My previous review of this manuscript argued that while the technical improvements of the CARIBIC-AMS described therein are very impressive, there was not enough detail provided to actually show that this instrument has promise as a quantitative, autonomous PM sensor in the UT/LS. The revised manuscript has addressed adequately most of these shortcomings but still falls a bit short on the overall quantification front. The authors make a good case that in a lot of ways this is still a case of "COVID-19 hangover", complicated by the general challenges of the CARIBIC platform. Given the overall consistency of the data presented, I am reasonably convinced that the instrument got UT/LS PM right when it was working within a factor of 3 or so if one accounts for the CE, transmission and other uncertainties, which for what the paper aims to show is sufficient. Hence, I recommend publication once a few last comments listed below are addressed. Hopefully after some additional characterizations the next paper will show significant improvements on these fronts, so that the CARIBIC-AMS can truly be a benchmark for CTMs for years to come.

These comments listed below include some responses/clarifications to previously raised points in the discussion which are not germane to the actual review of the revised text, so for simplicity I am going to highlight the actual action items in *italics*:

- *While in the main text it is made clear that the DLs are typical averages under flight conditions, this is not made explicit in the abstract. Also, neither the text nor the abstract make clear that these are CE=1 DLs (I would assume), and that they might change for different ambient conditions. I think both of these points should be added since there is in general a lot of confusion on AMS DLs in general in our community.*
- I would maintain that instrument blanks are the most accurate way to determine/validate detection limits, provided they are taken under realistic conditions. A 2 hour long blank, as described in Drewnick et al (2009), is certainly not realistic, since as shown in that paper the ionizer basically cleans out after 30 min or so, but that does only partially apply to short blanks like the ones presented by the authors. *So I do think that the figure presented in the response showing agreement between the background variability method and the blanks is valuable, and I would ask the authors to consider including it in the paper.*
  Also, to clarify, my previous request for "stratospheric" data was a typo, I just meant high altitude data, I apologize for the sloppy wording.
- *I also think the DL discussion would read better if "noise of the background" would be replaced with "short-term variability of the background". While for high m/z it is indeed electronic/counting noise, this is e.g. certainly not true for ions such as SO^+ or CO_2^+, which have a high, highly variable, but not necessarily noisy background.*
- *Regarding the presentation of the instrument operation, I want to clarify that what I specifically meant in my previous review was a block diagram showing the control boxes/computers and the connected powerboxes. And then ideally (in a different color/style) which piece of software runs on which computer and what the call hierarchy*

is. I do think this would help the presentation, which is hard to follow at the moment, especially for e.g. a second year student. Such a diagram would also highlight how certain design choices do improve the overall reliability/fail-safe behaviour of the system, which is a bonus for a paper like this.

- Regarding ePToF, a couple of clarifications:
    - The time offset for ePToF and PToF is slightly different (see Williams et al, 2016)). Given the different triggering electronics used by the authors it might not be consistent with the Aerodyne findings, but this could be mentioned.
    - While the ePToF dutycycle is indeed higher than PToF, both the spread of the signal over 25-50 bins and the noise in the inversion at low signal levels does result in quite high DLs vs MS mode, so I would not be surprised if regular ePToF operations is ultimately not implemented. Best of luck in any case.
- The inclusion of the inlet loss calculations is very much appreciated, although as the authors state themselves the main uncertainty here is in the exact shape of the AMS lens transmission. Which could be included in Figure 4d for reference (potentially along with some others, e.g. the Molleker 2020 curve in case a different AMS inlet was flown).
- I do understand that in-field calibrations are not an option given the CARIBIC way of doing things. *What I was asking for (possibly in a very confusing way) in my previous comments is for the in-between deployment calibration data. It does sound that instrument optimization efforts resulted in very minimal coverage on that front. Still, if there were e.g. calibrations available every 6 months over the 2018-2020 period, that would I think still be more illuminating than the single point in time comparison provided.*
- Regarding the SI discussion, I think we mostly agree and I have no issues with the current text. What is worth pointing out, however, is that the "rapidly changing airbeam over the first hour of flight" most likely has indeed nothing to do with any inherent sensitivity change (hence the lack of change in the SI) but is just CO^+ signal from the high LVOC background in the instrument (if my assessment is correct, the same trend will be observed in the CO_2^+ background signal). If this checks out it could be mentioned.
- Regarding the uncertainty of the nitrate comparison in Fig 7a, I don't think I agree with the 35% uncertainty. If this was ambient nitrate (with the usual uncertainties in terms of mixing state, CE and organic/inorganic contributions) that would certainly be the appropriate uncertainty. But here the authors are putting pure AN particles (their analytical standard) into their respective instruments. *So the uncertainty should really come down to the stability of the AMS and the CPC, and not include either RIE or CE, just the overall IE uncertainty. Per Bahreini, that is 10% (explicitly stated in the SI) and should be used here.*
- Regarding the discussion about Collection efficiency (CE), the terms is typically used with two different meanings:

- Operationally, as in CE = Vchem/Vphys for a given particle sizer. This definition also then includes any transmission/inlet issues and does NOT allow for significant errors on the particle sizer side, since it takes it as ground truth.
- Physically, as a defined, rigoruous vaporizer particle bounce correction independent of the agreement with any other instruments. That is what e.g. the Middlebrook et al, 2012 parametrization addresses.

The authors are clearly talking about the operational definition, which for the purposes of proving adequate quantification severely limits which comparisons can be used (only chemical sensors, basically). *Now, based on the UT acidity measurements reported by Nault et al (2021), the UT is almost as acidic (pH<0) as the lower stratosphere. So it would be very surprising if the CE (in the "particle bounce correction" sense) would be anything but 1, not 0.5 for any of the ambient data presented in this paper.* Looking at the other plots in Joppe et al (2025), it does appear that using CE=1 would in general improve the median agreement (as expected) while worsening it in the larger plumes (which is likely related to transmission losses/and or mismatch with the UHSAS size range). Given the challenges, this is a nice finding and supports the overall quality of data of the prototype.

- *One item where a clarification would be appreciated is this sentence in the conclusions: "The time resolution of 30 seconds allows for detection of small-scale spatial and temporal structures on the order of 500 m.". Is this meant to be in the vertical, e.g. while climbing? If so please clarify, because at Mach 0.82 or so (typical cruise speed of a jet airliner), 30 s is about 7 km in the horizontal.*

References

A. M. Middlebrook, R. Bahreini, J. L. Jimenez, M. R. Canagaratna, Evaluation of Composition-Dependent Collection Efficiencies for the Aerodyne Aerosol Mass Spectrometer using Field Data. *Aerosol Sci. Technol.* **46**, 258–271 (2012).

R. Bahreini, B. Ervens, A. M. Middlebrook, C. Warneke, J. A. de Gouw, P. F. DeCarlo, J. L. Jimenez, C. A. Brock, J. A. Neuman, T. B. Ryerson, H. Stark, Atlas, E, J. Brioude, A. Fried, J. S. Holloway, J. Peischl, D. Richter, J. Walega, P. Weibring, A. G. Wollny, F. C. Fehsenfeld, Organic aerosol formation in urban and industrial plumes near Houston and Dallas, Texas. *J. Geophys. Res.* **114**, D00F16-D00F16 (2009).

F. Drewnick, S. S. Hings, M. R. Alfarra, A. S. H. Prevot, S. Borrmann, Aerosol quantification with the Aerodyne Aerosol Mass Spectrometer: detection limits and ionizer background effects. *Atmospheric Measurement Techniques* **2**, 33–46 (2009).

S. Molleker, F. Helleis, T. Klimach, O. Appel, H.-C. Clemen, A. Dragoneas, C. Gurk, A. Hünig, F. Köllner, F. Rubach, C. Schulz, J. Schneider, S. Borrmann, Application of an O-ring pinch device as a constant-pressure inlet (CPI) for airborne sampling. *Atmospheric Measurement Techniques* **13**, 3651–3660 (2020).

B. A. Nault, P. Campuzano-Jost, D. A. Day, D. S. Jo, J. C. Schroder, H. M. Allen, R. Bahreini, H. Bian, D. R. Blake, M. Chin, S. L. Clegg, P. R. Colarco, J. D. Crounse, M. J. Cubison, P. F. DeCarlo, J. E. Dibb, G. S. Diskin, A. Hodzic, W. Hu, J. M. Katich, M. J. Kim, J. K. Kodros, A. Kupc, F. D. Lopez-Hilfiker, E. A. Marais, A. M. Middlebrook, J. Andrew Neuman, J. B. Nowak, B. B. Palm, F. Paulot, J. R. Pierce, G. P. Schill, E. Scheuer, J. A. Thornton, K. Tsigaridis, P. O. Wennberg, C. J. Williamson, J. L. Jimenez, Chemical transport models often underestimate inorganic aerosol acidity in remote regions of the atmosphere. *Communications Earth & Environment* **2**, 1–13 (2021).

Williams et al, EPToF Measurements,17th AMS Users Meeting, Portland, USA, http://cires1.colorado.edu/jimenez-group/UsrMtgs/UsersMtg17/Williams_UsersMeeting2016_eptof.pdf