Useful work that needs some clarifications
R Subramanian (Referee)


(Disclosure: I worked closely with Professor Zimmerman when she was a postdoc at Carnegie Mellon University and we co-authored a few manuscripts published over 2018-2021.)

Instrument inter-comparisons are essential to the research community, as we need to make sure we are investigating real atmospheric composition differences, not manufacturing or design artifacts. So this is a useful submission from a carefully-conducted study investigating urban black carbon, which is now finally recognized by the WHO as an air pollutant of interest on which more work is required. Papers like this will help build a database on which future WHO guidelines can be based.

The work can be published after some revisions/clarifications as noted below.

- Lensing is important, it should at least be mentioned. Lensing can lead to inflated eBC values using manufacturer’s default MAC values especially during wildfire periods; impact on source apportionment is uncertain (whether lensing has a wavelength dependence). E.g. Saleh et al. https://doi.org/10.1002/2015JD023697 and Bond et al. https://doi.org/10.1029/2006JD007315
- Effect of filter loading (BC mass per unit filter area) should be considered, which can explain key results - e.g. noisier MA300 (0.15 lpm) data at lower BC values, better performance (lower NRMSE) during wildfire periods, and greater precision of AE31 in a previous study (as sample flow rate is 2-5 lpm).
- Drinovec algorithm considers the effect of variable flow rates, and flow rate variability should impact all wavelengths equally. So it is unclear why flow rate variability should result in the Drinovec algorithm not working for MA300 (unless they did not include that factor) nor why it performs quite well for UV absorption but not for IR absorption. How did the authors verify flow rates in the MA300?
- Was Fig 1b for SD of MA300 averaged across all three units and Fig S8 for unaveraged...
data? The latter seems more representative - though perhaps you should (a) normalize response of each MA300 (as that is a known bias, not noise) and then calculate SD for normalized response.

- Table 1 shows individual MA300 data, which are useful (given N=3), but Sec 3.2.2 discusses apparently the average of the three units - which masks the significant variability in device performance. The latter may be more useful to the reader as people may buy just one unit ($10k is a lot of money!), and whether they get unit B or unit C makes a huge difference.

- Lines 364-367 - results for filter loading effect not shown elsewhere; OA hygroscopicity speculative. Suggest deletion or clarification.

- Suggest running source apportionment for a week before and after the wildfire period to minimize the effect of seasonality on fossil fuel BC. Also, GDI vehicles could also contribute BC especially in urban areas, not just diesel vehicles, as shown by this excellent paper: https://doi.org/10.1021/acs.est.5b04444

- Lines 395-396 - is that for the regular, non-wildfire period? (Also, just type out “wildfire” and “regular” or “non-wildfire”, using acronyms/abbreviations is annoying and this does not reduce the word count.) What is the MDL below which source apportionment is not robust?

- If the Drinovec algorithm is not appropriate for the MA300, why are source apportionment results with this algorithm discussed in the main text? Delete or move to SI.

Minor comments on language/phrasing/oddities in the marked up PDF, attached.

Please also note the supplement to this comment: https://amt.copernicus.org/preprints/amt-2022-278/amt-2022-278-RC1-supplement.pdf