times, e.g. P.6 L.22-30 and P.13 L.3-5. We added a hint to the caption of Figure 3.

Figure 5: Why are the measured intensity not symmetric? How does this asymmetry affect instrument elevation angle calibration?

If the fibre bundle would be exactly located in the focus of the lens and under ideal conditions, the image of the lamp would exactly pass through the center of the space between the fibres. However, all this conditions might be not completely fulfilled which leads to a more asymmetric intensity distribution. This is also mentioned in the paper P.5 L.32 – P.6. L7. We added a hint to the caption of the Figure.

Figure 7: How do you make sure the laser level is aligned with the optical axis of the instrument to calibrate the elevation angle?

See answer to the previous comments regarding the near lamp measurements.

Figure 9: Please mention what is the red dashed line in the figure?

Thanks for this hint. The red dashed line indicates the median center of the horizon scans. This information is given in text. To make this clearer we also added this information to the caption of the Figure.

Figure 11: This shows that it is the best method? Why is this not recommended exclusively?

Indeed, this figure suggests that this method is the best, since the presented fit is quite smooth. Nevertheless, the used setup was not optimal and especially the distance between stripe and telescope should be larger. Problems with the used setup are discussed in section 3.5.1. Further, pros and cons are discussed in the conclusion section of the paper concluding that the white stripe scans are a well-suited method if the setup is stable and optimal.

Figure 12: Panel c: I think it might be better to show average than sum? Same with Figure 13-15.

Since an average is just a normalised sum and the Gaussian fit will not be influenced by an additional normalisation, we decided to keep the sums in the plot. Therefore, we did not modify the figure.

Figure 12-15: I think it would be better to combine these into one figure. Also, all these figures are likely not needed.

Thanks for this suggestion. However, we believe that all these figures are important to give a good overview on the different characteristics of the different instruments and their FOVs, we decided to keep the figures. However, we combined them into one figure.

Figure 18: What is mean of fit errors? Instrument label is confusing?

The „mean of the fit errors" is the mean of all fit errors of the Gaussian fits which were applied to the intensity curves as explained in the text (section 4.1., especially P.12. L.28-30). As explained there, the fit error measures the quality of the Gaussian fits and the shape of the measured intensity curves.
Regarding the instrument labelling: the labels are according to Table 1. The numbers in brackets give the number of available lamp scans as indicated in the caption of the figure.

Figure 19: There are instruments that are not listed in Table 1. Make the markers little larger. How is the expected horizon calculated?

As already explained above (first minor comment), in total 28 instruments reported horizon scan data. However, Table 1 only lists instruments which explicitly participated in the far lamp measurements.
The marker size was increased. Additionally, the instrument labels were adjusted according to Table 1 (where instrument IDs were removed).
The estimation of the expected horizon elevations is explained on P.13. L.23-24. and P.10 L.27-30. We slightly revised the explanation of the estimation on P.13 and added more details.

Figure 20 and 21: Add the expected horizon on the plot. There are instruments in the plot that are not listed in Table 1. Either add them to Table 1 or remove from the figure.

Regarding the number of instruments which performed horizon scans, please see our previous answers regarding this topic. The expected horizon elevations were added to the revised versions of figures 20 and 21. Additionally, the instrument labels were adjusted according to Table 1 (where instrument IDs were removed).

Figure 24 and 25: Why are the error bars for different instruments so different? It seems like there were different number of measurements for different instruments. I think it would be more appropriate to include standard error of mean as the error bar.

The reviewer is right, since the error bars are indicating the standard deviations, the number of measurements plays a role here. However, also the performances of the individual instruments determine the actual sizes of the error bars. Nevertheless, we agree to your comment and replaced the standard deviations with the standard error to represent the error bars. The intercepts and slopes are almost unaffected as well as the relative sizes of the errors bars amongst each other for most of the instruments.