Review of "Quantifying the uncertainties in thermal-optical analysis of carbonaceous aircraft engine emissions: An interlaboratory study"

This article presents an interesting and pertinent study on the estimation of measurement uncertainties on the total carbon (TC), elemental carbon (EC) and organic carbon (OC) contents, measured with the instrument Sunset 5L owned by five certified laboratories worldwide. Six equally shared punches from 20 samples probed from the exhaust of a helicopter engine were distributed for analysis among the chosen laboratories. Each instrument user respected the NIOSH5040 analysis protocol. The database was analyzed with statistical multilevel models to identify or predict the uncertainty bias among the samples.

The article is well-written and fits the topic of this journal. The obtained results are highly interesting for research topics such as nvPM emissions measurement protocols, atmospheric measurements, and aviation emissions, measurements, and protocols. There are a few arguably contrasting points that deserve to be put in a better light or clarified.

General remarks to clarify

1. The text does not clearly state when the samples were obtained. Is it the same work as Olfert et al. (2017) or another specific study? Please clarify this aspect.

   The samples were collected on the same engine in the same facility as that reported by Olfert et al. However, they were collected as part of a separate emissions measurement campaign, conducted in Oct. 2016. Olfert et al. obtained their results as part of the MANTRA campaign, which was conducted in March 2015. We have clarified this by stating the following in the manuscript:

   Emissions were collected from the exhaust of a helicopter turboshaft engine using a single point sample probe, in a subsequent study to MANTRA (reported by Olfert et al., 2017), on the same model of engine and in the same facility.

2. Despite the significant experimental work in the referenced work Olfert et al., 2017, this article does not state which engine operating conditions were used for the obtained samples. It may not seem relevant to the authors, but why not have well-identified conditions in which nvPM was produced by the engine? I think this information is essential since the same operating conditions of the engine were used for three sets of samples loaded with 50, 100 and 250 $\mu g/m^3$ of soot, while the last sample loaded with 500 $\mu g/m^3$ of soot was obtained by increasing the RPM of the engine. It is well known that changing the engine's operating conditions will impact the structure and morphology of soot particles/nvPM. Isn't this contradictory with what the authors state in the paragraph from the introduction containing Lines 74 to 77? The reader can find additional information about the sampled particles on the filter by identifying the

operating conditions in the specified article if the work is common and even though the detailed statistical analysis did not identify any correlation between the filter loadings and uncertainties (lines 183-184).

All nvPM samples were collected at high power conditions for the Gnome engine. The samples loaded at mass concentration of 50, 100 and 250 µg/m^3 of nvPM were obtained with the engine running at a steady 22,000 rpm. To produce the higher nvPM concentration required for the samples loaded at mass concentration of 500 µg/m^3, the engine was operated at a steady 23,000 rpm. In both cases, the engine is at high power, and this modest adjustment to the engine's operating condition is not anticipated to impact the structure and morphology of the nvPM particles, as compared to reducing to a low power condition such as the 13,000 rpm reported in Olfert et al. (2017). This is supported by Saffaripour et al. (2020) which shows no significant change in the morphology of the particles from the same engine model between 21,000 and 22,000 rpm.

The following has been added to the manuscript:

"All nvPM samples were collected at high power conditions for the Gnome engine. All the samples loaded at mass concentration of 50, 100 and 250 µg/m^3 of nvPM were obtained with the engine running at a steady 22,000 rpm. To produce the higher nvPM concentration required for the samples loaded at mass concentration of 500 µg/m^3, the engine was operated at a steady 23,000 rpm. Saffaripour et al. (2020) demonstrates that there is no significant change in the morphology of the particles from the same engine model for such modest changes in the rotation speed."

Saffaripour, M., Thomson, K. A., Smallwood, G. J., & Lobo, P. (2020). A review on the morphological properties of non-volatile particulate matter emissions from aircraft turbine engines. Journal of Aerosol Science, 139, 105467.

3. The filter holder from Figure 1 contains two filter holders in series. Was the second filter analyzed for some residual TC content, as presented in the work of Corbin et al. (2020)?

The reviewer is correct in noting that two filters in series were used. The front filter is used to collect the sample for TC, EC, and OC analysis. It is known that quartz filters adsorb gas phase organic artifacts, and the second quartz filter is used to correct the OC and TC measurements from the front filter for the gas phase organics that were adsorbed on the front filter.

The following has been added to the manuscript:

"Quartz filters adsorb gas phase organic artifacts, and following the procedure outlined in Corbin et al. (2020), the data from TOA of the quartz filter in the second filter holder shown in Figure 1 is used to correct the OC and TC measurements from the front filter for the gas phase organics that were adsorbed on the front filter."

4. It is surprising that the different loadings of the samples do not affect the uncertainty measurement of the three quantities measured by the Sunset instrument. This finding deserves a more detailed discussion since studies show that the loading of the filter impacts the uncertainty measurement of the thermo-optical analysis measurements.

   The sampling times were adjusted such that the loadings were similar for all filters, regardless of the source concentration. Minor loading differences were observed, but they were small relative to differences in the mass concentration.

   The following has been added to the manuscript:

   "To compensate for the different mass concentrations used for loading the filters, the sampling time durations were adjusted such that the mass loadings were similar for all 20 filters."

5. The use of the word structure (lines 183, 225, 229) and structural trends (line 227) can sometimes be misleading in the text for readers who are not specialized in multilevel statistical analysis. Please be more specific where it is the case; such as data/uncertainty structure or something that fits better in the context.

   The terminology around "structured" errors has been removed from the manuscript in favor of discussion around systematic biases or effects. The biases also have some surrounding description to clarify the author's intention:

   "Results for EC and TC exhibit a consistent bias (or systematic error (JCGM, 2008)) across the different filters, where a laboratory that measured a value above average generally did so for all of the filters."

Specific comments:

Figure 2 - to which sample corresponds to the obtained data? it is worth mentioning.

The data shown does not correspond to any particular measurement in this work. The caption has been updated accordingly:

"Figure 2. Representative example of a TOA thermogram for nvPM emissions collected from the engine used in this study. Shown are the thermal protocol for aircraft engine emissions (SAE, 2018; Lobo et al., 2015a), the sample temperature, the FID signal, and the laser transmission measurement."

Figure 6 - what represents the error bars in the bottom graphs with the Relative value [%] since it is mentioned that the error bars are excluded for clarity?

The error bars correspond to laboratory-reported uncertainties on individual points. The caption has been updated:

"Error bars in the lower panels correspond to expanded (k = 2) uncertainties reported by the laboratories and, while only included for select points, were similar across all of the data."

Line 37: remove on from the sentence "… mass on collected …"

This change has been made as recommended.

Line 110: "darkness …" can be replaced with "coverage …" to differentiate from dark uncertainty

Coverage is now used in connection with expanded (k = 2) uncertainties. We rather use "loading":

"The loading of most filters was visually homogeneous, which further supports this decision."

Line 136: remove the from the sentence "… and the their uncertainty …"

This sentence was removed amongst the other changes.

Line 199: The authors mentioned, "These filters coincide with cases where the overall variance is larger and represent a minority of cases." Please be more specific when selecting a criterion for the value of the variance to eliminate the sample in question. Either be specific and justify why this selection was made or mention if you referred to data outliers.

The data for these samples were not excluded from the analysis, such that no specific criterion was applied to remove the data or identify them as outliers. Further, the updated statistical model also no longer relies on laboratory-reported uncertainties to determine within-laboratory variation, such that these variances are less relevant to the analysis (only introduced as weak prior information). Overall, the quoted sentence added more confusion than clarification and has thus been removed from the revised manuscript.