Reply to Second Review of "Local comparisons of tropospheric ozone: Vertical soundings at two neighbouring stations in Southern Bavaria" by Trickl, et al.

ThomasTrickl

September 14, 2023

General remarks

Given the substantial changes made during the revision I am astonished to read the lines of Reviewer 2. Nevertheless, I tried to make amendments in order to improve the clearness.

One again, I print the lines of the reports in italics and my replies in normal text.

Report 1 (July 31, 2023):

Some minor corrections

1) line 30. Correct to "...for low to moderate ozone concentrations"

Corrected.

2) line 543: '... differentiating the backscatter signals.". Please cite a relevant publication.

I found this phrase in line 515: I cited our 2020 lidar paper.

3) line 575: "... Scheel for 3 km". please provide a relevant citation.

I already wrote "see Introduction", which refers to the statement "personal communication around 2010".

Report 2 (July 20, 2013):

Summary:

The revised version of the manuscript accepted many of suggestions for improvement. Thank you. However, my main major comment was largely ignored and therefore, the paper remains slightly confusing and in some places misleading. I do object to adjusting ozone sonde profiles to lidar data and then using these adjusted profiles to evaluate the quality of the lidar. By adjusting the ozone profiles to the lidar, there is already the implicit assumption that the lidar is a suitable reference. Once this assumption has been made, the result cannot be used to prove that the lidar is a suitable reference.

I am astonished to read that we adjust the MOHp sonde data to the lidar data implicitly assuming that the lidar is a suitable reference. Indeed, the lidar is a suitable reference as demonstrated in Sec. 3.1! We made clear statements about this, e.g. in lines 314-316 in the reviewed version. Nevertheless, thank you for pointing this out! I tried to clarify this more explicitly.

In the "Results" section we first describe the highly successful comparison with the FZJ ECC sondes that, together with the UFS data prove the excellent performance of the lidar. The agreement with the mountain stations has been routinely verified over many years (as mentioned several times, starting with the Introduction), which is confirmed and statistically evaluated in the current study. There is an indication that the lidar calibration could be free of bias. The unique side-by-side comparison of lidar and ECC sondes plus the station data yield just minor concentration offsets of these sondes, but an impressive agreement in vertical structure.

Based on this successful validation of the lidar we start the comparison with the MOHp sonde. Despite the distance of 38 km between both sites we find structural agreement between the soundings, but also offsets that change from sonde to sonde. It is hard to believe that these offsets are caused by accidentally vertically agreeing atmospheric differences. This is now discussed in the introduction to Sec. 3.2. We evaluate both the agreement of the lidar with the mountain sites and the offsets statistically. Since the agreement of lidar and in-situ ozone is convincing we conclude that the offsets are a sonde-specific issue. Only after this conclusion we try to evaluate an upper limit of the lidar uncertainty based on the retrieved structural differences.

This does not mean that I consider any of the instruments of deficient quality. Each of the measurements systems has their strengths and weaknesses. Evaluating the quality of each means describing their differences (biases or offsets) relative to each other (without adjusting one or the other). The attribution of the offsets should be done based on additional information, for example the surface sites, but other information may be suitable as well.

I am really astonished to read this sentence since this is what was actually done (e.g., lines 387-396 of the old manuscript)! As mentioned above, I added some information including the routine comparisons with the summit sites and UFS over many years of ozone sounding with the DIAL. Many examples have been shown in our earlier publications. In this manuscript we even present a statistical analysis for all the three years of comparison.

The revision still does not clarify the hierarchy of references. This is a source of the confusion. While the authors may have a clear picture, which instruments they consider the most reliable as reference for which purpose, this does get lost in the order data are being presented and discussed.

As pointed out above I do not agree. However, although this hierarchy is clearly visible I add more explicit statements.

I consider this manuscript an important paper for the ozone monitoring community. However, the analysis and presentation of the material confuses issues more than it helps elucidate the current state of observing capabilities. I would strongly urge the authors to take another look. In its current stage, the manuscript still requires major revisions.

Detailed comments:

Lines 283 ff: I am not sure, whether you have understood my comment during the initial review: It is not appropriate to adjust data and then call an agreement outstanding. Either data have to be adjusted to reach agreement (i.e. there is no agreement), or the agreement (without adjustment) is outstanding. You cannot have it both ways!

It is very difficult indeed to follow this argument. The outstanding agreement is not claimed for the unshifted data. I revised that part.

ECC ozone sondes are generally considered absolute instruments, i.e. they do not require calibration (against a known ozone reference). This implies that the term "uncalibrated" is not applicable here. Rather than showing an agreement that was forced, the magnitude and the character of the differences need to be described. This may be important information for the ECC community.

Thank you for this remark. As can be found in Sec. 2.3 the ECC sondes were prepared as described and (as I learnt) no ground calibration was done. This is what "uncalibrated" means. However, the entire paragraph has changed and "uncalibrated" no longer exists.

Lines 307 ff: Same as comment on Lines 283ff. In addition, which instrument is evaluated against which. The start of this paragraph reads: "For quantifying the quality of the lidar measurements

…". Next you proceed to correct the ECC ozone sonde profiles based on the lidar comparison. Then, they describe the small offsets between the two systems. Lastly, in that paragraph they conclude: "This result justifies to use the lidar as a quality standard in the comparisons with the MOHp Brewer-Mast sondes described in the following sections." It is not appropriate to adjust one instrument to another and then use the resulting agreement as quality justification in the comparison with a third instrument. I do agree that the lidar is a good instrument. I disagree with the logic of the argument.

Thank you! I tried to describe the hierarchy clearer than before. The lidar agrees better with UFS than the ECC sondes. However, I think that offsets of 1.5 ppb ± 1 ppb are not a bad result for the ECC sondes. As to the comparisons with MOHp I see reasonable agreement after removing the offsets. It is an important question why this approach looks rather suitable. Given the amount of material analysed I cannot believe in accidentally constant atmospheric offsets. Of course an altitude-dependent atmospheric component is present and cannot be removed. As mentioned above, this is now more explicitly discussed.

Lines 337f: "After adding 5.8 ppb the sonde results (cyan curve) match the lidar values well for altitudes above 2.1 km." This is the same issues as above, i.e. that you cannot have it both ways. However, here it is easier to correct. You could simply write: "The Brewer Mast ozone sondes show a low bias of 5.8 ppbv relative to the lidar above 2.1 km."

Accepted.

Lines 354ff: In the discussion of the summer profiles, this is same issue not as easily corrected. Most importantly, it is no longer possible to evaluate the uncertainty of the lidar measurements (lines 349f).

This is obvious because of the atmospheric component. Thus, the analysis can give just an upper limit.

Lines 428: Same issue as above and something I pointed out in the initial review. To be clear what I mean: The original ozone sonde data show a bias relative to the lidar. You remove this offset from the sonde data. And then you state that "The analyses for 2018 do not reveal a significant bias …". Quite to the contrary, your analysis did find a bias and you removed it. Therefore, your statement is misleading. Unfortunately, the source of this confusion is your basic approach to remove the bias from the sonde measurements rather than just sticking to describing it.

What means "describing"? We find that the approach of a constant bias rather suitable since it is verified in so many examples. Of course, the explanation of the bias is an interesting topic for future research. In any case, I modified this sentence.

Figures 11 (formerly Figs 11-13): This Figure relates to the fundamental presentation challenge that I tried to raise in my initial review and above. It is difficult to evaluate differences, if they have been removed in an earlier stage. What do the remaining differences tell the reader? This is unfortunately misleading and has not been addressed.

It is reasonable to assume that the Brewer-Mast sonde reproduces the vertical structure of tropospheric ozone, apart from the bias. This is verified in most of the comparisons, although local differences exist at some altitudes. We clearly point out that the uncertainties are overestimated due to the atmospheric issues.

Line 33: You did not address my initial question. What does the uncertainty, the value after the +- refer to? You now use the term "maximum of deviations". In the text, you actually specify the *standard deviation without specifying the "maximum of deviations". It would help the manuscript to make these two identical.*

We now take the standard deviation throughout the Abstract since this is the quantity derived in the analyses.

Line 134: Your answer to my initial comment is interesting. Please include this brief discussion in the manuscript at the appropriate location.

Information on sonde comparisons is given in the Introduction. One sentence was added to the Discussion.

Lines 446ff: I don't believe you understood what I meant in my initial comment. Although the distance between the stations is fixed, ozone gradients in the atmosphere are not. If winds are blowing along the line connecting both stations, the spatial separation is probably negligible. If winds are blowing orthogonal to that line, then the spatial separation is quite important, despite their relative proximity.

I fully agree. We discuss here just two examples with substantial discrepancies. However, I have problems in assuming that an atmospheric structure extends over the entire troposphere to explain the constant bias. I added a sentence on this.

Technical comments:

Line 26: Delete "just"

Removed

Line 133: Change to "(corresponding to more than 2.5 ppb)"

Changed.

Line 175: Delete "just"

Deleted.

Line 263f: Change to "(indicated by low relative humidity)"

Changed.

Line 266: Delete "must be assumed"

Deleted.

Line 425: "severest" -> "most severe"

Changed!

Report 3 (July 31, 2023)

General remarks:

The content of the manuscript has already been well described by the previous three referees who had reviewed the first version of the manuscript, such that I will constrain my review mostly to the major and minor revisions demanded by referee #2. Regarding the minor revisions, the authors have responded and revised the manuscript appropriately, however, the new manuscript still lacks in the way the comparisons between lidar and ozonesondes were made. The major critics of referee#2 was that in multiple figures and descriptions in the text bias corrections to the ozonesonde data were applied without given the reason/cause for that. Therefore, it would be more interesting to properly describe the bias and show profiles of the actual differences. I fully share these demands of referee#2. Unfortunately, in the revised manuscript the authors did not revise the *text and figures adequately, neither replied in a satisfying way to these major critics made by referee#2, which I fully underline. In the revised manuscript the authors still follow their original methodology to do unexplained bias corrections to the ozone data. When doing such corrections then solid arguments have to be given, which are still missing. At present, the reader easily gets the impression that the bias corrections are just artificial "corrections" to adjust the comparisons to get a better agreement of the Lidar with the sondes, however, this would finally not improve the trust in the Lidar data, but just do the opposite. The last is certainly not in the authors their own interests, and therefore I strongly recommend to going for a second revision of the manuscript, but now follow the major comments given by referee#2 more strictly.*

In tried to describe the hierarchy of the instruments in more detail.

The offset corrections bring both instruments in substantially better agreement. Since the lidar agrees well with the ECC sonde and the station data we can assume that the lidar produces highly accurate results, at least under conditions up to moderate ozone. The agreement with the station data persists throughout the period under investigation. Thus, in this altitude range we are sure about what we are doing. Our concern has been the range above 6 km, particularly in summer. Here, our results suggest a reasonable agreement, apart from some issues in 2009.

Of course, we do not ultimately know if there are systematic differences in air-mass composition between the lidar an MOHp. However, since the agreement of the profiles in structure is so good on average it is hard to believe that such a difference exists. We cannot fully remove the atmospheric variability. Thus, out statistics yield just an upper limit of the uncertainty of the lidar.