RC2: 'referee report', Anonymous Referee #4

Experimental studies of the atmosphere using aircraft are extremely important and multi-aircraft experiments are often performed to expand the range of measurements or the spatial or temporal scales. Sometimes comparisons between these different aircraft platforms are performed and are often very instructive for those involved since they improve the measurements and identify any issues with the data processing or instruments. However, they are rarely published. It is therefore good to see that the authors are trying to provide this for the major study carried out above the Amazon region. These exercises are often very important and allow the data sets from both aircraft to be combined and integrated together. This is very useful and the paper achieves this aim by providing statistical comparisons between the platforms. In this sense it provides a useful contribution to the ACRIDICON-CHIVA experiment.

Response: We thank the reviewer for all the valuable comments and appreciate the suggestions the reviewer made. Below, we have added our responses to the comments submitted.

Unfortunately, it does little more and this is an opportunity missed. It would have been very good to see a more insightful discussion of the instrument performance, pre and post flight calibration details and what happens if these are not carried out. What, if any ground comparisons were carried out and how useful these were to the overall performance of the instruments? Were data analysis approaches compared and what did these yield? A more detailed discussion of these topics would provide some real insight and information for others carrying out similar work, whether in a single aircraft project or when multiple aircraft are being used together. I would strongly recommend that this is carried out in a revised manuscript and that some of the sections are removed such as the extrapolation of the size distribution to smaller sizes and the radiation sections.

Response: Thank you very much for your suggestions. We changed the manuscript to provide more insights about the instruments' performance: 1) we included more details in the aerosol chemical composition comparison section 3.3.3 and added two figures (Fig 11 and 12); 2) we expanded the discussion in the trace gas and aerosol number concentration section 3.2 ; 3) we modified cloud probe comparison section to include details on the data processing (Line 587-596

and Line 628-636); 4) we ameliorated the comparison results with an additional linear regression approach (Table 4). More details are provided in the responses to specific comments.

Comments Line 193-207: Given that there are some discrepancies in the AMS measurements it would be very useful to have more information on the inlets and sample tubing for the two instruments, particularly the pressure controlled inlet systems. Were the instruments calibrated before and after each flight or if not when were the calibrations performed? Were all the instrument parameters (ionization efficiency/air beam, flow rates etc) varying in a consistent way throughout the experiment? How was the CE determined?

Response: The constant pressure inlets used by both G1 and HALO AMS were very similar to the design by Bahreini et al. 2008 (the reference had been included in section 2.1.3). Both AMS instruments were calibrated before and after the field deployment and also once a week during the field campaign (line 210-211). More details about the AMS measurements are given in separate AMS papers from the respective groups (Schulz et al., 2018; Shilling et al., 2018). A short summary is now included in the supplemental material. For example, the CE of the G1 AMS was determined by comparing AMS data to UHSAS and FIMS data. We confirmed the CE=0.5 by comparing mass loadings observed at the T3 site to the G1 data. For HALO AMS, CE of 0.5 was applied, as recommended by Middlebrook et al. (2012) for low nitrate conditions.

Line 414-420: No comment is made about the two sets of points at the start of the comparison which show enhancements in aerosol number in both aircraft at separate times, presumably one shortly after another. This gives rise to an increase in the uncertainty statistics but not the regression since the values are relatively low. It might also be good to discuss the breadth of points in the CPC regression since it could almost be argued that the pairs of points fall around two different regression lines.

Response: We included further discussion of the CPC difference in lines 443 – 457 with additional plots in Figure 6.

Lines 434-439: If you can demonstrate that the aerosol sources are systematically different in the two profiles from the G1 and HALO then I don't see any justification for including the plot in the

paper since there is no information to be gained. I suggest a clearer and more detailed explanation of why the aerosol sources in the two measured profiles are different and then a statement stating that this is the reason for not including the comparison, or if this cannot be satisfactorily demonstrated the statement of causality should be removed.

Response: Based on the aerosol number concentration, chemical composition, and CO concentration data, we believe that the G1 and HALO were sampling different air masses at altitudes between 2000 and 3000 m. Thus, we excluded these data points on replotted figure 8 and revised the corresponding discussion in section 3.3.1.

Line 471: I would recommend the removal of section 3.3.3. This is already a long paper and contains considerable amounts of detailed information. This section doesn't really show any comparison as such, it simply says that extrapolating a particle number size distribution below 100 nm based on optical particle size distribution information alone will underestimate the particle concentration if there is a small aerosol mode. In deep convection such particles can be activated and so extrapolations are to be treated with caution in environments where this occurs. A comment to this effect in the previous section is important as a caution but reducing the text would certainly help also.

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

Line 519 and following: Despite Section 3.3.4 being titled aerosol composition there is no comment about the chemical composition only a focus on the transmission of one of the AMS inlets. The implication from what is written is that the aerosol is predominately organic. Some discussion of the composition and any difference between the two instruments discussed. This is particularly the case if the inorganic components are above the detection limit since one could

then test the effectiveness of the ion balance to derive ammonium concentrations. It would also be good to include some comment on the Collection Efficiency that is used and how this was calculated.

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

The quality of the English, particularly through the cloud section is rather poor. This needs to be significantly improved before publication.

The revised manuscript has been reviewed by many native speaking co-authors and the cloud section is edited by a professional editor.

Lines 618-619: It is always difficult to compare cloud probes between aircraft due to the spatial and temporal distances between the two aircraft. Nevertheless, this section does fall short of any detailed insight at all. It is stated that "The difference between the G1 CDP and FCDP may be due to the data post-processing". The implication is that this wasn't checked out in detail. There is clearly no information here that can be used by a reader that would be remotely useful. I suggest that much more detailed analysis is provided for this to be useful. Why weren't the corrections for coincidence and shattering applied in a consistent manner?

Response: Thank you very much for your suggestions. We have modified the text accordingly. "The difference between the G1 CDP and FCDP is mainly due to the data post-processing. The G1 CDP used an earlier version of the data acquisition system from Science Engineering Associates, with limited capability to store the particle-by-particle (PBP) data for further processing. The CDP had an 800-μm-diameter pinhole placed in front of the sizing detector to minimize the coincidence up to 1850 cm^{-3}. On the other hand, FCDP was equipped with new electronics and PBP data was locally stored on a flash drive onboard the Linux machine. For the G1 flights, a constant probe-dependent adjustment factor was applied to FCDP to adjust the coincidence further. The G1 CDP and FCDP operated with redesigned probe tips to minimize the shattering effect. An additional algorithm was applied to the FCDP data to eliminate particles with short interarrival times."

Line 632 and following: This section says almost nothing at all and could be removed.

Response: This radiation comparison section is mainly to illustrate the challenges of comparing two radiation instruments deployed on two aircraft, including many factors which affect the accuracy of the measurements. And we also confirmed the effects of the difference in spectral sensitivity of the radiometers using the NCAR tropospheric ultraviolet and visible (TUV) radiation model.

Line 652: Uncertainty Assessment: This section is extremely qualitative and non specific. As written it serves very little purpose. Instead I would recommend a much more detailed examination of uncertainties embedded with each of the sections and for this to be made more quantitative.

Response: We have modified the uncertainty assessment section contents (page 23, section 4) and revised Table 4 to be quantitative. The information about the sources for the discrepancy between the two measurements can be useful for users to understand data uncertainty and for future field campaign planning.

Minor comments: (I stopped writing the minor corrections after a while since the latter part of the paper needs a significant revamp if it is to remain).

Line 85: uncertainty ranges

Response: corrected

Line 101: issues

Response: corrected

Line 1100: delta

Response: corrected

Line 155: section 2.1.3: were the CPCs from the G1 and the HALO run side by side on the ground for a period? If so it would be good to report this. When were the instruments calibrated relative to the field experiment? This isn't said explicitly.

Response: Unfortunately, we never got a chance to run CPCs from the G1 and HALO side by side on the ground. The aircraft were parked far apart, and such a comparison would have required un-mounting and relocation of one CPC, which was not practical during tightly scheduled field campaign. All CPCs were calibrated before and after the field campaign and checked at least once a week during the deployment.

Line 188-189: which have a refractive index

Response: corrected

Line 178-192: when were the UHSAS instruments calibrated relative to the flight periods?

Response: The UHSAS was calibrated before and after the field deployment and checked with PSL's once a week during the deployment.

Line 227: should read in present tense "are discussed"

Response: corrected

Line 234: needs to be rewritten "working independently and electronics produce shadowgraph"

Response: changed to "working independently. The 2DS electronics produce shadowgraph…"

Line 263: not sure about the use of the word "proven"

Response: changed to "examined".

Lines 247-268: How were the sample volumes of the HALO probes established? This is stated for the G1 but not HALO.

Response: line 257-258, "The sample area of the CCP- CDP was determined to be 0.27±0.025 mm^2 with an uncertainty of less than 10% (Klingebiel et al., 2015)."

Lines 276-278: It is not clear how this is actually achieved.

Response: Dr. Long have provided more details about the tilt correction in his paper (Long et la., 2010). We modified the sentence to make it clear, "Additionally, the angular offset between the

actual orientation of each radiometer's detector and the level position from the navigation data has also been determined and corrected after the installation for each deployment."

Line 304: stacked pattern

Response: corrected.

Lines 308-309: "Due to the different aircraft speeds, the flight distance between two aircraft flight paths continued increasing from 15 min to 1 hour" I do not dispute that the distances between the flight paths continued increasing but since the G1 took off first and the HALO is the faster aircraft I cannot see how the increase in time between the aircraft is due to the different aircraft speeds.

Response: We have revised the sentence: "Due to the different aircraft speeds, the time difference between two aircraft visiting the same part of the flight path varied, increasing up to 1 hour at the end of the path, as shown in Figure 3."...

Line 321: present

Response: corrected.

Line 323: intervals

Response: corrected

Line 337: "The linear regression achieved a slope was near 1" should be "The linear regression achieved a slope that was near 1"

Response: corrected

Line 340-342: This is a good way of presenting the uncertainty though I am surprised that you didn't use the orthogonal distance that would also represent the variability in x.

Response: Thank you very much for your good suggestions. We have included the orthogonal regression in new Table 4.

Line 356: when (the) G1 flew

Response: corrected

Line 418: change “rest” for “remaining”

Response: changed to “rest of the 10-15%”

Line 387-393: I am unsure why the regression statistics are presented including the point with high CO measured by the G1 but not by the HALO in Fig 5b. By all means present the data point but it does seem a little strange to include it in the reporting of the agreement.

Response: We replotted the CO comparison and modified the discussion in lines 397-419.

Line 439: sources

Response: corrected.

Line 470: “has a reduced spatial resolution”

Response: corrected.

Line 718: strophic?

Response: changed to “lower troposphere”