Interactive comment on "Comparison of Aircraft Measurements during GoAmazon2014/5 and ACRIDICON-CHUVA" by Fan Mei et al.

Anonymous Referee #5 Received and published: 26 August 2019

Mei et al. provide a comparison of datasets from two research aircraft obtained during a coordinated comparison effort. Comparisons between calibrated instruments are quite useful for evaluating whether the estimated uncertainties for each instrument do accurately represent the data quality, which is of course paramount to the usefulness of the data. Ideally, analysis of such comparisons could be used to better understand estimated uncertainties and possibly reduce those uncertainties. This paper takes on a significant effort because the authors compare all of the possible measured parameters between these two aircraft (> 10 parameters). In general, I feel that the paper would be more useful if the scope were somewhat smaller with more significant analysis and discussion of the differences between a subset of the measurements. There are a few useful recommendations for measurements going forward, but also some of the disagreements between measurements which might be considered significant are not explored enough to understand if the measurements can be reconciled. At the same time, I don't think it is reasonable to ask the authors to change what they see as the purpose of the paper, but suggest that in the future such comparisons may better serve the community by going more in-depth on a smaller group of the measurements.

We thank the reviewer for the thoughtful comments and suggestions. We also agree that the community may benefit from in-depth comparison of a smaller group of the measurements in the future. Our responses to the specific comments are described below.

I have some suggestions and changes that I would like to see the authors address. These are listed below.

Table 3: Recommend instead of highlighting only slope and R2 that the systematic differences in measurements are calculated (two measurements could be perfectly correlated with a slope of 1 yet have a huge offset and differ on average by a large fraction). Can you also include something about the expected agreement based on the uncertainties of each instrument?

Response: The fitting slope and R2 in Table 3 assume the correlation between two measurements: the G1 measurement is equal to HALO measurement. Thus, the listed slope is from y=slope* x equation without the offset, and the R2 is based on the equation 1. We modified the line 336 "assuming that two measurements from the G1 and HALO have the 1:1 relationship." We also tried the orthogonal regression to relate the measurements from the G1 and HALO, which shows the similar results as Table 3. We added another table (new Table 4) to further illustrate the uncertainties of each instrument.

Line 365: Was the fact that the G1 sensor data point bad here known before the comparison and would it have been thrown out? If so, recommend removing this point from the figure as it does represent what is thought to be good data.

Response: We replotted Figure 4(c) without the bad data points, then changed the sentence (line 366). The initial data quality control did not exclude the G1 sensor data as questionable data i.e. chilled mirror sensor wetted by cloud droplets.

Section 3.2: Ozone: Table 3 shows a minimum ozone value of 0.5 for G1. Is this correct or a typo? It seems there is a slope and offset between the ozone instruments. Difference between the means is about 17%, which I think exceeds what is expected (~ 5% each instrument). I doubt the explanation that sampling losses in the tubing could account for the difference as O3 is not too difficult to sample. Please state clearly whether the differences observed between the O3 instruments exceeds what is expected for the sensors themselves, and what evidence there is to suggest sampling loss is to blame. Possibly, a leak of cabin air into the sample line affected one of the instruments.

Response: We agree that the sampling loss is not the main reason causing the measurements difference. We have edited the lines 386-390 "As mentioned in section 2.1.2, each instrument has a 2 ppb accuracy (or 5%) on the ground based on a direct photometric measurement measuring the ratio between a sample and ozone free cell. The in-flight calibration showed that the variation of each instrument could raise to 5-7% (or 2-3.5 ppb). Thus, the difference between the averaged ozone concentrations – 4.1 ppb is close to the instrument uncertainty."

CO: Recommend removing the outlier CO point if you have good reason to believe it was not coincident. At the same time, I don't see how the explanation on 389-391 about "different operation principles" has anything to do with lack of coincidence between the measurements. Please clarify if the disagreement is because of bad coincidence or if you think the instruments really do not measure the same thing.

Response: We removed the outlier CO points from the altitude between 2000-3000 m, which we believe the G1 and HALO are sampling different air mass. We agree that the "different operation principles" should not cause the significant difference in the measurements. We modified the manuscript lines 397-419.

Line 418: Kind of weak discussion here about CPC difference. Seems like HALO is systematically lower. It would be useful to understand something about the difference rather than just state that it can be attributed to the typical uncertainties and other unknown factors. The comparison between UHSAS does not support it being an issue with the isokinetic inlets.

Response: We included further discussion of the CPC difference in between line 443 – 457 with additional figures in Figure 6.

Figure 9: Why does HALO UHSAS look so much noisier?

Response: The airborne version of the UHSAS does have an issue with maintaining constant volumetric sheath flows (discussed by Kupc et al. 2017), which directly affects concentration since the sample flow is not directly measured but calculated as the difference between total and sheath flows.

Line 488 / Fig. 11: I don't see the value of this comparison. It is stated in the text that the UHSAS < 50 nm is not a measurement, but rather an extrapolation of the distribution down to sizes the UHSAS cannot measure, and that this extrapolation could easily be invalid during e.g. a nucleation event. Therefore, I don't understand when the extrapolated UHSAS data would ever be of use for scientific analysis. The fact that the extrapolated UHSAS distribution deviates from the FIMS measurements sometimes does not even require the UHSAS instrument to determine this. One could just extrapolate the FIMS data using the UHSAS sensitivity range and look at the difference between the FIMS measurements.

Response: We moved this section to the supplemental material. The original objective was to emphasize the importance of expanding size distribution measurements to below the 50 nm range on an airborne platform using advanced instrumentation (e. g. FIMS). We learned from many modelers that they typically use extrapolations of the UHSAS size distribution in their models due to scarcity of real data. Thus, we compared the real measurement from FIMS to the UHSAS based extrapolation.

Line 512: What is referred to here had been done for decades on other aircraft and has been referred to as NMASS. Recommend citing the relevant papers for that here and earlier in the paper (i.e. lines 478 – 487). Most recently: Williamson et al., AMT 11, 3491-3509, 2018.

Response: Thank you very much for your suggestions. The section is modified in the supplemental material.

Line 519/section 3.3.4: There is no actual discussion of the chemical composition, just the mass/volume. Recommend removing the reference to chemical composition here and earlier in the paper (e.g. abstract and introduction).

Response: More discussion of the chemical composition was added to the manuscript in section 3.3.3, pages 19-20.

L 650: How about a calculation with TUV to test whether the different sensitivity ranges can account for the 10%? ———-

Response: We used the tropospheric ultraviolet and visible (TUV) radiation model from NCAR website (https://www2.acom.ucar.edu/modeling/tropospheric-ultraviolet-and-visible-tuv-radiation-model) and estimated the weighted irradiance at 15:42:00 on Sep 9 2014. Note that the modeling output is limited to the range between 315 to 900 nm. It is different from the irradiance spectral range (400-2700 nm) in the G1 or the 300-1800 nm from HALO. The difference between the two aircraft measurements was 24.1 W/m^2 at that time, and the modeling suggested the irradiance difference between 315-400 nm was 13 W/m^2. Although we can't estimate the difference between 1800-2700 nm with TUV. We have shown that the difference in spectral range of the instruments is the main contribution to the difference in the comparison.

Editorial type notes: Line 101: issures -> issues

Response: corrected.

Line 146: change comma to period

Response: corrected.

Line 304: ‘paten’ -> ‘pattern’?

Response: corrected.

Line 326: ‘Tables’ -> ‘Table’

Response: corrected.

Line 418: ‘ rest of the 10-15. . .”

Response: corrected.