Author responses to reviews of amt-2021-320: "Aircraft-engine particulate matter emissions from conventional and sustainable aviation fuel combustion: comparison of measurement techniques for mass, number, and size"

by J. C. Corbin et al.

1. RC2

I struggle with the assessment of this work: while the all the methods and results presented are of respectable scientific quality, I think there is a lack of focus in terms of relevance and scope for AMT. There is no novelty in concepts or data treatment and it is not clear what the real scientific value of the study is. For regulatory purposes there is little value due to the non-compliant sampling system, non- existent pre experiment calibration etc. The scientific value is also limited – I understand the argument for connecting ground measurements to cruise at altitude data, but for that purpose, a more focused effort with a better experimental design that would allow tracking down sampling/ conditioning from instrument issues would be beneficial. With the current manuscript one gets the impression that it is a side product of a bigger effort and was not carefully thought through when the experiment was conducted – which is not necessarily a problem if the reader does not get this impression, but I currently do.

Our third-last paragraph in the introduction provided some justification. This paragraph was followed by a misplaced paragraph describing the measurement campaign – that misplaced paragraph has now been moved to Methods, and a new sentence added to the third-last paragraph. The full paragraph is now (new text underlined):

> The standardized system components are not easily adaptable for use on aircraft for measurement of cruise level nvPM emissions. Consequently, there are no comparable in-flight engine-emissions data available for developing and validating models that predict cruise nvPM-emissions based on engine certification data. Particle size distribution measurements are also

not included in the standardized system, which are important for assessing the effects of fuels, operating conditions, and engine technologies on the environmental impacts of PM emissions. Thus to advance our understanding of aircraft engine emissions and the factors that control them as well as to develop a large and consistent observational data base, it is important to evaluate the relative performance of other diagnostic instruments that are not prescribed in the standardized protocol but meet these needs. Such instruments must be evaluated for their response to nvPM and total PM emissions from aircraft engines using standardized and non-standardized systems, and for measurements at the engine exit plane and downstream of the engine in the near field, since these instruments are typically used with minimal change to their operating parameters for a wide range of sampling conditions. Very limited data are available in the literature for this purpose, and no data have yet been published for SAFs.

Thus, this manuscript features one aspect of the detailed analysis that is one facet of a large collaborative project. The manuscript, with its analysis of the response of instruments to variations in the properties of the particulate emissions with fuel type, has implications for in-flight measurements of SAF emission factors, standardized vs. non-standardized measurements, and total vs. non-volatile PM emissions.

Major comments:

The comparison of the mass measurement is somewhat biased experimentally (due to distance to the engine, dilution, detection limits and long lines etc.) to higher thrust levels. At these thrust levels it is not a major surprise that there is not much variability in instrument responses (little OC, larger aggregate sizes, soot properties less influenced by fuel type etc.). I also tend to disagree with the authors conclusion that a 30-50% difference is a “comparable” especially for the near real time in situ instruments such as MSS LII and CAPS. Would be good to point this out to the reader, or even split the discussion for cruising relevant (i.e. 50 -70% thrust) and near idle thrusts this might improve the lack of relevance pointed out above.

The bias to "higher" thrust levels is only caused by the rejection of some test points at 23% thrust. Some 23% data was retained, and the remainder of the data spans 40% to 83%. This range of thrusts is substantial.

We agree with the reviewer's "disagreement" that 30-50% is not really "comparable". We did not intend to imply that a 30-50% disagreement is not statistically significant. We believe that it is significant and implies a systematic bias (e.g. calibration drift or imperfect line-loss corrections) between the instruments. The reviewer may have the impression that we believed otherwise because our discussion focussed on the larger disagreements of the SMPS and filter-based instruments (up to a factor of 2).

When we searched the manuscript for the word "comparable" we could not find that word used to imply no statistical significance. We do agree that we made that implication by omission. We modified Section 4.4.1:

> The agreement of the real-time measurements to within 30 % is ~~notable considering the different types of instruments used~~. larger than the calibration uncertainties of the individual instruments, and suggests an influence of systematic biases (e.g. in instrument calibration or penetration corrections). There is no evidence of systematic differences between absorption and LII measurements, which might have been hypothesized if coatings of volatile PM on the light-absorbing nvPM had enhanced absorption.

Here we also added the underlined sentence to introduce a new hypothesis about why the measurements might differ.

We have not observed any systematic differences by thrust. Figure 11 shows this for N1 thrusts from 40% to 83%. Any differences between instruments are larger than differences between thrusts. So, we have not taken the reviewer's last suggestion.

It would beneficial to show the comparison of measured concentration as a function of CO_2 (at least in the SI)

All requested information was provided in the supplementary data file. The measured CO_2 increment ranged from 0 ppm to 929 ppm, with median 384 ppm.

We take this comment to be related to the comparison of the mass instruments, for example in Figures 8 and 9. We agree that the relevant axis for a mass instrument comparison is mass concentration. However, the instruments shown in Figures 8 and 9 were located on different sampling lines and experienced different levels of dilution. Therefore, we were forced to compare these instruments in terms of EIm rather than mass concentration.

SMPS EIm derivation: this work makes the impression that an SMPS measures the volume size distribution with high precision and there is furthermore no need to apply a size dependent effective density (which I believe is crucial for larger sizes). It would be beneficial for the discussion to elaborate on this based on previous experiences on helicopter or jet engines [...]

The reviewer is correct that we omitted a description of the SMPS PSD mass integration in our Methods section. We now added the following paragraph:

> Finally, the SMPS PSDs were converted to equivalent mass concentrations by the integrated PSD approach, described in detail by Momenimovahed and Olfert (2015). In brief, the equivalent mass of each SMPS-reported mobility diameter was calculated using an effective density of 1000 kg m^{-3}, which has been shown to produce better than 20% accuracy relative to more complete, size-resolved effective densities (Durdina et al., 2014).