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.
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.

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?

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.

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 includ-
ing the comparison, or if this cannot be satisfactorily demonstrated the statement of causality should be removed.

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.

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.

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

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?

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

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.

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

Line 101: issues

Line 1100: delta

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.

Line 188-189: which have a refractive index

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

Line 227: should read in present tense “are discussed”

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

Line 263: not sure about the use of the word “proven”

Lines 247-268: How were the sample volumes of the HALO probes established? This is stated for the G1 but not HALO.
Lines 276-278: It is not clear how this is actually achieved.

Line 304: stacked pattern

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.

Line 321: present

Line 323: intervals

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

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.

Line 356: when (the) G1 flew

Line 418: change “rest” for “remaining”

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.

Line 439: sources

Line 470: “has a reduced spatial resolution”

Line 718: strophic?