Reviewer 2

This study explores the uncertainties involved in measuring equivalent black carbon mass concentration using micro-aethalometers. Under controlled laboratory conditions, five micro-aethalometers (MA200/MA300) were evaluated for measuring black carbon aerosols with known mass concentrations, generated using a reference mass standard coupling MISG and CPMA. This study successfully isolates 3 types of uncertainties and develops a comprehensive uncertainty model that considers factors such as mass concentration, sampling intervals, and flow rates. Additionally, this study presents an open-source algorithm designed to optimize sampling strategies based on targeted uncertainty levels. The experiments were carefully designed and the manuscript is well written. The following issues need to be addressed.

Major Comments.

1. The title should be revised, as this study is limited to laboratory conditions and does not consider ambient-related factors such as temperature, relative humidity, and low EC/TC ratio samples. An alternative title can be "Laboratory-based uncertainty quantification and post-processing strategies for micro-aethalometers measuring black carbon". The following statement should be included in the conclusions. "It should be noted that the uncertainty model derived in this work is based on laboratory-generated, non-volatile soot under stable conditions. Its applicability to ambient measurements, where factors such as fluctuating temperature, relative humidity, and the presence of externally mixed non-BC aerosols or coatings on BC particles (leading to low EC/TC ratios) can significantly influence the measurement, requires further validation."

We agree with the final statement's intention and had emphasized this in our Introduction. We have now also added an edited version of these statements to the conclusions.

In general, while the reviewer's points are valid, the objective of this manuscript was not to characterize the overall slope of the line relating the external CERMS reference and the aethalometer response but to characterize (1) the intercomparability of the sensors and (2) the noise profile of the sensors. In theory, these two quantities should not depend as significantly on what is being measured. For example, if the particles absorb more light due to absorption enhancement from coatings, this should affect all of the devices similarly and can be accommodated by using a separate calibration factor. This work then describes the remaining base uncertainty of the instruments.

Rather than changing the title, we have added the following statement to the abstract:

> *"This represents a minimum uncertainty estimate, given a well-understood aerosol."*

At the beginning of the conclusions, we now emphasize that we use an ideal aerosol for measurements:

> *"Experiments were performed using an ideal aerosol – composed of laboratory-generated, non-volatile soot – to compare five different aethalometers with reference mass concentrations generated using a CPMA-electrometer reference mass standard (CERMS)."*

We have also added a clarifying note to the conclusions:

"It should be noted that these expressions only capture repeatability and inter-device variability. They do not intend to account for other systematic artifacts, including cross-sensitivity to scattering, humidity, and temperature. These effects are often location- and time-specific and should be addressed by using appropriate calibration factors, as is currently common practice. Rather, these expressions give the minimum uncertainty that can be achieved after the aethalometers have been calibrated for the properties of the specific particles being measured. We also do not assert that this model will account for all uncertainties in the measurements, as, while systematic artifacts in the data can largely be removed by way of calibration (similar to calibration to remove inter-device variability), the calibration factors used to perform this correction will themselves have uncertainties (whether due to physical fluctuations during the measurements or incomplete knowledge of the artifacts) that must be considered alongside the uncertainties here."

We also added scattered comments throughout the text clarifying some of the limitations emphasized by the reviewer, in the introduction and in Sec. 3.1.

2. The main finding of this study is the identification of three types of uncertainties: multiplicative inter-device bias, random noise due to dual-spot correction, and additive systematic bias. This finding must be explicitly stated in the abstract.

This has been stated more clearly in the conclusions:

"Three types of measurement errors were characterized. First, device effects were multiplicative, consistent with other studies in the literature. Second, while the dual spot correction algorithm was found to be effective in correcting biases in the measurements, the correction caused an increase in random errors in the measurements. This noise was well-modelled as a signal independent Gaussian noise source. Finally, Poisson noise was present in the measurements, which is likely attributable to the discrete arrival of the particles on the filter. This noise will exhibit a particle size dependence."

This has also now been stated in the abstract:

"Uncertainties were found to scale with mass concentration, with Gaussian, Poisson, and multiplicative components. The multiplicative errors between devices are approximately 10 % in the best case of long sampling times and/or high mass concentrations."

3. At the end of Section 1, the authors state that "Since our aerosol model represents a simple source of eBC, with negligible content of non-absorbing PM and stable gas-phase composition, our results provide a lower limit on between-instrument reproducibility." Compared to mini-CAST, MISG tends to generate samples with a high EC fraction (>90%). Did the authors verify the actual level of EC fraction in this study? If the EC fraction reported in this paper is indeed greater than 90%, then the results apply only to source emission scenarios. For ambient atmospheric aerosols, where the EC fraction is typically less than 10%, could the authors provide insights into how the

EC/TC ratio might affect the uncertainty of micro-aethalometers when applied to ambient aerosols? Although this is not the primary focus of the current study, understanding this relationship would be valuable.

Previously work at the same operating conditions characterized the EC fraction as roughly 95%, with some uncertainty. The authors noted that the operating conditions were not specified in the previous manuscript. So, these details have been added to the manuscript alongside a reference to the previous work where the EC/TC was quantified:

> *"For testing, the aethalometer chamber was periodically filled with soot, after which the inlet to the chamber was closed and the concentration of particles in the chamber was allowed to decay slowly over time, as shown in Figure 1. BC particles were generated using a Mini Inverted Soot Generator operated on ethylene (MISG; Argonaut Scientific; (Kazemimanesh et al., 2019)) operating with 0.1 SLPM of ethylene and 10 SLPM of air. This is expected to produce particles with EC/TC values in excess of 0.9 (Kazemimanesh et al., 2019)."*

We otherwise acknowledge that organics will impact the measurand. As we noted in our first response, EC/TC (refractory carbon as measured in thermo-optical analysis) and eBC (light absorption) are correlated properties only because soot possesses both features. If a sample contains infrared-light-absorbing, non-soot carbon, then EC/TC ratios and eBC (as measured by aethalometers) will diverge even in the absence of any measurement uncertainty. This is also true if a sample contains soot with varying degrees of organic coatings (often described as "lensing/E_{abs}") or soot of varying degrees of graphitization (described sometimes as maturity). These deviations are not uncertainties but reflections of the complexity of light-absorbing carbon. They have been discussed in detail elsewhere, e.g., Corbin et al. 2019 (https://doi.org/10.1038/s41612-019-0069-5),

In response to the reviewer's concerns, we have added the following text when introducing the model:

> *"As noted in the introduction, this model represents a minimum uncertainty that does not take into account various causes of artifacts that can occur in aethalometer measurements (such as particle size, composition, and rapid changes in gas composition). Such additional artifacts will add further uncertainty."*

4. The statement in Section 2 that "The correction factor is taken as that provided by the instrument firmware and applied during post-processing" is incorrect. The k value is derived by solving the system of equations from Eq. 1, which was obtained under different flow rates at the two sampling spots. For each data point at time t, a corresponding k value can be calculated. The k value is not preset in the instrument's firmware.

We did not use a static value of the correction factor, but rather the value of the correction factor provided by the instrument for each measurement interval. We have clarified this in the manuscript:

"The correction factor is taken as that provided by the instrument (firmware v1.1) at each time and is applied to the measurements during post-processing."

5. The authors mentioned, "We tested five microAeths (two MA200, three MA300, AethLabs, USA) that were previously deployed at ambient monitoring sites across Canada." Did the authors calibrate the flows of the MA200 and MA350 before the experiments? The accuracy of the flow readings can significantly affect the inter-device bias.

Yes. The flows were calibrated according to the manufacturer instructions. We have added clarifying notes to this effect:

"The flowrate of the micro-aethalometers was calibrated immediately prior to testing according to the manufacturer instructions. The micro-aethalometers themselves were calibrated by the manufacturer approximately a year prior to testing."

6. Did the authors check the firmware version of the five micro-aethalometers? The reviewer's experience with the MA series indicates that the firmware version can influence the inter-device bias.

Yes. We were using firmware version 1.1 for all of the instruments. The firmware version number was included alongside a previous change:

"The correction factor is taken as that provided by the instrument (firmware v1.1) at each time and is applied to the measurements during post-processing."