Response to Reviewer #1: AMT-2022-278

The authors would first of all like to thank the reviewer for the insightful comments on the work we have submitted for publication, and the editor for the opportunity to improve the manuscript. Under each reviewer comment (in red) there is a summary of the response (black text), in addition to the text from the manuscript (in blue) that was modified, if applicable.

(1) 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

Author's Response (MC): We thank the reviewer for highlighting the importance of the lensing effect on an aethalometer's measurement. Even though mixing state plays a critical role, the fact that IR specific light absorption is only due to pure BC has been well accepted in the literature. Since, source apportionment calculations utilize the two-wavelength measurement, where light absorption measurement in lower wavelengths can be significantly impacted by the lensing effect due to the presence of BrC compounds, we agree that this effect may cause uncertainty in eBC apportioned components. We included this discussion in the introduction section. (**L62-L67**)

In addition, the mixing state of aerosol can influence the light absorption measurements of bulk aerosol (Bond and Bergstrom, 2006; Saleh et al., 2015; Healy et al., 2015). At lower wavelengths (near-UV), light absorption measurements were found to be enhanced (by a factor >2) due to the presence of brown carbon (BrC) during wildfire-affected aerosols in Canada (Healy et al., 2015). The same study identified minimal light absorption enhancements due to the lensing effect at higher wavelengths (near-IR) regions. As such, light apportionment-based BC measurements and their source apportionment can be further complicated by the bulk aerosol's source and mixing state.

Also in Section 4, we mention that our study could not explicitly estimate the impact of lensing in light absorption estimations, particularly during the Wildfire period. We identify this as a source of uncertainty in our work and highlight as a potential important area for future work.

Section 4: This study did not take into account the lensing effect of BC, which has been identified as being particularly relevant during wildfire periods (Healy et al., 2015), and can impact the light absorption coefficient measurement. Evaluating how lensing impacts the b_{abs} measurements of MA300 is an important area for future work.

(2) 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).

Author's Response (MC): Thank you for this helpful suggestion. We have adopted the filter loading calculations you suggest. The results of the filter loadings were estimated for all the devices for individual periods (**Table S6**). For individual devices, filter loadings were estimated from real-time flow measurements and then averaged for the regular and wildfire periods. We then use instrumental noise estimates (as suggested by Reviewer #2) and MAE to identify instrumental variability during the wildfire episode. We verified that larger errors in Wildfire periods are likely related to higher filter loading and updated the text accordingly. The discussion on filter loading was included in **Section 3.1.2 (L359-L368).**

Text from Section 3.1.2

As mentioned in Section 3.1.1, filter loading during a high BC event can lead to measurement errors that vary by each instrument's sensitivity. Mean filter loading (eBC mass per unit filter area, see Table S6) during the Regular period were 1.63 μg cm^{-2}, 1.46 μg cm^{-2}, 1.96 μg cm^{-2} and 5.8 μg cm^{-2} for devices MA300A, MA300B, MA300C and AE33, respectively. For the Wildfire period, the filter loading increased across all the devices (to 5.49, 4.85, 5.79 and 19.0 μg cm^{-2}, respectively) as expected. During the Wildfire period, filter loading per unit volume of air sampled increased by a factor of 2 in AE33 (see Table S6). However, for MA300 units, this factor varied (2.4 for MA300A, 2.27 for MA300B and 1.9 for MA300C). Even though, all these devices measure the same environment, we find differences in the aerosol loading on MA300 filters. We hypothesize this variability might be occurring due to the variability in sampled airflow and instrumental noise. By studying the instrumental noise and filter loading estimates, we find that the error contribution in MA300's eBC measurement can be sensitive to their exposed concentration range. This observation aligns with a previous study on MA-series aethalometers, where the impact of high eBC concentration has been found to impose large errors from more pronounced filter loading corrections (Alas et al., 2020).

Table S6: Filter Loadings experienced by aethalometers during the measurement periods.

Device	Filter Loading (μg/cm^2)		Filter Loading per unit volume of air sampled (μg/cm^2/mL)	
	Reg	WF	Reg	WF
AE33	5.8	19.0	0.002	0.006
MA300A	1.63	5.49	0.016	0.055
MA300B	1.46	4.85	0.015	0.049
MA300C	1.96	5.79	0.020	0.058

(3) 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?

Author's Response (MC): This is a great point raised. The Drinovec algorithm considers non-linearities in the flow measurements in estimating the loading correction factors (Equation 4 of this manuscript).

$$\frac{F_2}{F_1} = \frac{\ln(1 - k \times ATN_2)}{\ln(1 - k \times ATN_1)}$$

The reason for the impact of variability in flow rate across different wavelengths of light absorption measurement has been explained by Drinovec et al., (2015). Briefly, filter spot advancement is driven by the lowest wavelength, i.e. UV (until a set attenuation value is reached). Corresponding measurement signals from other channels tend to be lower as wavelength increases. The differences in wavelength-specific attenuation values determine the face velocity ratio factors (referred as FVRF by Drinovec.) which was found to significantly impact the k-values. AE33's onboard correction mechanism uses flow dependent-corrections using Drinovec et al., (2015), and any loading correction estimates will be sensitive to small errors in face velocities. As compared to AE33, the MA300 has significantly lower face velocities, which can increase the overall sensitivity of the correction method (modified Drinovec method, Section 2.4) tested on MA300s. MA300's on-board correction algorithm (Section B of Supplementary Information) is independent of the flow measurements, however, comparatively higher flow variability has been observed in MA300, as compared to AE33. Raw data from both MA300 and AE33 includes real-time flow

measurements from spot 1 and spot 2, which have been used to calculate the percentage deviation in MA300's flow compared to the setpoint. We have added discussions on the flow measurements and their relationship to the MA300-adopted modified Drinovec correction algorithm in **Section 3.4**. (**L469-L475**).

From the onboard mass flowmeter readings, we find that filter spots 1 in all the MA300 units were drawing comparatively lower airflow with wider variability. MA300 uses a sampling flow rate of 150 mL min^{-1}, and ideally, 2/3rd (100 mLmin^{-1}) is split to filter spot 1. MA300's flow setpoint deviation ranged from -9.7% to 2.4% for MA300A, -8.9% to 5.8% for MA300B, and -14.2% to 1.4% for MA300C. In contrast, AE33 was run at 5 L min^{-1} (with set airflow of 3333.33 mL min^{-1} on filter spot 1) and had smaller deviations from the setpoint (-0.44 % to 3.2%). A high range of flow variability can lead to additional noise in corrected eBC measurements when a flow-based correction technique, such as the Drinovec et al. (2015) algorithm, is adopted in MA300 devices.

Also, in the conclusion (**L605-L606**) we now discuss the regular flow audits we have performed in this work.

Since MA300 operates at very low-flow conditions, we regularly audited and calibrated the flow (twice a month) and we recommend that MA300 users conduct routine flow audits while doing continuous sampling, particularly in a highly polluted environment.

(4) 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.

Author's Response (MC): In Fig 1b, our goal was to look at the between-unit SD of MA300, which could be considered a measure of unit-to-unit variability. We created a pool of MA300's measurements (for absolute and normalized data separately) for each bin (of width 1 μg/m^3) of AE33 reported eBC data. We then plot the standard deviation of the pooled measurement against the AE33's measurement. The slope of the linear fit corresponds to the linear response of MA300's variability across the concentration range.
In our earlier submission, Fig 1b was for SD of MA300 averaged across all three units and Fig S8 for unaveraged (pooled) data. As per the suggestion, we replaced Fig 1b with the normalized responses of MA300's pooled data. Additionally, we updated supplementary information (Fig S7), to show the pooled MA300 variation (non-normalized). Normalized eBC concentrations, when fitted against normalized binned AE33 concentrations showed a 5% unit-to-unit variability. The normalization approach we used is summarized in the equation below, and is explained in **Section 2.7** (**L270-L274**) and supplementary **Section E**. In addition to the normalized response, in the updated discussion, we have also reported the calculated variability of 21% from the non-normalized measurements of eBC. The updated discussion can be found in **Section 3.1.2 (L337-L345)**

Normalization process discussed in Section 2.7

Relative accuracy assessments were performed using the slope of the linear fits. The measured and derived parameters had different scales or ranges over wavelength channels, MA300 units, and periods. Hence, to remove measurement bias and to focus on instrument's variability, we choose to normalize the data for individual groups by scaling with respect to their range of measurement, also known as min-max normalization (Géron, 2022).

From Section E of Supplementary Information:

$$N_{normalized} = \frac{N - N_{min}}{N_{max} - N_{min}}$$

For any parameter N, normalized quantity ($N_{normalized}$) was derived by the respective range, i.e. the difference in the maximum (N_{max}) and minimum (N_{min}) values.

From the Section 3.1.2:

First, we normalize the absolute measurements from MA300 units by the range of concentration (max – min criteria) to only consider the measurement bias and eliminate any device-specific noise. Next, we created a pool of MA300's measurements (for absolute and normalized data separately) for each bin (of width 1 μg m^{-3}) based on AE33 reported eBC data. The standard deviation of the pooled measurement from MA300 were fitted against the AE33's measurement. The slope of the linear fit corresponds to the linear response of MA300's variability across the concentration range (Figure 1(b)). The slope of this linear fit line is 0.049, and can be interpreted as an approximately 5% variability across MA300 units of hourly eBC mass concentration. When compared against non-normalized measurements, MA300 exhibited 21% unit-to-unit variability (see Figure S7). In Figure 1(b), we also see that a linear fit cannot totally explain MA300's unit-to-unit variability (R^2 = 0.6) and depend on the observed concentration range, which suggests estimates of unit-to-unit variability can change based on the range of eBC concentration.

Updated Figure 1 (b):

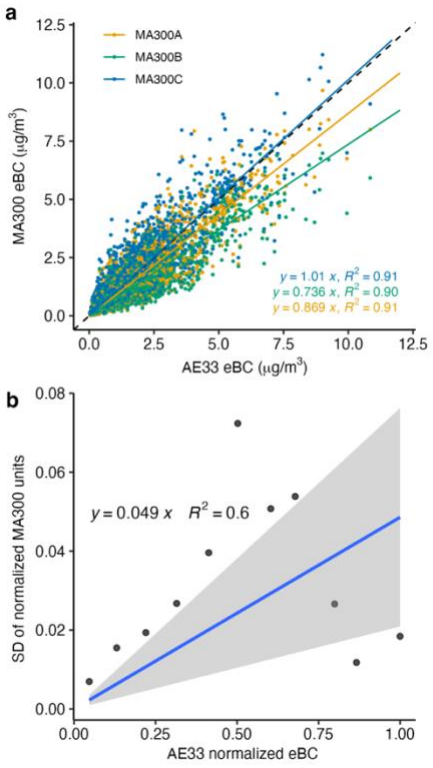

Figure 1 Scatter plot of eBC mass concentration for individual MA300 units A,B and C vs AE33. The dashed line represents the 1:1 line, and solid colors are the regression fit lines for the individual MA300 units; (b) Linear relationship of multi-unit pooled Standard Deviation (SD) from normalized MA300 measurements for each μg m^{-3} of normalized AE33 eBC concentration. The fit line (in blue) represents the linear response of MA300's variability across the concentration range. The shaded region represents the 95% CI of the fit.

Updated Figure S7

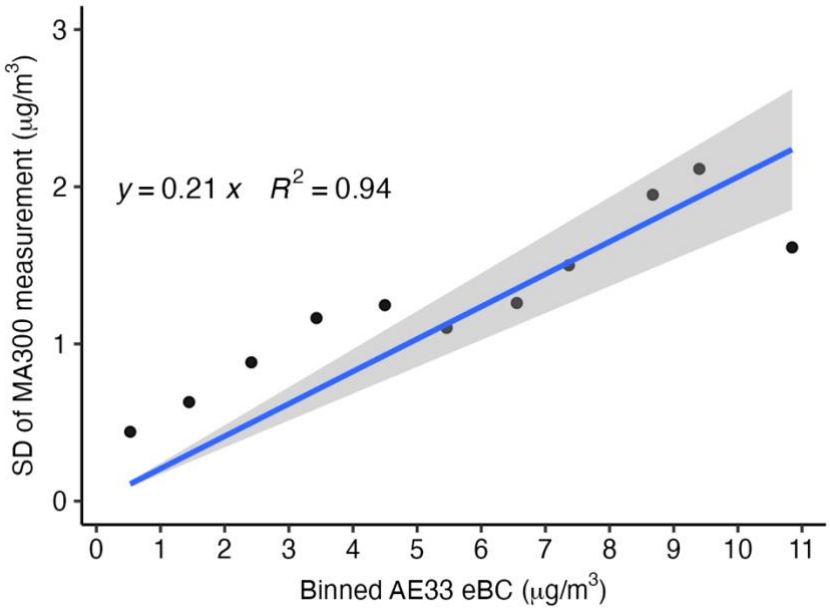

Figure S 1Multi unit pooled standard deviation from MA300 measurements for each one µg. m^{-3} of eBC concentration measured by AE33. The fit line (in blue) represents the linear response of MA300's variability across the concentration range. The shaded region represents the 95% CI of the fit

5. 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.

Author's Response (MC): This is a good point and is aligned with the objectives of this work. Table 1 serves the purpose of providing readers with a statistical summary of the measurements made in this study. Based on the individual unit's response, we want to highlight the inconsistency observed in MA300 units. A handful of studies have focused on the performance evaluation of multi-wavelength black carbon (BC) measurements from Dual-Spot-based micro-aethalometers (e.g. MA300). However, a one-to-one comparison of individual MA300 units and the reference method may not clarify the overarching performance of such micro-aethalometers. As we understand, the growing focus on BC emission and its source estimates across the globe might lead to multi-unit aethalometer measurements. Therefore, both overall and individual metrics are important for understanding the extent of measurement uncertainty and exploring the reliability for future users of micro-aethalometers.

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

Author's Response (MC): We thank the reviewer for pointing this out. As mentioned in the response to Major Comment #2, a discussion on filter loading is now included in **Section 3.1.2 (L357-L366).** In the updated discussion in Section 3.5, we establish the impact of strong loading effects and RH on the MA300's measurements. Additionally, a more detailed explanation of aerosol aging and its relationship with hygroscopicity has been included. Section 5 (**L485-L495**)

In our study, all three MA300s were influenced by strong filter loading (Table S6) in addition to RH changes (45% to 95%) during the sampling periods. Being a near-road emission measurement site, our measurements captured complex aerosol mixtures of various mixing states. During the regular period, local traffic during the daytime contributes to fresh BC-enriched aerosols, which can be hydrophobic in nature (Sarangi et al., 2019; Wang et al., 2020) and by night-time, these fresh BC-enriched aerosols can evolve by aging and change their morphological and optical properties. In contrast, during the wildfire smoke-affected days, the measurement site experienced increased quantities of aged aerosols through long-range transport from the Pacific Northwest. These claims align with our calculated α values, as shown in Fig. 3. With the abundance of organic aerosol components during wildfire days, coated BC particles have been found to dominate and often enhance light absorption in lower wavelengths due to the presence of BrC (Healy et al., 2015). This wildfire smoke-affected BC particles can be mixed with a significant fraction of secondary organics, which can be hygroscopic in nature (Wang et al., 2020).

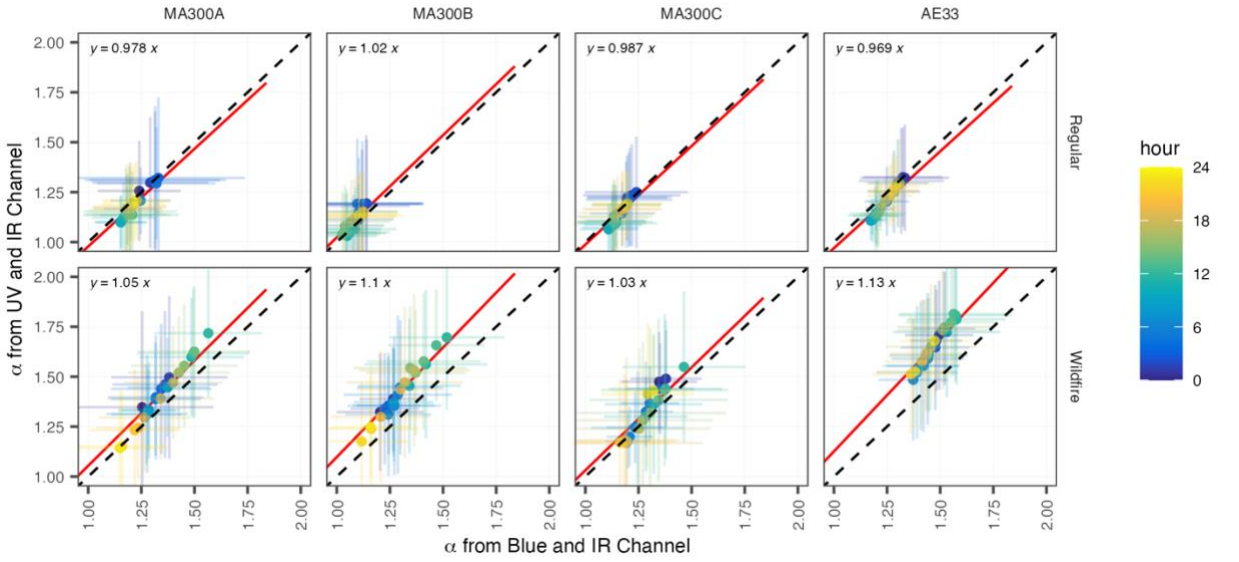

Figure 3: AAE (by the hour of the day) measured by different aethalometers during the Regular and Wildfire period. Average hourly AAE values derived from the Blue-IR wavelength pair (on x-axis) and the UV-IR wavelength pair (on y-axis) with error bars representing respective standard deviation. Red lines represent the linear relationship (forced through the origin), and the dashed line is 1:1.

(7) 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

Author's Response (MC): This is a great suggestion. We have performed Source Apportionment on a week of data before the wildfire period (August 24 to August 30) and on a week of data after the wildfire period (October 14 to October 20) and included these new results in the discussion (**Section 3.5.1, L527-L532**). In this study, we have also used traffic data measured by MetroVancouver. Our apportioned eBC_{ff} results overlapped with peak traffic periods, however the fuel type of vehicles could not be specified. We used general fossil-fuel based traffic sources in the updated discussion.

To verify the impact of seasonality on eBC_{ff} during the wildfire period source apportionment results, we considered two additional week-long periods (August 24 – August 30 and October 14 – October 20) before

and after the wildfire smoke-affected period (Figure S12). We found that the fossil-fuel component dominated the eBC mass throughout the day and eBC_{ff} range was similar to the Regular period. We find slight difference (increase in pre-wildfire period) in the eBC_{ff} concentrations in the late night hours. Both pre- and post-wildfire weeks followed almost similar diurnal eBC_{ff} profile as the Regular period. Hence, no seasonality in eBC_{ff} were identified with this analysis.

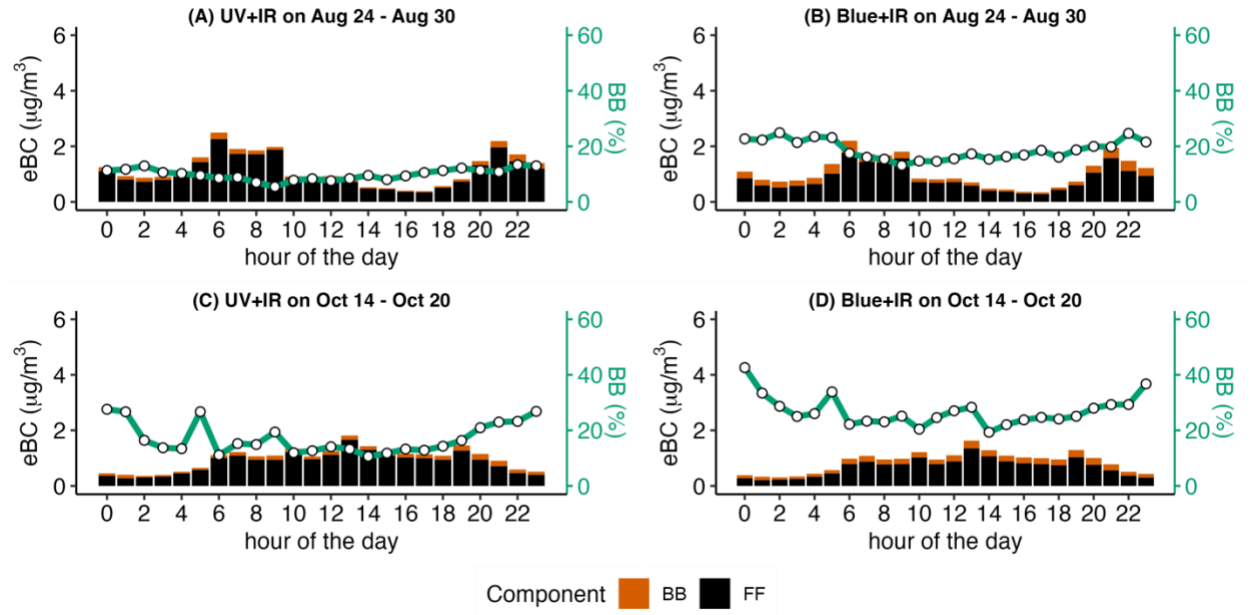

Figure S 12: Source apportionment results (similar to Figure 4 in the main text) but for two specific periods. Panel A and B are pre-wildfire period: August 24 to August 30. Panel C and D are for post-wildfire period: October 14 to October 20.

(8) 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?

Author's Response (MC): We thank the reviewer pointing this out. As recommended, we have updated the manuscript "Reg" to "Regular", and "WF" to "Wildfire". Based on particle free measurements, we calculated MDL of MA300B and MA300C as 0.21 and 0.43 µg m^{-3} (MA300A had measurement error, hence excluded). Clearly instrumental sensitivity played a big role in determining the MDL of MA300. Considering the lowest identified MDL value of eBC measurement, we argue that any source apportionment should not be conducted below the MDL of the black carbon concentration, which is 0.21 µg m^{-3}. However, results from sensitivity analysis (**Figure S10** and **S11**), reflected that, different combinations of α pair can provide negative $b_{abs,bb}$ and $b_{abs,ff}$ compounds which can again be treated as artifacts of source apportionment method.

(9) 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.

Author's Response (MC): We thank the reviewer for this suggestion. In the revised manuscript, Figure 6 have been updated to present the results from MA300's onboard correction only. The results from the responses of modified Drinovec correction were moved to SI. (Figure S13).

Minor Comments

Line 21: improves by 10% or comes to within 10% of reference monitor?

Author's Response (MC): Here we refer that the results of source apportioned components improved by 10% when we choose Blue-IR method as compared to UV-IR method. This has been clarified in the Abstract (**L24-L25**).

When the Blue-IR based source apportionment technique is adopted instead of the UV-IR, there is a 10% (on average) decrease in the percentage difference of the apportioned components from the reference monitor.

Line 53: coating by non-BC material produces a lensing effect that complicates these measurements. Not considered in this manuscript, must be at least mentioned. e.g. Saleh et al. https://doi.org/10.1002/2015JD023697 and Bond et al. https://doi.org/10.1029/2006JD007315

Author's Response (MC): We thank the reviewer for pointing this out. We have added the following lines in the main text to explain the importance of coating on BC particles. (**L62-L67**)

In addition, the mixing state of aerosol can influence the light absorption measurements of bulk aerosol (Bond and Bergstrom, 2006; Saleh et al., 2015; Healy et al., 2015). At lower wavelengths (near-UV), light absorption measurements were found to be enhanced (by a factor >2) due to the presence of brown carbon (BrC) during wildfire-affected aerosols in Canada (Healy et al., 2015). The same study identified minimal light absorption enhancements due to the lensing effect at higher wavelengths (near-IR) regions. As such, light apportionment-based BC measurements and their source apportionment can be further complicated by the bulk aerosol's source and mixing state.

Line 56: did you mean to write soot?

Author's Response (MC): We thank the reviewer for catching this error. Yes, we meant soot and not aerosol. We have made this change in the manuscript. (**L59**).

Fossil-fuel-based aerosol sources (e.g., diesel vehicles) generate soot, which tends to absorb light uniformly across the spectrum (Bond and Bergstrom, 2006)

Line 145: no mention of lensing effects? scattering of *light by* loaded aerosols

Author's Response (MC): We thank the reviewer for pointing this out. We have updated **Section 2.3** (**L156-L163**) where we include discussion on lensing effects.

Filter-based light absorption techniques are subject to measurement artifacts due to scattering on the filter, scattering of light aerosols loaded on the filter surface or due to some particles being shadowed by others (Weingartner et al., 2003). In addition to these measurement artifacts, current aethalometer real-time correction algorithms do not consider light absorption enhancement occurring from the lensing effect, particularly at lower wavelengths due to light-absorbing organic components. Present designs of a stand-alone aethalometer are not equipped to estimate light absorption enhancement in real-time, as they can not distinguish the aerosol mixing state and focus on bulk aerosol properties. Therefore, aethalometers require proper estimation of loading compensation factors and multiple scattering factors for accurate measurement (Virkkula et al., 2007; Weingartner et al., 2003; Virkkula et al., 2015).

Line 169: It is important to note that MA300 corrections do not include any filter leakage parameters and use higher MAC values as compared to AE33. – why?

Author's Response (MC): We are using the manufacturer provided MAC values for both MA300 and AE33. Wavelength-specific MAC values are usually provided by the manufacturer based on laboratory tests carried out while developing the instrument. It is worth noting that the MA300 devices used in this study operated on firmware v1.10, which did not include leakage parameters for BC estimation. We anticipate that this will be addressed in future firmware upgrades. In the updated manuscript, we remove this line and modified the text as we found this line redundant. (**L187-L189**)

Even though the MA300 includes flow measurements in the raw data files, it does not consider any lateral filter leakage parameters and flow values in their correction algorithm. Wavelength-dependent MAC values for both MA300 and AE33 models (Table S1) were taken from the user manuals provided by their respective manufacturers.

Line 234: This is a strange metric and not used anyway, so delete.

Author's Response (MC): We have considered this suggestion and removed bR2 estimates from the manuscript and SI.

L266: "calculated", not "estimated" (which means an approximation)

Author's Response (MC): We thank the reviewer for pointing this out. We have updated in the manuscript accordingly.

L275: doesn't this simply mean that the MA300 performs better at higher concentrations (aka higher filter loadings)? You could calculate the filter loading (microg/cm2) at the different flow rates (5 lpm for AE33 vs 0.15 lpm for MA300).

Author's Response (MC): This is a great point and we further investigated our work to reframe the error matrices used in this work. In our previous calculation, we calculated normalized error as, $NRMSE = \frac{RMSE}{mean(AE33)}$ for individual periods. However, we recognize that choosing mean value for individual periods does not take into account the variability in the concentration. Upon further investigation, we found that the range of hourly eBC concentration during the Wildfire period is much wider than that of Regular period. In such cases, mean based NRMSE calculation may not be sufficient. Hence, we adopt a range based NRMSE estimation, which allows for a more meaningful comparison of MA300's performance in both Regular and Wildfire periods. The normalization process is based on the range measured by each individual device and has been explained in detail in the updated manuscript (**supplementary Section E**), and previously mentioned in the response to major comment #4. We calculated average NRMSE of MA300 units as 8.5% during the regular period and 12.5% during the wildfire period. We hypothesized that the increasing relative error during the wildfire period could be related to the filter loading effect. As suggested by the reviewer we calculated the filter loadings (in μg cm^{-2}) for individual instruments and compared the results with the errors calculated (**Table S6**). Since AE33 and MA300 samples at different flowrate, we normalized the filter loading estimates by their set flow. We verified that larger errors in Wildfire periods are related to higher filter loading and low face velocity and updated the text accordingly (**L316-L329**).

Increased absolute error during the Wildfire period can be attributed to the higher observed absolute concentration of eBC which results in more potential for large absolute differences with respect to the reference measurement. Average normalized errors (NRMSE) for the three MA300s were calculated as 8.5% and 12.5% during the Regular and Wildfire period, respectively. Higher MAE and NRMSE during the wildfire period indicate that the MA300's errors have increases in both absolute and relative terms. This means, MA300's relative accuracy can deteriorate in a highly polluted environment. Since, the aerosol sampling process between the AE33 and MA300 can differ (filter mechanism, flowrate), it is possible that, MA300's measurement errors are associated with filter loading interactions. Differences in sampling flowrates (for MA300 150 mL min^{-1} and for AE33 5000 mL min^{-1}) can change differences in face velocity, and hence change the particle penetration depth into the filter (Moteki et al., 2010). Further, from the real-time estimates of eBC concentration, we calculated the device-specific filter loadings (in μg cm^{-2}) for the corresponding flow rate of measurement (see Table S6) and found that Wildfire periods were subjected to higher loadings (almost 3.5 times than Regular period). When filter loadings of MA300s and AE33 were normalized with their set flow rate, we found that MA300s were experiencing significantly higher filter loading (0.054 μg cm^{-2} mL^{-1}) than AE33 (0.006 μg cm^{-2} mL^{-1}) during the wildfire period. Higher filter loading and lower face velocity can lead to large measurement errors in MA300's results, particularly during wildfire period.

L280 not absolute, but variability in

Author's Response (MC): Mentioned variability in AE33 measured eBC concentration in the updated manuscript. (**L331**)

The linear fit results in Figure Figure 1(a) indicate that the variability in AE33 reported eBC concentrations were well captured by the MA300 units during the whole campaign.

L289: was the shown plot 1b for SD of average MA300 in each AE33 bin and S8 plot is for unaveraged SD for each bin? 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.

Author's Response (MC): This comment is linked to Major Comment #4 and has been addressed.

L309-310: "The absorption enhancement in the UV channel reflects the elevated contribution of organic compounds originating from wildfire smoke" -- only if it is normalized. Absolute increases in UV absorption can simply reflect absolute increases in BC.

Author's Response (MC): We thank the reviewer for pointing this out. We computed normalized b_{abs} values from the AE33 for all channels and for both the periods and showed the enhanced absorption during the Wildfire period in UV channel in Figure S8. In all the wavelengths, we performed Welch's t-test and found the enhancements during Wildfire period were statistically significant ($p < 0.05$). The updated discussion can be found in **L383-L387**.

To further check the differences in relative spectral absorption, for the Wildfire and Regular period, we compared the normalized b_{abs} from the AE33 (Figure S8). We used Welch's t-test on the regular and wildfire data separately for all the channels and found the observed spectral light absorption enhancement to be statistically significant ($p < 0.05$). The UV had the largest absorption enhancement during the wildfire period, reflecting the elevated contribution of organic compounds originating from wildfire smoke (Healy et al., 2019; Laing et al., 2020).

Figure S8 from Supplementary Information

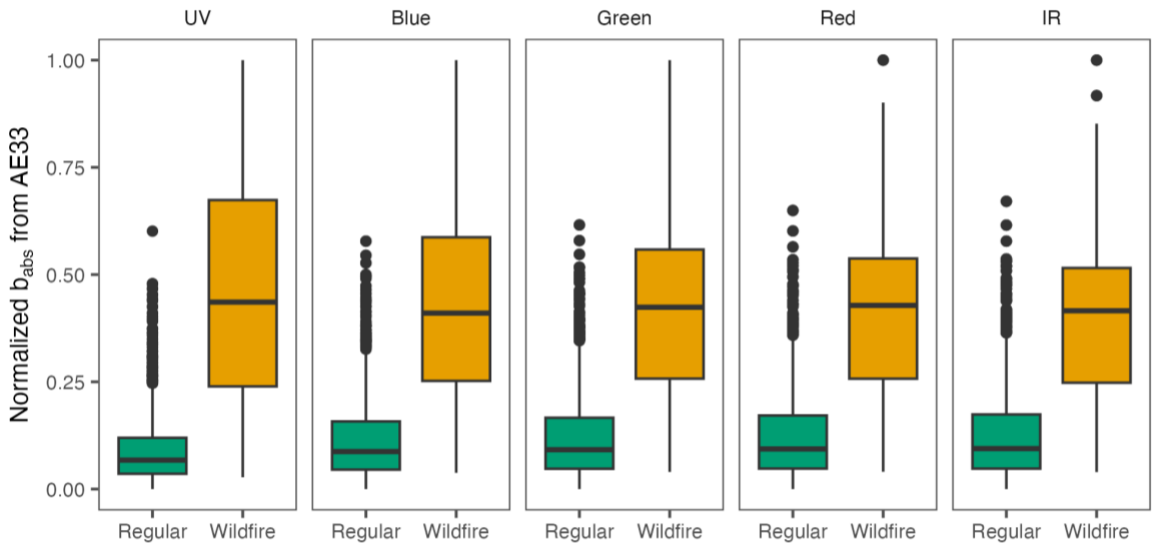

Figure S 2: Box plot of normalized b_{abs} from AE33 over the five channels categorized by the measurement period. Increased median values during the Wildfire period can be identified from the horizontal lines of each boxplot.

L319-320: However, the overestimation percentage decreases to 10% for lower wavelength light absorption (babs,UV) during the Reg period and interestingly changes to underestimation during the WF period by 9% -- has this been reported before? seems like a significant finding - if it is statistically significant.

Author's Response (MC): We thank the reviewer for highlighting the importance of our findings. We could find enough literature on this particular device. However, Li et al., (2021) showed that, for less absorbing BrC, MA200 (sister model of MA300) overestimates light absorption measurement. In the updated manuscript (**Section 3.2.2 L398-L404)** we have included results of statistical significance test (unpaired t-test) with a significance level of 0.05. t-test results revealed most of the findings are statistically significant ($p < 0.05$), except for $b_{abs,UV}$ during the Wildfire period ($p = 0.075$). However, we mention is as a borderline significance – i.e. underestimation is real but with slightly lower probability.

We find that the MA300-reported absolute measurements overestimated $b_{abs,IR}$ by 20% and 18% during the Regular and Wildfire periods, respectively, when compared against AE33. However, the overestimation percentage decreased to 10% for the UV channel ($b_{abs,UV}$) during the Regular period and interestingly switched to a 9% underestimation during the Wildfire period. Unpaired t-test results revealed most of these findings to be statistically significant ($p < 0.05$) except for $b_{abs,UV}$ during the Wildfire period ($p = 0.075$). The underestimation of $b_{abs,UV}$ during the Wildfire period appears to be borderline statistically significant with a p-value of 0.075, indicating that there is some evidence to suggest that the underestimation is real; however, the result does not reach the usual threshold for statistical significance.

Figure 3: binning these data like Fig 1b might be more robust? And maybe use different scales for different wavelengths, as the magnitudes vary.

Author's Response (MC): We thank the reviewer for this suggestion. However, in the updated manuscript, this figure has been removed. The unit-to-unit variability has been explained with the help of Figure 2,

instead, which shows the slope of linear fit of normalized b_{abs} values from MA300 units and from AE33. We determine the wavelength specific unit-to-unit variability comparing the coefficient of variation (CV) of slope values determined from Figure 2.

L331: just typing out "regular" and "wildfire" might improve readability.

Author's Response (MC): We thank the reviewer for pointing this out. We have updated from "Reg" to "Regular" and "WF" to "Wildfire" in the manuscript.

L336: Lower variability in the high pollutant period can further be explained by the errors estimated in the linear fits of MA300 vs AE33. -- this sentence is confusing.

Author's Response (MC): We thank the reviewer for mentioning this. This line was in reference to old Figure 3, which has now been removed and our analysis with normalized b_{abs} values have been presented.

L337: but the AE33 - descendant of AE31 - is your reference method! So this is not a good comparison. What was the reference instrument used by Muller et al.?

Author's Response (MC): In the updated manuscript (**L418-L422**), we removed the comparison with Muller et al, and included a new comparison study done on multiple units of AE33 by (Cuesta-Mosquera et al., 2021).

Previously, Cuesta-Mosquera et al. (2021) tested 23 units of AE33 in both laboratory and ambient settings, assessing the instrument's performance before and after maintenance. They found that, after maintenance, AE33 tends to slightly underestimate (slopes slightly reduced from 1) for ambient aerosol measurements at wavelengths 590, 660, and 880 nm, but any wavelength dependency of the unit-to-unit variability of AE33 was not reflected.

L354: How did you verify the flow rate? You earlier mentioned the Drinovec et al. algorithm includes the effect of flow rate, so should not these fluctuations be accounted for by that method?
Finally, the flow rate fluctuations should impact all wavelengths equally, and yet the two methods are about equally better than the raw data for the (noisier!) UV. I am not sure this interpretation is well supported by these results, though it is fair to say that empirically, the onboard correction is a better technique.

Author's Response (MC): This comment was captured in Major Comment #3 and was addressed above.

L360: "noiser"?

Author's Response (MC): This line has been removed in the updated manuscript.

L364: Stronger than in the AE33?? Where was this shown?

Author's Response (MC): We have performed the loading effect calculations for individual devices and included the results in supplementary information (Table S6). Our results show that MA300s were experiencing significantly higher filter loading (0.054 μg cm^{-2} mL^{-1}) than AE33 (0.006 μg cm^{-2} mL^{-1}) during the wildfire period. To improve clarity, we have cross-referenced Table S6 where we discuss this point. (**L506**)

In our study, all three MA300s were subjected to strong loading effects (Table S6) in addition to RH changes (45% to 95%) during the sampling periods.

L365: "wildfire smoke affected aerosols can become hygroscopic with aging as compared to freshly emitted soot particles from diesel vehicle emission. Hence, the combined effect of imposed noise due to filter loading correction, and highly loaded hygroscopic aerosol" -- very speculative

Author's Response (MC): In the updated manuscript, in Section 3.3 (**L442-L447**) we include discussion the presence of aged fossil fuel derived particles with the variations of AAE values. And then compare the understanding of aerosol hygroscopicity with respect to wildfire emissions in the discussion of Source apportionment results in Section 3.5 (**L5487-492**)

From Section 3.3

The peak of AE33's α distribution was found to be 1.69, similar to Zotter et al. (2017) for SA calculations. In Figure 3, we present each device's hourly mapped α values calculated by UV-IR and Blue-IR pairs. Day-time α values for both wavelength pairs during the Regular period were found to be lowest and closer to unity, representing aerosol sources from traffic sources Healy et al. (2019); Bernardoni et al. (2021). In contrast, nighttime α values were found to be highest during the regular period, which could be attributed to the local wood-burning sources (Healy et al., 2019).

From Section 3.5

Being a near-road emission measurement site, our measurements captured complex aerosol mixtures of various mixing states. During the regular period, local traffic during the daytime contributes to fresh BC-enriched aerosols, which can be hydrophobic in nature (Sarangi et al., 2019; Wang et al., 2020) and by night-time, these fresh BC-enriched aerosols can evolve by ageing and change their morphological and optical properties. In contrast, during the wildfire smoke-affected days, the measurement site experienced enhanced quantities of aged aerosols through long-range transport from the Pacific North-West. These claims align with our calculated α values, as shown in Figure 3.

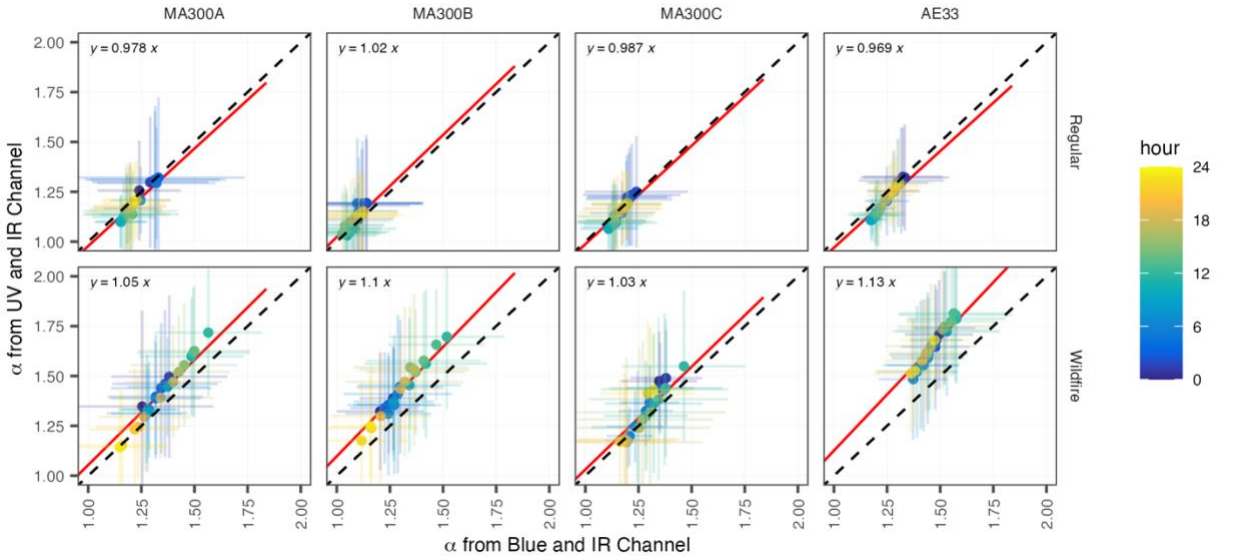

Figure 3. Ångström Exponent (α) (by the hour of the day) measured by different aethalometers during the Regular and Wildfire period. Average hourly AAE values derived from the Blue-IR wavelength pair (on x-axis) and the UV-IR wavelength pair (on y-axis) with error bars representing respective standard deviation. Red lines represent the linear relationship (forced through the origin), and the dashed line is 1:1.

L380: daily eBCf f concentrations were in the range of 0.6 – 1.9 µg m−3 during Reg period and in 1.0 – 3.8 µgm−3 during WF period. – lensing

Author's Response (MC): In the updated manuscript, we highlight the effect of lensing elsewhere (**L520-L526**).

During the Wildfire period, the enhanced eBC concentration was heavily influenced by biomass burning components, eBCbb; however, we hypothesize that the enhancement of eBCff may be due to two major factors. Firstly, there may have been a real increase in the number of heavy-duty vehicles during the early morning and evening hours that coincided with the Wildfire period, increasing the eBC emission from fossil fuel-based sources. Secondly, during wildfire smoke-affected days, aerosols can be a complex mixture of fresh and highly aged components, and the presence of BrC (Wang et al., 2019) and the lensing effect (Healy et al., 2015) have been found to impact the bulk aerosol light absorption measurements, particularly in the lower wavelengths.

L391: The *estimated" contribution. Why not run the source apportionment for five days on either side of the wildfire period?

Author's Response (MC): We thank the reviewer for this suggestion. We ran a source apportionment analysis as mentioned previously in the response to Major Comment #7.

L393: "be lower"

Author's Response (MC): We thank the reviewer for this correction. We have incorporated the change.

L394: In the case of biomass burning-based source contribution, eBCbb, 22% underestimation was identified during the WF period using Blue-IR based SA instead of UV-IR. Utilizing Blue- IR method based source apportionment resulted, 41% overestimation of eBCbb component -- these sentences are contradictory. I am having a hard time following this argument.

Author's Response (MC): We thank the reviewer for highlighting this. Our results identify that overestimation of eBC_{bb} during the Regular period using Blue-IR pair can be flawed (Figure S5 and Figure S11 (Panel A)). However, we choose to report our observations and discuss the failure of Source Apportionment results with low b_{abs} signals with the help of sensitivity analysis performed. These were discussed in **Section 3.5.1, L540-L548.**

It is important to note that derived absolute eBC_{ff} and eBC_{bb} components are dependent on the absolute b_{abs} inputs in the Aethalometer model and inherently, $b_{abs,UV}$ measurements are higher than $b_{abs,Blue}$. The effect of different input levels of b_{abs} and α pair on the Aethalometer model was explored through a sensitivity analysis and has been presented in Figures S10 and S11. For test purposes, we used AE33's mean b_{abs} concentrations for UV, Blue and IR channels from the Regular and Wildfire period with α_{bb} range 1.6 -- 3.0 and α_{ff} range 0.8 -- 1.5. Sensitivity analysis results show that apportioned $b_{abs,bb}$ and $b_{abs,ff}$ can often get negative values or even higher than the input b_{abs} values, which is an established flaw of Aethalometer model Grange et al. (2020). For a clean environment, lower b_{abs} input can cause large errors in the estimates of $b_{abs,bb}$ and $b_{abs,ff}$. Hence, we claim that source apportionment should not be conducted below the MDL of the black carbon concentration, which has been found as 0.21 μg m^{-3} for MA300.

Figure 6: if the Drinovec correction is not appropriate, why show it here?

Author's Response (MC): We thank the reviewer for this suggestion. In the revised manuscript only MA300's instrument data were included (Figure 6). Results from modified Drinovec correction have been moved to supplementary information (Figure S13).

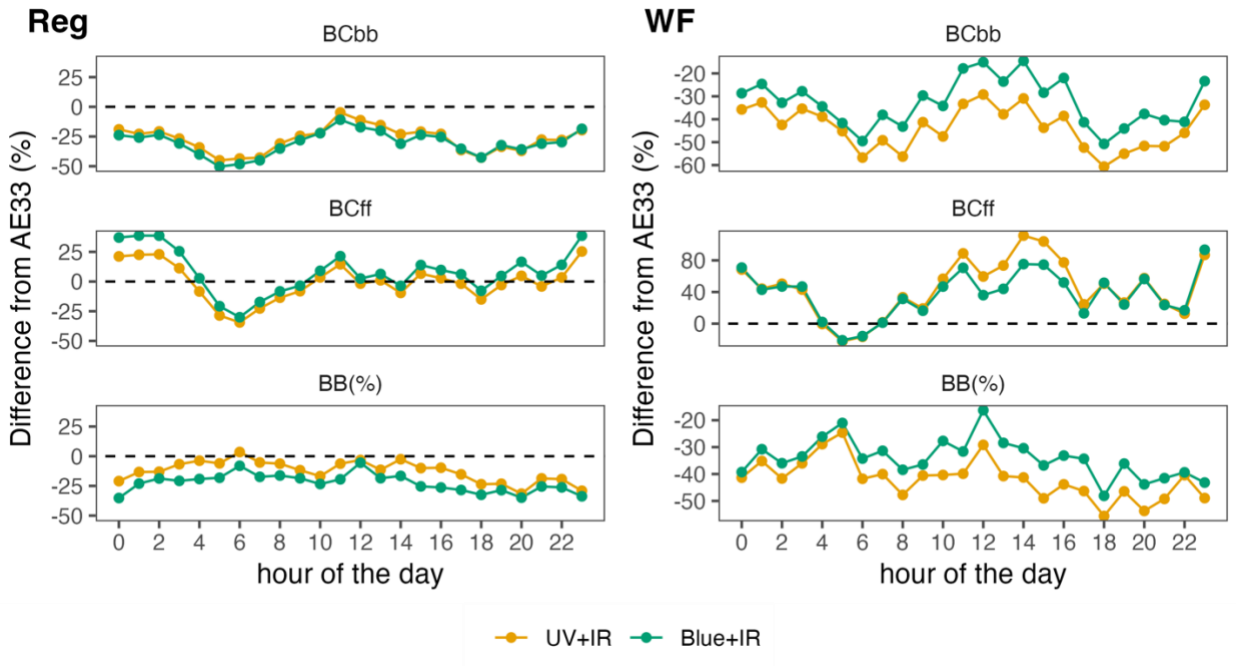

Figure S13 Percentage difference of hourly SA results in between MA300 and AE33. For MA300s, the average of onboard corrected data for the two periods (Regular and Wildfire). SA results from UV-IR and Blue-IR wavelength pair have been evaluated separately

L441: "lower"

Author's Response (MC): Updated in the revised manuscript.

L449: However, we find that MAE of eBC measurement from MA300 (compared to the AE33) can be in the range of 0.44 – 0.98 µgm−3 depending upon the measurement period. We observe larger MAE during high pollution conditions (e.g., WF period in this study). Based on these findings, caution may be required when MA300s are used to capture spatial or temporal differences in eBC below the 0.98 µg m−3 threshold. -- no, that's not the MDL.

Author's Response (MC): We agree with the reviewer that this is not the MDL. Here, we were emphasizing on capturing the spatial or temporal differences in absolute eBC measurements using MA300s. However, this cannot be treated as a limit of detection for MA300, but a concentration level to identify meaningful differences in measurement. We have now added a sentence that clarifies that the MDL should also be taken into account in experimental design. As discussed in Major Comment #8, we have now calculated the MDL following (Hyslop et al., 2022).

L454: I thought it was 1-minute data.

Author's Response (MC): We thank the reviewer for pointing this mistake out. Yes, it is 1 min data. Updated to 1 min in the revised manuscript.

L457: We were able to eliminate the drifted signals through post-processing the raw data using statistical outlier detection method. – this seems odd.

Author's Response (MC): Yes. This was highlighted in Section 2.4 (**L198-L200**) with the aim to raise awareness about the LED performance in MA300 units.

We observed that the MA300 sensor data was affected by both drift and post-filter- change transient effects. The drift in the photometer data was removed by calculating statistical outliers before calculating ATN and has been explained in detail in the Supporting Information (Section C).

L468: we hypothesize that an inconsistency is flow fluctuation in MA300, which is a key variable in deriving the real-time loading correction parameter. – a key roadblock.

Author's Response (MC): We thank the reviewer for this suggestion. We have updated the line (**L621**) accordingly.

We hypothesize that inconsistency in flow fluctuation in MA300 is a key roadblock in deriving the real-time loading correction parameter.

L470-476:
More like 21%,
these are reference methods for you, come on,
sounds like an LED problem, not a filter loading problem.

Author's Response (MC): We incorporated the suggestions made by the reviewer and updated the text (L630-L636)

Characterizing unit-to-unit variability can speak to instrument precision and may be particularly important for use cases where multiple MA300s are simultaneously deployed to measure a pollution event. We reported the precision of MA300 eBC in terms of unit-to-unit variability (based on normalized responses) as 5%. This value is slightly higher than that reported for other aethalometer models: 4.3% for AE31 (Müller et al., 2011), 0.5% for AE33 (Cuesta-Mosquera et al., 2021). This variability can increase to 21 % when absolute measurements are considered, reflecting individual instruments' sensitivity and noise characteristics affecting precision. For the multi-wavelength babs, the highest unit-to-unit variability was found in the UV channel (8%) with large instrumental noise, which is consistent with previous studies on multi-wavelength aethalometers (Cuesta-Mosquera et al., 2021; Müller et al., 2011). The unit-to-unit variability in the UV channel was not identified to be significantly varying with filter loading impacts and hence could be occurring due to problems associated with LED light sources or detectors.

L479: That should be semi-volatile organic compounds (because they condense on the filter material otherwise), but I think it is not a good hypothesis - it is possible there were denuder losses. Not going through Vecchi to verify, sorry.

Author's Response (MC): We have considered the presence of semi-volatile organics in the updated manuscript. (**L650**).

Light absorption measurements in the UV channel can also be sensitive to interference from the volatile to semi-volatile organic compounds on the filter tape Vecchi et al. (2014) or from the other light-absorbing non-BC combustion particles, which affect lower wavelengths more than higher ones.

Response to Reviewer #2 AMT-2022-278

The authors would first of all like to thank the reviewer for the insightful comments on the work we have submitted for publication, and the editor for the opportunity to improve the manuscript. Under each reviewer comment (in red) there is a summary of the response (black text), in addition to the text from the manuscript (in blue) that was modified, if applicable.

(1) Section 2.6: I have a fundamental issue with the methodology applied here. In particular, the assumption that biomass burning particles have a fixed absorption Angstrom exponent. This is known to be false. The AAE for biomass burning varies tremendously dependent on the fuel source and burn conditions. Any apportionment that assumes a fixed value for the AAE for biomass burning is, in my view, inherently flawed. I realize that the authors here are not developing this method and are taking these values and the approach from the literature. Nonetheless, in my view the method is poorly equipped for true source apportionment since one of the end members is unknown/non-constant. Related, it is not clear why the authors would state on L199 that the AAE for BC is 1 but then in their apportionment use a value of 0.9. It is established that BC has an AAE that is close to 1 but that is not necessarily exactly 1 depending on the particle size or coating state. It would be nice to see a recognition of these issues in the discussion of the apportionment method.

Author's Response (MC): We appreciate reviewer's comment on methodology discussed in section 2.6. The fact that 'AAE (α) for biomass burning cannot be fixed', is absolutely correct and many studies have recognized that fixing α values in any apportionment method will lead to uncertainty in estimates. As is, source apportionment using two-component Aethalometer model is widely accepted and used by regulatory bodies and scientific studies as a standard approach. Thus, we focus on presenting the MA300's performance for a standard method, and include more detailed sensitivity analysis around α assumptions and expanded discussion of caveats in the methodology (**Section 2.6, L228-250**). Given the wide usage of this method, we hope this expanded discussion will be useful for the MA300 user community.

One of the major use cases of multi-wavelength aethalometers is to perform source apportionment (SA) of eBC mass concentration. Source characterization of eBC is usually achieved by the widely used "Aethalometer model" developed by Sandradewi et al. (2008b). For the AE33, this SA model is built into the device software for estimation of the real-time contribution of biomass burning (BB%) to total eBC mass. The majority of aethalometer-based SA studies have used this method to characterize sources of eBC originating from fossil fuel or transportation sources (referred to as eBCff) and biomass burning or wood burning sources (referred to as eBC_{bb}) (Sandradewi et al., 2008b; Healy et al., 2017; Rajesh and Ramachandran, 2017; Zotter et al., 2017; Grange et al., 2020; Deng et al., 2020; Bernardoni et al., 2021). This model is based on the principle that eBC emitted from biomass burning (wood burning, wildfire) sources will tend to show enhanced absorption in the near-UV region of the light spectrum compared to fossil fuel (transportation, liquid fuel) sources. The components of eBC derived from the Aethalometer model strongly rely on a fixed pair of Absorption Ångström exponent inputs (α_{ff} and α_{bb}). However, in reality, fixing α_{ff} and α_{bb} does not capture the real-world variability in α from different fuel and burn conditions, leading to inaccurate estimates (Healy et al., 2017). Ambient aerosol is often mixed with volatile organic compounds, undergoes aging processes, and forms BrC components. As a result, bulk aerosol light absorption from highly mixed environments often contributes to lower wavelength light absorption (Saleh et al., 2015; Healy et al., 2015). Ideally, values of α_{ff} and α_{bb} should be derived from radiocarbon-based 14C analysis of the aerosol samples (Zotter et al., 2017; Sandradewi et al., 2008a). Due to the limitations of onsite α_{ff} and α_{bb} measurement, we use the values from Zotter et al. (2017) for two wavelength pairs (UV-IR and Blue-IR), which were verified using multiple instrument comparison and radiocarbon-based analysis. The constrained values of αff and αbb used in this study are as follows:

1. UV-IR with α pairs as 0.9 (α_{ff}) and 2.09 (α_{bb})
2. Blue-IR with α pairs as 0.9 (α_{ff}) and 1.75 (α_{bb})

To understand the impact of input parameters on the Aethalometer model, a sensitivity analysis on the choice of α has been performed for different combinations of b_{abs} inputs as experienced by AE33 during the Regular and Wildfire periods and discussed in Section 3.5.1.

Edited L199 for clarity. – we referred to the "theoretical" value of α for pure BC as 1 and have updated the text accordingly. (**L220**)

In theory, a pure BC aerosol particle is a strong absorber over the whole spectrum (near-UV to near-IR); hence it would show a weak spectral dependence (αBC = 1).

(2) L221: In my view, it is not correct to refer to "eBC_bb" and "eBC_ff" from the source apportionment. The method is not separating different BC sources, but is separating absorption and specifically separating BC (the absorption properties of which are largely independent of the source) and the contribution of brown carbon (BrC). As such one cannot apportion "eBC_bb." One can apportion "eBC" and "eBrC" if some assumed value for the MAC for BrC is determined. Or one can apportion the absorption. But one cannot apportion the absorption into BC from fossil fuel and biomass burning. As currently presented I do not think that this is accurate or correct.

Author's Response (MC): We thank the reviewer for highlighting this point. We have clarified this distinction in the manuscript, and emphasized that the labels eBC_{bb} and eBC_{ff} should more correctly be interpreted as eBC and eBrC. The outcome of the present source apportionment method (Sandradewi et al., 2008) assumes that these are characteristic of fossil fuel combustion and biomass burning, although they could be inferred as empirical sources. However, we have chosen to keep the labels eBC_{bb} and eBC_{ff} to facilitate comparison with past applications of this source apportionment model. In the updated manuscript (**Section 2.6, L250-L255**) we have acknowledged that this method first separates the light absorption contribution and then estimates the eBC_{bb} and eBC_{ff}.

Equations for the source apportionment (SA) can be found in Supplementary Information (Section D). It is important to note that the Aethalometer model operates on near-UV and near-IR babs values using equations S3-S7 and at first, it separates the contribution of light absorption to biomass burning source (babs,bb) and fossil fuel source (babs,ff). Next, eBC components (eBCbb and eBCff) were derived using Equations S8 and S9 and by dividing the babs components with instrument specific MAC(880 nm) values (Table S1). The rationale of using similar MAC values for determining both the eBC components has been discussed by Zotter et al. (2017).

(3) Fig. 2 vs. Fig. 1: The slopes in these two plots should be, largely, directly related. The main reasons that the slopes would change is because of (i) noise differences between wavelengths and (ii) the difference in the assumed MAC values. In particular, the slopes in Fig. 2 should be pretty similar to the slopes in Fig. 1 multiplied by the ratio of the MAC values between the MA300 units and the AE33 units. The authors might specifically acknowledge this relationship. They might consider normalizing the slopes to the value in the UV or IR, as the focus here seems to be on the wavelength-dependence of the slopes and not the absolute values (since, again, the absolute values just come directly from Fig. 1). Related, I personally did not find the discussion in Section 3.2.2 particularly valuable, as it is really just reiterating the results already shown for the eBC concentrations, but here in absorption space. I suggest removing this section or refocusing it on the AAE. Or at least merging it with section 3.2.3, as I see these as largely redundant discussions.

Author's Response (MC): We thank the reviewer for bringing up this discussion. In Fig 1., our goal was to present MA300's capability in resolving eBC concentration (which is based on the IR channel's response **only**) for the **combined** Regular and Wildfire days. This comparison could be suitable for users who might be utilizing only eBC concentrations from the MA300. In Fig 2., we focus on exploring MA300's additional capabilities, i.e. multi-wavelength b_{abs}, which could be helpful for advanced users who are interested in aerosol science. b_{abs} measurements in different wavelengths can vary based on the sample aerosol composition and are often valuable for characterizing/estimating aerosol's source. The reviewer rightly pointed out the relationship between the ratio of MAC and b_{abs} slope, which we have included in the discussion of the slopes in the updated manuscript in **L421-L423**. As suggested, we also normalized the b_{abs} values by rescaling (using each device's min-max range for normalization), converting multi-wavelength b_{abs} values onto a standard scale to compare the relative performance for individual periods and updated our results accordingly. The reviewer's suggestion for merging Section 3.2.2 and Section 3.2.3 was valuable, particularly after adopting data normalization. In the updated Section 3.2.2 we present the following.

1. Absolute measurements from MA300's b_{abs}
2. Assessment of MA300's unit-to-unit variability using normalized b_{abs} values

Additionally, **Section 3.3**. has been added to discuss the results from α . The updated Section **3.2.2** has been reframed as follows.

3.3.2 Comparison of MA300's multi-wavelength b_{abs} with AE33

To assess the performance of the MA300 b_{abs} measurements across all five wavelengths, we used unit-specific normalized measurements of each wavelength's b_{abs} for individual periods and compared them to the normalized wavelength-specific b_{abs} measurements from the AE33. As shown in Table 1, the absolute b_{abs} measurement range can significantly vary based on the measurement period or wavelength; a direct comparison of absolute values may not provide insights into measurement differences between wavelengths. We find that the MA300-reported absolute measurements overestimated $b_{abs,IR}$ by 20% and 18% during the Regular and Wildfire periods, respectively, when compared against AE33. However, the overestimation percentage decreased to 10% for the UV channel ($b_{abs,UV}$) during the Regular period and interestingly switched to a 9% underestimation during the Wildfire period. Unpaired t-test results revealed most of these findings to be statistically significant ($p < 0.05$) except for $b_{abs,UV}$ during the Wildfire period ($p = 0.075$). The underestimation of $b_{abs,UV}$ during the Wildfire period appears to be borderline statistically significant with a p-value of 0.075, indicating that there is some evidence to suggest that the underestimation is real; however, the result does not reach the usual threshold for statistical significance. The underestimation of $b_{abs,UV}$ during the Wildfire period can lead to erroneous source characterization results as UV and IR light absorption estimates are the primary inputs for the aethalometer source apportionment algorithm (Sandradewi et al., 2008b). When compared to the AE33, MA300 b_{abs} measurement errors were found to be higher during the Wildfire period across all the units (Table S3(b)). The mean absolute error for the $b_{abs,UV}$ ranged between 35.7–40.0 Mm^{-1} during the Wildfire period, which was three times the range observed during the Regular period. Normalized errors (NRMSE) were found as 14.4 – 18.4 % and 7.9–8.4 % during the Wildfire and Regular periods, respectively. The lowest absolute errors were found in the $b_{abs,IR}$ measurements (3.4–5.4 Mm^{-1} during the Regular and 7.0–13.2 Mm^{-1} during the Wildfire period).

As shown in Figure 2, the linear fit of individual MA300 units vs AE33's hourly averaged normalized multi-wavelength b_{abs} revealed significant variability within MA300 units. It is important to note that, in Figure 1, we present the linear performance of MA300 units (with respect to AE33) in measuring eBC for the whole campaign, which corresponds to the IR channel measurement only. Equation 5 shows that the linear relationship of MA300's $b_{abs,IR}$ and AE33's $b_{abs,IR}$ will be directly related to eBC measurements multiplied by the ratio of MAC_{IR} values between MA300 and AE33. During the Regular period, the slope ranged between 0.80 and 0.99, while during the Wildfire period, it ranged between 0.71 and 1.16 (Figure 2). Previously, Cuesta-Mosquera et al. (2021) tested 23 units of AE33 in both laboratory and ambient settings, assessing the instrument's performance before and after maintenance. They found that, after

maintenance, AE33 tends to slightly underestimate (slopes slightly reduced from 1) for ambient aerosol measurements at wavelengths 590, 660, and 880 nm, but any wavelength dependency of the unit-to-unit variability of AE33 was not reflected. Here, to assess the unit-to-unit variability of MA300s across the five channels, we used the coefficient of variation (CV) of the normalized slopes from the three units (Figure 2). The highest amount of unit-to-unit variability was in $b_{abs,UV}$ (CV ≈ 8 %). Underestimation of $b_{abs,UV}$ combined with the observed largest variability in normalized measurements will impact the SA results, particularly during the Wildfire period. However, the variability in the Blue channel was found to be low (CV ≈ 4%) and slope values were much closer to 1 during the Wildfire period, which makes it a potential near-UV wavelength of choice for the SA studies using the MA300.

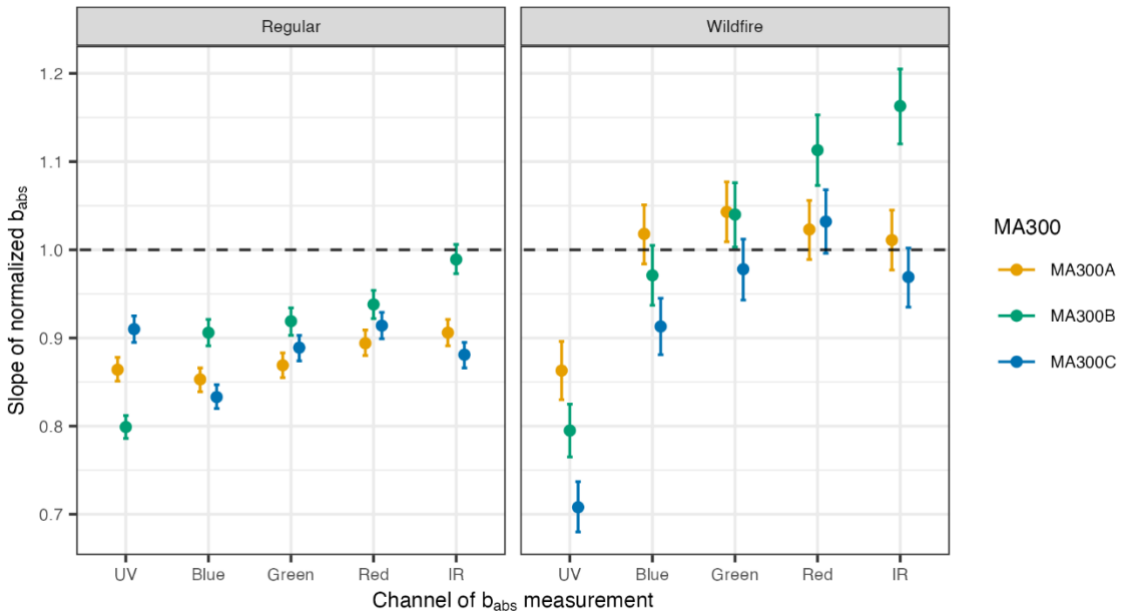

Figure 2 Slope of regression fit from the linear fit of normalized multi-wavelength b_{abs} values of MA300s vs AE33 during the Regular and Wildfire period. The dashed horizontal line is a slope of 1. The error bars show the 95% confidence interval of the linear fit.

(4) L358: The authors state that "in Section 3.3, we identify that the UV channel absorption measurements, babs,UV , are subject to higher error than measurements at higher wavelengths." This is not robustly true. In particular, it is evident in Fig. 2b that the slope relative to the reference instrument is closer to unity for the UV channel compared to other wavelengths. So I do not understand what the authors mean here when they say that the UV has higher error. Higher error relative to what? What is the metric? Also, in Fig. 4 the UV channel has a slope closer to unity than the IR channel and has a higher R2 value.

Author's Response (MC): Based on the restructured manuscript, particularly in **Section 3.2.2**, we discuss the claim we made regarding the UV channel in detail. We agree that our earlier claims were not totally clear from our discussions. In the revised manuscript, we present absolute and normalized errors for all five measurement wavelengths (**L416-L419**).

The mean absolute error for the $b_{abs,UV}$ ranged between 35.7–40.0 Mm^{-1} during the Wildfire period, which was three times the range observed during the Regular period. Normalized errors (NRMSE) were found as 14.4 – 18.4 % and 7.9– 8.4 % during the Wildfire and Regular periods, respectively. The lowest absolute

errors were found in the $b_{abs,IR}$ measurements (3.4–5.4 Mm^{-1} during the Regular and 7.0–13.2 Mm^{-1} during the Wildfire period).

Absolute errors were quantified using Mean Absolute Error (MAE in Mm^{-1}), and Normalized errors (NRMSE in %) were recalculated following min-max criteria and have been explained in detail in the supplementary information (**Section E**).

Section E from Supplementary Information

$$MAE_{ij} = \frac{1}{n_j}\sum\left|P_{AE33,j} - P_{MA300,i,j}\right|$$

$$RMSE_{ij} = \sqrt{\frac{\sum_1^n\left(P_{AE33,j} - P_{MA300,i.j}\right)^2}{n_j}}$$

$$NRMSE_{ij} = \sqrt{\frac{\sum_1^n\left(N_{AE33,j} - N_{MA300,i.j}\right)^2}{n_j}}$$

$$N_{normalized} = \frac{N - N_{min}}{N_{max} - N_{min}}$$

*P_{AE33} and P_{MA300} represents absolute measurements of parameters (eg. eBC) from AE33 and MA300 respectively for i^{th} MA300 device at j^{th} period (Regular or Wildfire). N_{AE33} and N_{MA300} represents the normalized values for individual instruments. For any parameter N, normalized quantity ($N_{normalized}$) was derived by the respective range, i.e. the difference in the maximum (N_{max}) and minimum (N_{min}) values.

Additionally, the slope of the linear model of normalized b_{abs} values (MA300 vs AE33) provided a basis to compare the relative performance across devices for each wavelength and measurement period. Our new results reflect the higher absolute errors in $b_{abs,UV}$, irrespective of MA300 unit or period. During the Wildfire period, we find that normalized $b_{abs,UV}$ had the lowest slope (average slope = 0.78), indicating underestimation by approximately 20%, and the highest amount of unit-to-unit variability (CV = 8%). Our latest results confirm that the UV channel absorption coefficient measured by MA300 performed the poorest compared to other channels (**Figure 2,** see page 4 of this document).

(5) L384: The authors need to demonstrate that these statements and conclusions regarding biomass burning contributions during the Reg period are robust to a rigorous assessment of measurement uncertainties and to uncertainties inherent in the assumptions made during the apportionment. What if the BC AAE were 1.2 rather than 1.0, for example? This is within the realm of reasonableness. How would the conclusions here change? This points to a bigger need for a more rigorous assessment of uncertainties in the context of the apportionment that goes beyond noise characterization which dominates the manuscript.

Author's Response (MC): We thank the reviewer for highlighting the point on the robustness of the Aethalometer model. Based on the updated analysis on derived AAE (α) (**Section 3.3**), we identified that α values can vary significantly during the Wildfire period. The multimodal distribution of α during the wildfire period (**Figure S9**) confirmed the impact of biomass burning on the light absorption measurements. The uncertainty arising from using a fixed pair of α values (α_{bb} and α_{ff}) has been previously discussed in several studies (Healy et al., 2019; Zotter et al., 2017) and we have included these references where we discuss the method (**Section 2.6**) of source apportionment. However, to identify the variation of source

apportionment results on the choice of α pairs, we performed a sensitivity analysis by varying the values of α_{ff} and α_{bb} for four test combinations b_{abs} values: AE33 reported average b_{abs} values from UV-IR and Blue-IR pair for the Wildfire and Regular periods. The results of the source apportionment were discussed in **Section 3.5.1,** and the outcome of the sensitivity analysis has been included in supplementary information (**Figure S10** and **S11**). We chose to explain the impact of source apportionment result by considering average b_{abs} as reported by AE33 (from Table 1) for both the periods. We find that apportioned results were also impacted by the input b_{abs} values. For low input of b_{abs} values (b_{absUV} = 29 Mm^{-1} and b_{absIR} = 10 Mm^{-1}), we find the apportioned $b_{abs,bb}$ values to be mostly negative and not sensitive to the changes in α_{bb}. On the other hand, $b_{abs,ff}$ gets overestimated beyond the input values. Source apportionment done on a clean environment can have unrealistic estimates of $b_{abs,bb}$ and $b_{abs,ff}$. Hence, we claim that source apportionment should not be conducted below the MDL of the measurement device. The updated text as follows.

It is important to note that derived absolute eBC_{ff} and eBC_{bb} components are dependent on the absolute b_{abs} inputs in the Aethalometer model and inherently, $b_{abs,UV}$ measurements are higher than $b_{abs,Blue}$. The effect of different input levels of b_{abs} and α pair on the Aethalometer model was explored through a sensitivity analysis and has been presented in Figures S10 and S11. For test purposes, we used AE33's mean b_{abs} concentrations for UV, Blue and IR channels from the Regular and Wildfire period with α_{bb} range 1.6 -- 3.0 and α_{ff} range 0.8 -- 1.5. Sensitivity analysis results show that apportioned $b_{abs,bb}$ and $b_{abs,ff}$ can often get negative values or even higher than the input b_{abs} values, which is an established flaw of Aethalometer model Grange et al. (2020). For a clean environment, lower b_{abs} input can cause large errors in the estimates of $b_{abs,bb}$ and $b_{abs,ff}$. Hence, we claim that source apportionment should not be conducted below the MDL of the black carbon concentration, which has been found as 0.21 μg m^{-3} for MA300.

Figure S10 and S11 included in the supplementary section.

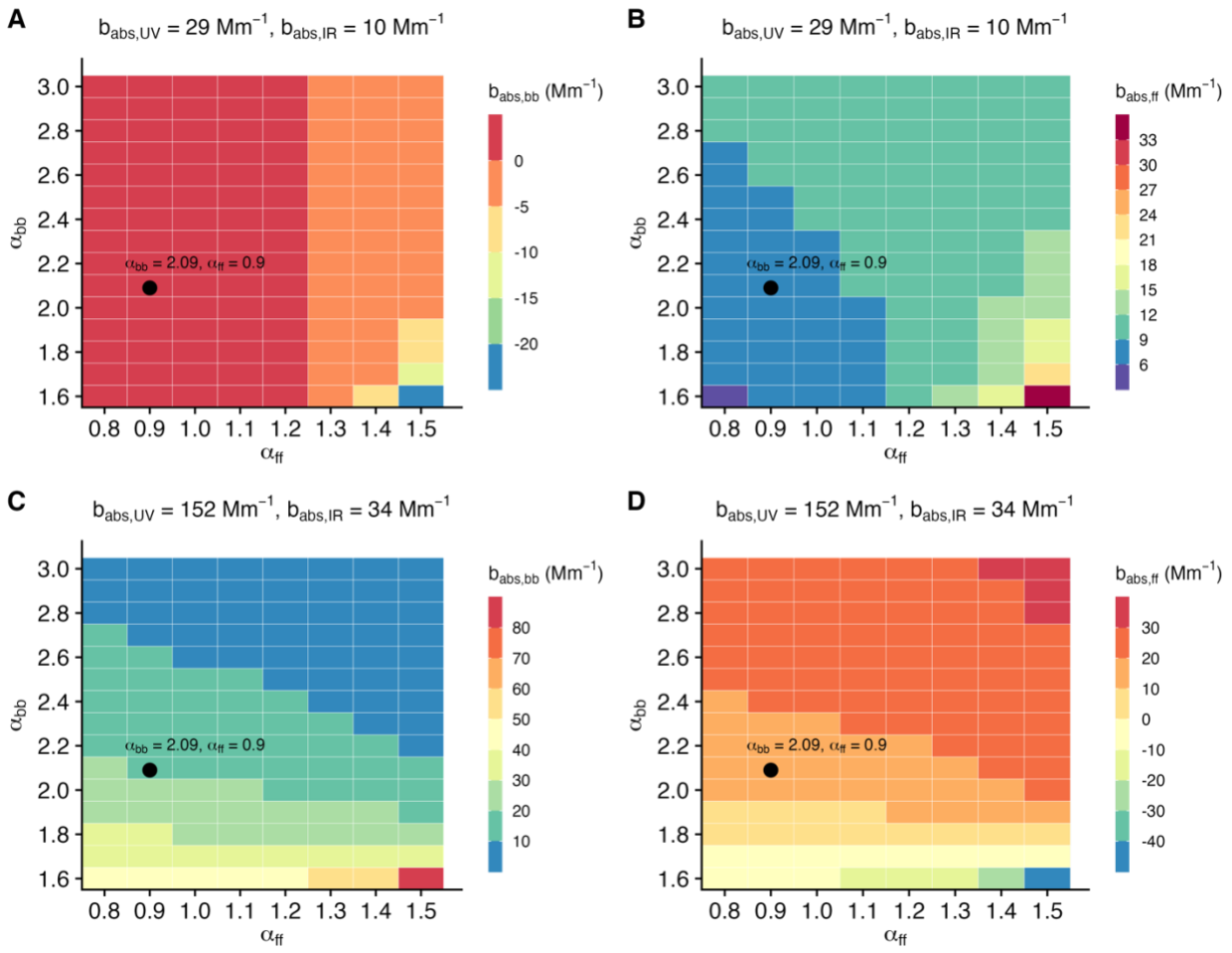

Figure S 10: Sensitivity analysis on UV-IR wavelength pair done with α_{ff} range of 0.8-1.5 and α_{bb} range of 1.6-3.0. Panel A and B represents the $b_{abs,bb}$ and $b_{abs,ff}$ components derived from the AE33's average b_{abs} values for the regular period. Panel C and D represents the same but for the AE33's average b_{abs} values for the wildfire period. Black dot represent the α pair considered in the source apportionment method.

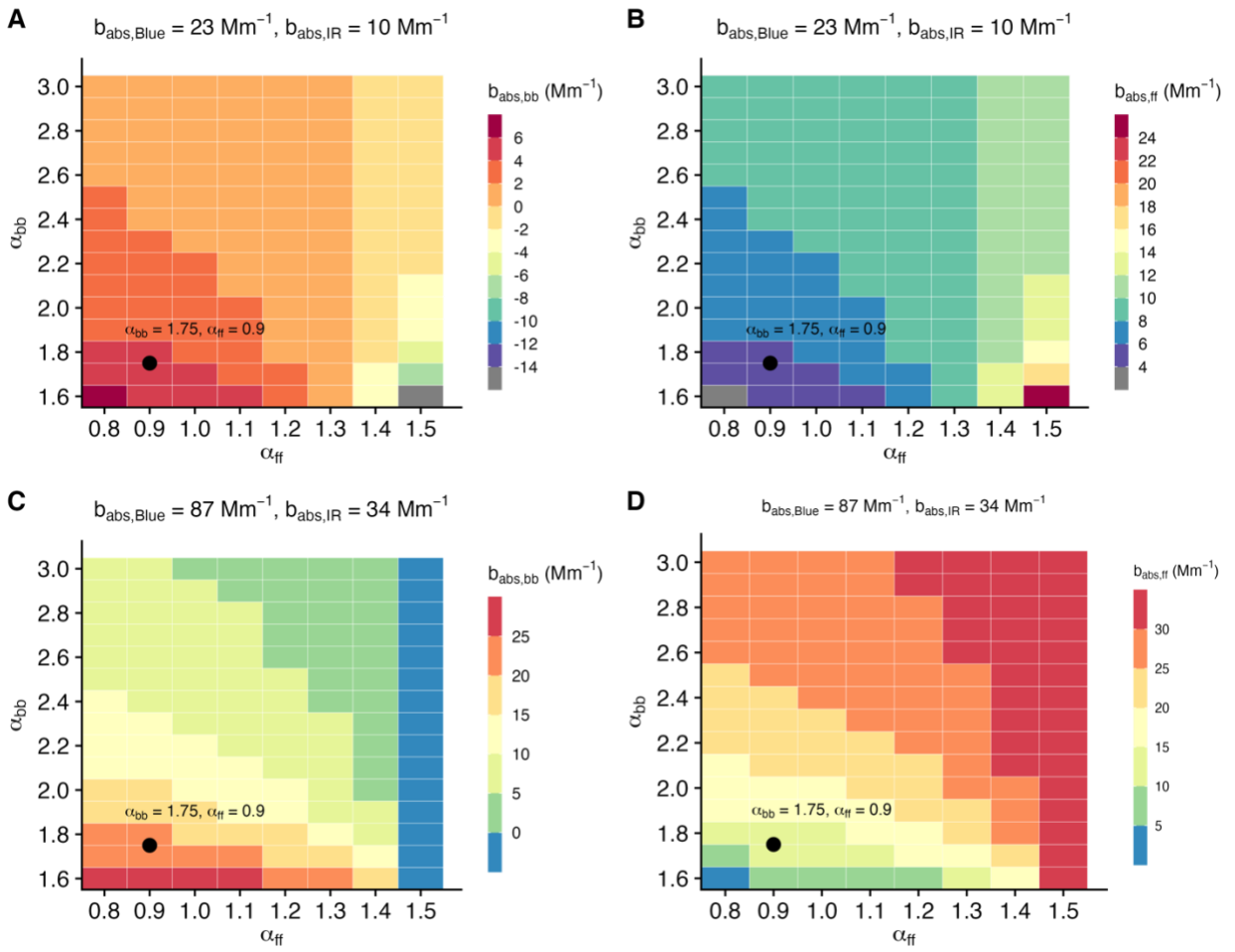

Figure S11: Sensitivity analysis on Blue-IR wavelength pair done with α_{ff} range of 0.8-1.5 and α_{bb} range of 1.6-3.0. Panel A and B represents the $b_{abs,bb}$ and $b_{abs,ff}$ components derived from the AE33's average b_{abs} values for the regular period. Panel C and D represents the same but for the AE33's average b_{abs} values for the wildfire period. Black dot represent the α pair considered in the source apportionment method.

(6) Section 3.4.2: I strongly encourage the authors to reframe this entire discussion to focus on the AAE values and how they compare between the instruments. Everything in the apportionment comes back to the AAE. The apportionment is just a mathematical transformation that then brings in the assumptions regarding AAE values for different particle types. A focus on the inherent measurement, the AAE, rather than subsequently derived properties would, in my view, greatly strengthen the manuscript. To me, I see the apportionment discussion a distraction from the core assessment of instrumental performance.

Author's Response (MC): We appreciate the reviewer's comment on restructuring the manuscript's narrative. We have included a detailed analysis of AAE (α) values derived from MA300 units and AE33 and included in updated **Section 3.3**. First, we present a distribution of hourly averaged α values from all four devices separately for two periods in Figure S9. Then we present the relationship of α values with respect to diurnal variations, as sources of BC could change with respect to time of the day. We also found large variability in MA300's α values. The discussed topic on α and it's variability has been further used in explaining the Source Apportionment results in Section 3.5.1. The updated text and figures are as follows.

3.3 Absorption Ångström Exponent (α)

The strength of spectral light absorption of aerosols is considered one of the most important parameters in understanding an aerosol's impact on earth's radiation balance and can be derived from aethalometer measurements (Zotter et al., 2017; Bernardoni et al., 2021). In addition, α values are used for determining fossil fuel and biomass burning source contributions in eBC from the Aethalometer Source Apportionment Model (Sandradewi et al., 2008a; Healy et al., 2019). In this section, the exponent of a power-law fit (Equation 6), Absorption Ångström Exponent (α), was derived for two wavelength pairs - UV (λ = 375nm) & IR (λ = 880nm) and Blue (λ = 470nm) & IR (λ = 880nm) using the hourly averaged babs values. In the literature, α has been calculated by several combinations of wavelengths. Grange et al. (2020), reported α by curve fitting all absorption wavelengths, as shown in Figure S3 (for AE33) and in Figure S4 (MA300 units). However, most studies (Garg et al., 2016; Zotter et al., 2017; Healy et al., 2019; Rajesh et al., 2021) have focused on reporting α by choosing two extreme wavelength pairs on the measurement spectrum, as we choose here. The distribution of α values (based on a UV-IR pairing) from the three MA300 units and the AE33 for the Wildfire and Regular period are shown in in Figure S9. During the Regular period, we observed a unimodal distribution with a α peak close to 1.13 (from AE33). This suggests that Regular periods were mostly experiencing a strong single source of aerosol from the nearby traffic emissions. For MA300 measurements, we find the distribution to be wider than AE33. During the Wildfire period, the distributions broadened and were multi-modal. The peak of AE33's α distribution was found to be 1.69, which is very close to that previously recommended as an optimal α_{bb} value by Zotter et al. (2017) for SA calculations. In Figure 3, we show each device's hourly mapped α values estimated by UV-IR and Blue-IR pairs. Daytime α values for both wavelength pairs during the Regular period were found to be lowest and closer to unity, representing aerosol sources from traffic sources (Healy et al. (2019); Bernardoni et al. (2021)). In contrast, nighttime α values were found to be highest during the regular period, which could be attributed to local wood-burning sources (Healy et al., 2019). During the Wildfire affected days, the scenario becomes the opposite; the highest α values were during daytime and the lowest during nighttime. We speculate that, in our measurement site, aerosol light interaction can change significantly by time-of-day as dominant sources and additional oxidation processes fluctuate. The error bars in the α measurement (Figure 3) were consistently higher in MA300-based measurement as compared to the AE33, which we believe to be contributed by the errors from b_{abs} measurements. Additionally, we find the Blue-IR based α values were consistently lower during the Wildfire period. Slope of fit line (in Figure 3) were 1.05, 1.1, 1.03 and 1.13 for MA300A, MA300B, MA300C, and AE33 respectively, indicating lowering of α values. This is in line with the differences observed in b_{abs} measurement in UV and Blue channel during the wildfire period.

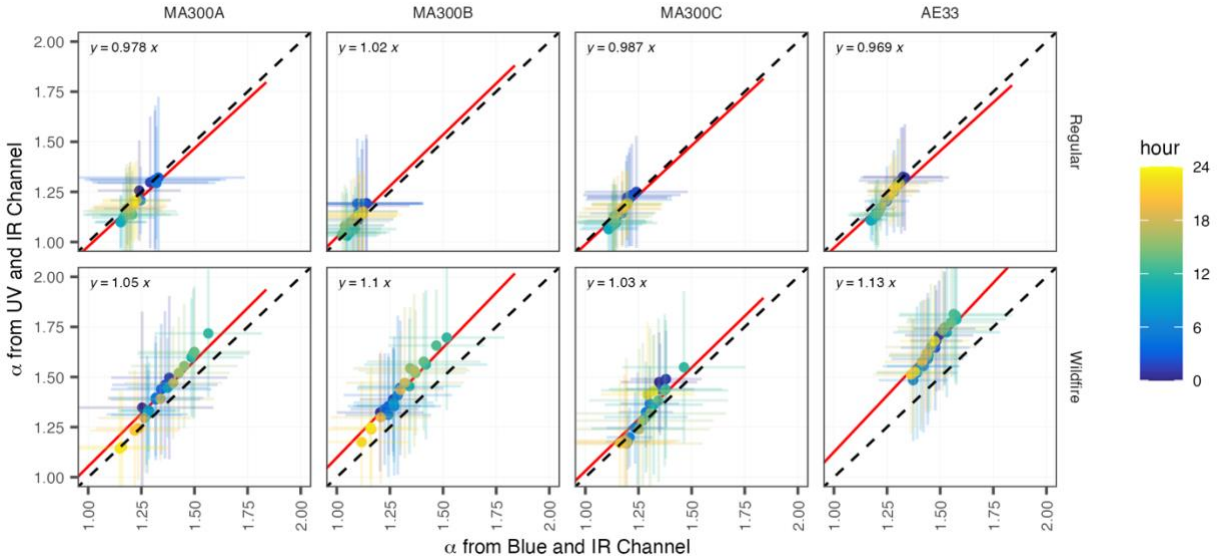

Figure 3: Ångström Exponent (α) (by the hour of the day) measured by different aethalometers during the Regular and Wildfire period. Average hourly AAE values derived from the Blue-IR wavelength pair (on x-axis) and the UV-IR wavelength pair (on y-axis) with error bars representing respective standard deviation. Red lines represent the linear relationship (forced through the origin), and the dashed line is 1:1.

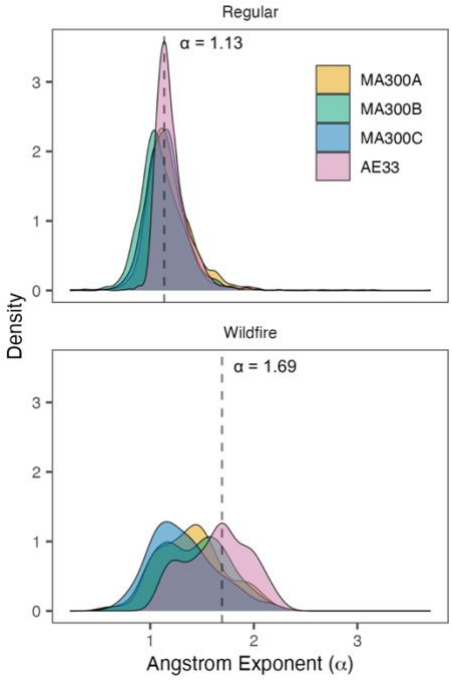

Figure S9: Distribution of Ångström Exponent (α) from the four aethalometers during regular (top) and wildfire (bottom) period. The α values have been calculated from hourly estimates of UV and IR b_{abs} values. Vertical dashed line represents the peak of the distributions from AE33 measurements for individual period.

Other Comments:

L40: Rather than making the distinction between "traditional" and new measurement methods, I suggest that the authors simply say that there are three main methods for characterization. The cited paper regarding LII is from 2006, and thus it has been around long enough and used by enough people that it could easily be considered part of the "traditional" canon.

Author's Response (MC): We appreciate the reviewer's point and have updated the main text (**L40-L46**) in explaining the different black carbon measurement processes.

There are three main processes used to quantify mass concentration of BC: (1) as elemental carbon (EC) mass concentration derived from thermal-optical analysis of aerosol deposited on filters (Bauer et al., 2009)(e.g. Sunset thermal-optical OC-EC analyzer), (2) as equivalent black carbon (eBC) measurements derived from light absorption of aerosol collected on a filter (Hansen et al., 1984) material (e.g. Aethalometer, Multi-Angle Absorption Photometer) or from photo-acoustic measurements (e.g. the Photo Acoustic Soot Spectrometer), and (3) the laser-induced incandescence (LII) technique has been used to measure refractory BC (rBC) concentration after the development of the Single Particle Soot Photometer (SP2) instrument (Schwarz et al., 2006).

L52: I suggest stating that at 880 nm light absorption is predominantly due to BC rather than it being "only" due to BC, as the former is more formally correct.

Author's Response (MC): We agree with the reviewer's suggestion and have updated the text in **L53-L54.**

The aethalometer reported BC mass concentrations are derived from the light absorption measurements at an infrared (IR) wavelength (880 nm), as light absorption at 880 nm has been identified as being predominantly due to BC (Hansen et al., 1984).

L70: The authors might note that, if it is assumed that the scattering correction factor is wavelength independent (which is not a given) that the accuracy of the C parameter impacts primarily the determination of the absolute absorption and derived concentrations and not the apportionment.

Author's Response (MC): The scattering correction factor (C) is typically considered as fixed, as provided by the manufacturer based on the filter type used. Previous studies have pointed out that the major factor that influences C is the filter material (Bernardoni et al., 2021; Drinovec et al., 2015; Segura et al., 2014). However, as the reviewer pointed out correctly, C is not constant and has been shown to vary by the wavelength, and even time of day. Having a fixed C value in the aethalometer algorithm can therefore also influence source apportionment results (Bernardoni et al., 2021). In light of the reviewer's suggestion, we have acknowledged these points in the updated manuscript **(L77-L86)**. However, for the calculations we used the manufacture recommended fixed C values.

In the aethalometer's onboard correction algorithm, the manufacturer includes a standard value of C for all wavelengths depending on the type of filter installed, as C is found to be strongly dependent on the filter material used (e.g. C_{quartz} = 2.14 and C_{TFE} = 1.57) (Segura et al., 2014; Drinovec et al., 2015). However, Bernardoni et al. (2021) and Segura et al. (2014) estimated comparatively higher values of C with wavelength dependency by comparing different field and laboratory-based instrumental measurements. Additionally, wavelength-dependent C values were shown to depend on the aerosol's single scattering albedo (SSA), which can directly impact the light absorption estimates (Yus-Díez et al., 2021). Bernardoni et al. (2021) also identified limitations of using a fixed C value in the aethalometer source apportionment

results. However, deriving optimized C values is challenging, requires additional monitoring, and may not always be transferable as aerosol properties and filter-matrix interactions with light scattering can change by instruments operated in different regions.

L87: All AE33's have the same properties and thus "typically" can be removed.

Author's Response (MC): We thank the reviewer for pointing this out. We removed "typically" from the main text (**L100-L102**).

Although advanced aethalometers like the AE33 are widely used, they may not be appropriate in certain environments where portability and battery-powered operation are essential since these instruments are expensive, bulky, and require external pumps or an external power supply to operate.

L100: Further details regarding the nature of the "non-linearity arising from flow" would be welcome. As is, it is difficult for a reader to understand what is meant by this.

Author's Response (MC): We thank the reviewer for pointing this out. We now more clearly explain the term "non-linearity arising from flow" in the main text (**L109-L112**).

Currently, the MA300's onboard correction algorithm uses a linear loading correction method (Virkkula et al., 2007) applied to simultaneous dual filter spot (dual-spot) measurements. In contrast, the AE33 onboard algorithm uses a real-time dual-spot correction, that includes adjustments for real-time variations in flow rate (Drinovec et al., 2015), which can lead to non-linearities in the relationship between ATN and BC surface loading.

L154: What is meant by the "(0-120)?" It is unclear.

Author's Response (MC): We thank the reviewer for catching our omission. "0-120" is the operating ATN range for model AE33. To add clarity in the text, we removed "(0-120)" and added a line explaining the aethalometer ATN range (**L168-L170**).

Fresh filter spots will have an ATN value of 0, and continuous aerosol deposition on the filter spots will gradually increase ATN to a user-defined threshold value (typically 120 for AE33 and 100 for MA300) before moving to the next set of fresh filter spots.

L156: Convention is generally that sigma means a molecular cross section and that the MAC is simply referred to as MAC. I suggest adopting this convention.

Author's Response (MC): As per the suggestion, we removed sigma in the text and use "MAC" uniformly to refer to mass absorption cross-section. This has been adopted throughout the manuscript.

L160: This should refer to Equation 2, not 3.

Author's Response (MC): We thank the reviewer for catching this mistake. We have updated the equation number from 3 to 2.

Eqn. 2: It should be given in the main text. Also, what does the periods mean? Are they meant to indicate multiplication?

Author's Response (MC): We thank the reviewer for this suggestion. We have updated the text (**L178-L184**) by adding the nomenclatures of the equation and replaced periods with "×" to better indicate multiplication. Here is the modified text in the manuscript.

In equation 2, k and C refer to the loading and multiple scattering correction factors, respectively. In this work, a TFE coated glass fiber filter (model M8060) was used in the AE33; hence we used the manufacturer recommended scattering correction factor (C) of 1.39. AE33's aerosol-loaded filter spots comprised an area (A) of 0.785 cm2 . ΔATN1 (λ) refers to the change in ATN at loading spot 1 within the time change of Δt(= 1min). The recommended filter lateral leakage factor (ξ) was set to 0.01, representing 1% leakage of the tape. The wavelength-specific loading correction factor (k) is calculated by solving a non- linear equation consisting of flow (F) and attenuation measurements (ATN) at each time step from both filter spots (equation 4).

L161: Does the scattering correction factor, C, depend on wavelength? If not, why not, as scattering is a fundamentally wavelength-dependent property?

Author's Response (MC): As mentioned earlier in the text (**L80-L82**), the scattering correction factor, C has been identified to depend on wavelength (Bernardoni et al., 2021; Segura et al., 2014). These studies have acknowledged that C can vary based on location, time of day and even by aethalometer model and they are not yet adopted universally in the AE33's correction algorithm. Since our study is limited to simultaneous measurement to evaluate wavelength dependency during the study period, we used the AE33's manufacturer set constant C value (1.39) in this study. We made this clear in the updated manuscript (**L185-L187**).

In this work, a TFE coated glass fiber filter (model M8060) was used in the AE33; hence we used the manufacturer recommended scattering correction factor (C) of 1.39.

L170: The authors might note that the MAC values for both instruments are much higher than known (realistic) MAC values for BC, and thus are best considered as effective values rather than actual values. I think this is an underappreciated aspect of these instruments: the absorption measured is not the true absorption because if it were then realistic MAC values would be used. In many ways, the absorption coefficients determined here should be called "effective absorption coefficients" in the same way that the BC is referred to as "effective BC."

Author's Response (MC): We acknowledge that having a fixed set of manufacturer provided MAC values is a source of great uncertainty (Healy et al., 2017) in real estimates of BC. However, in view of the reader's suitability, we disagree with author's suggestion to changing aethalometer derived absorption coefficient to "effective absorption coefficient" to keep the nomenclature aligned with the literature. In the updated text (**L189-L192**), we now inform the readers about the Aethalometer's fixed MAC values and mentioned the uncertainty.

These MAC values may not represent the realistic MAC values of real-time sampled aerosol as MAC values can change by aerosol composition, monitoring site or even by measurement instrument (Healy et al., 2017). Hence, fixed MAC values used in aethalometers can contribute to uncertainty in reported eBC concentration.

Eqn. 6 should technically use a proportional too symbol, not an approximately symbol.

Author's Response (MC): We thank the reviewer for pointing this out. We have updated the equation.

L230: I suggest that the authors retain these high PM event days as they provide an additional test of the comparability of the two instruments, which is the overall aim of this paper. The exclusion of these leaves the reader wondering if the relationship between the two completely broke down during this period, leading to questions regarding the overall performance. Including this period as an additional case study would strengthen things, in my opinion.

Author's Response (MC): We thank the reviewer for this suggestion. We did not include these high PM event days in the main analysis as it will significantly change the conclusion and expand the eBC ranges during the Regular period. Aligning with reviewer's suggestion, we performed a case study specifically on these days and added our findings in the supplementary information (Section F). We also highlighted the importance/benefits of flagging the data points while performing long-term measurements. In the updated manuscript we provided detailed description of this point (**L264-L268**).

During the measurement period, we also identified unusually elevated concentrations of $PM_{2.5}$ for three days (October 31 - November 2, 2020) during the Regular period, which was attributed to local fireworks from Halloween celebrations. These three days of data were removed from the main analysis to increase the consistency of the data. Data from these days have been separately analyzed as a case study to check the performance of MA300 in high PM events and provided in the supplementary information (Section F).

Section F: A case study on MA300's eBC measurement during high PM event

During the Regular days, a specific high PM event was identified in this study. These days were October 31st to November 2nd and were impacted by the local fireworks from Halloween celebrations. We removed these data points from the main analysis to improve the data consistency in the Regular period. We find good linear association ($R^2 > 0.9$) across the MA300 units, when compared against the reference device AE33. Slope from the linear fit of these three MA300 devices were slightly higher than reported values in the main text (Figure 1(a)), however the patterns remain the same. The mean eBC concentration were found as 3.6, 3.0 and 3.9 $\mu g/m^3$ from MA300A, MA300B and MA300C respectively. Mean eBC concentration reported by the reference AE33 was 3.55 $\mu g/m^3$. These mean eBC concentrations were much higher than the rest of the days during the Regular period and highlights the importance of marking associated events while collecting data which could potentially impose bias in the analysis.

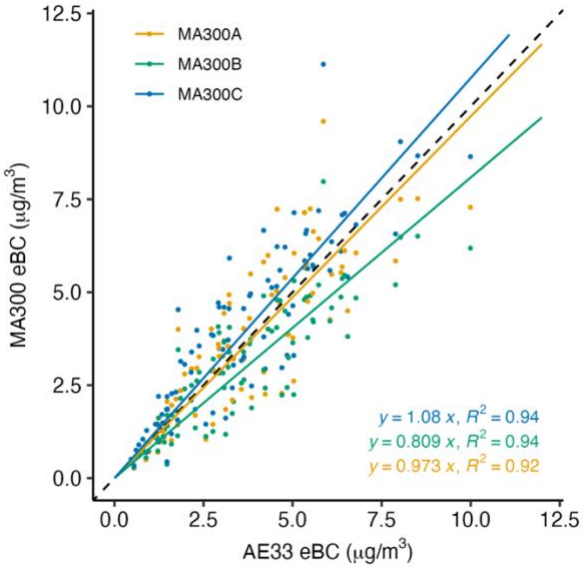

Figure S 14: Scatter plot of eBC mass concentration for individual MA300 units A,B and C vs AE33 during the high PM event.. The dashed line represents the 1:1 line, and solid colors are the regression fit lines for the individual MA300 units

L233: I strongly suggest referring to this as "relative accuracy" rather than "accuracy." This assumes that the AE33 is accurate, yet it has its own uncertainty.

Author's Response (MC): We thank the reviewer for pointing this out. We have updated this in the main text (**L270**).

The performance of the MA300 was assessed for both precision (via unit-to-unit variability) and relative accuracy (via linear regression against the AE33).

L234: R2 is not a measure of accuracy. It is a measure of goodness of fit, or in this case of the cross-precision of an MA300 and the AE33. I suggest this be revised. The slope is a measure of the accuracy (assuming that the AE33 is, in fact, accurate).

Author's Response (MC): We thank the reviewer for pointing this out. We have updated this in the main text (**L271**).

Relative accuracy was assessed using the slope of the linear fits.

L250: I am unfamiliar with the term "periodical average," and suspect many other readers will be too. I suggest this be defined.

Author's Response (MC): We thank the reviewer for pointing this out. We have defined these two periods in the **Section 2.1** (**L132-L134**) and changed the word "periodical average" to "average". We simply meant the average during the two periods of study in the paper.

We classify the campaign data into two distinct measurement periods based on the days Metro Vancouver issued air quality advisories: September 8 through 18 as "Wildfire" days and the rest as "Regular" days.

Also in Section 3 (**L280-L282**)

Data collected from AE33 and the MA300s (A, B, and C) during the campaign were separated into previously defined "Wildfire" and "Regular" periods for studying the aethalometer's performance in two different sources of aerosols.

L250: I think that grammatically this should be “during the Reg period” and “during the WF period.” This is a general statement for the remainder of the manuscript. It seems that most of the time the author does this.

Author’s Response (MC): We thank the reviewer for this suggestion. We have updated the manuscript to “during the Regular period” and “during the Wildfire period” to improve readability.

Table 1: Please report the actual wavelengths, rather than using e.g., “red.” Also, the number of sig figs is more than seems appropriate given the standard deviations.

Author’s Response (MC): We thank the reviewer for these suggestions. We have updated the table as suggested. However, we would like to highlight this fact that MA300 and AE33 b_{abs} measurements were not done at exact wavelengths, and hence we consider referring to the names of the channels instead of their actual wavelengths. We specify this information in the updated text in **Section 3.2** (**L382-L374**).

As shown in Table S1, the channel-specific wavelength may not match exactly in different aethalometer models. However, for simplicity, we adopted the MA300 measured wavelengths as a reference and the nearest wavelengths from AE33 were used for comparison. We present a statistical summary of multi-wavelength b_{abs} measurements from all four devices in Table 1.

L263: I believe these are average values for the diurnal profile. It would be useful to have this clarified.

Author’s Response (MC): We have included an explanation of the diurnal concentrations in the text (**Section 3.1, L289)**

Since eBC sources can vary within a day, eBC concentration for the measurement periods was again aggregated to estimate the average hour-of-day (diurnal) concentration.

L266: In what way are the R2 values “estimated?” Aren’t they simply “calculated?” This is a general statement that also applies to statements like “the estimated coefficients” such as on L269. These are calculated.

Author’s Response (MC): We thank the reviewer for pointing this out. We have updated the language in the manuscript accordingly.

L281: As there are only three slopes, I suggest just reporting these three rather than stating a range.

Author’s Response (MC): All three slopes are now reported, as suggested.

L295: If the authors still have access to the AE33 and MA300's I encourage them to put a filter on the inlet and measure the standard deviation for particle free air for their specific models.

Author's Response (MC): We have access to the three MA300s and we performed an assessment of MA300 devices for particle free air by adding a HEPA filter. In total, 450 data points of 5 min resolution were collected. After running the test, we find that the unit MA300A did not register any data due to "Optical Saturation" status. We discuss this limitation in the main text (**L352-L359**). The standard deviations from MA300B and MA300C were found as 0.04 μg/m^3 and 0.163 μg/m^3 respectively.

MA300's noise levels were separately assessed in the laboratory following the recommendations by Backman et al. (2017). Briefly, a HEPA filter was installed and the MA300 units were set to intake particle-free air samples at a frequency of 5 minutes for 36 hours with controlled weather parameters indoors. We could not report MA300A's noise level due to instrumental error (optical saturation); however, the noise level for MA300B was 0.04 μg m^{-3}, and for MA300C was 0.163 μg m^{-3}. The noise estimates for MA300 units were much higher (1.2 – 5 times) than AE33's reported noise value. Previously, Holder et al. (2018) reported that noise estimates in MA-series aethalometers could be much higher (1.5 – 5 times) than the reference instrument for 1-min averaged data.

Fig. 3: We know that the absolute absorption value should vary with wavelength, so I think that it is appropriate to use a distinct x-axis scaling for each wavelength considered. As it stands, by using a constant range it is difficult for the reader to see the data as well at the longer wavelengths.

Author's Response (MC): We have removed Figure 3 from the old manuscript. By normalizing the multi-wavelength b_{abs} values, we could discuss unit-to-unit variability of wavelength-dependent MA300's b_{abs} and determined this figure was redundant. In the revised manuscript, we include a discussion on α and include Figure 3 to discuss the variability in α , as mentioned earlier.

L333: The authors state that "The slope's variability in spectral measurements shows similar trends across the MA300 units, indicating the effect of instrumental sensitivity in resolving multiwavelength babs." It is not clear to me how instrumental sensitivity comes into play here. Can the authors expand or clarify?

Author's Response (MC): The focus of this section was to present the variability in the spectral b_{abs} measurements. We wanted to investigate if the unit-to-unit variability differs by wavelength. We also wanted to assess if unit-to-unit variability differed between heavily polluted days (Wildfire period) and cleaner days (Regular period). Previously, (Cuesta-Mosquera et al., 2021) investigated wavelength dependency of unit-to-unit variability of multiple AE33 units for ambient air data (with a cleaner environment) and found no clear trend. However, our results show significant differences in the unit-to-unit variability in MA300 depending on both wavelength of measurement and measurement period. In the updated manuscript **Figure 2** and **Section 3.2.2** (**L412-L425**) have been updated to explain the slope variability.

As shown in Figure 2, the linear fit of individual MA300 units vs AE33's hourly averaged normalized multi-wavelength b_{abs} revealed significant variability within MA300 units. It is important to note that, in Figure 1, we present the linear performance of MA300 units (with respect to AE33) in measuring eBC for the whole campaign, which corresponds to the IR channel measurement only. Equation 5 shows that the linear relationship of MA300's babs,IR and AE33's $b_{abs,IR}$ will be directly related to eBC measurements multiplied by the ratio of MAC_{IR} values between MA300 and AE33. During the Regular period, the slopes ranged between 0.80 and 0.99, while during the Wildfire period, it ranged from 0.71 and 1.16 (Figure 2).

Previously, Cuesta-Mosquera et al. (2021) tested 23 units of AE33 in both laboratory and ambient settings, assessing the instrument's performance before and after maintenance. They found that, after maintenance, AE33 tends to slightly underestimate (slopes slightly reduced from 1) for ambient aerosol measurements at wavelengths 590, 660, and 880 nm, but any wavelength dependency of the unit-to-unit variability of AE33 was not reflected. Here, to assess the unit-to-unit variability of MA300s across the five channels, we used the coefficient of variation (CV) of the normalized slopes from the three units (Figure 2). Unit-to-unit variability was highest in $b_{abs,UV}$ (CV ≈ 8 %). Underestimation of $b_{abs,UV}$ and high unit-to-unit variability will impact the SA results, particularly during the Wildfire period. However, the variability in the Blue channel was found to be low (CV ≈ 4 %) and slope values were much closer to 1 during the Wildfire period, which makes it a potential near-UV wavelength of choice for the SA studies using the MA300.

L335: The authors state "In Figure 3, we identify that the unit-to-unit variability ranged 20–23% during Reg and 17–19% during WF period." It is not clear to me that this is what the figure shows. The variability would be relative to the average of the three instruments and not relative to the reference instrument. Using the appropriate reference would change the slopes and numbers, but probably not the general conclusions.

Author's Response (MC): This comment relates to the previous one. We have removed figure 3 and followed the suggestion of restructuring the Section 3.2.2 and Section 3.2.3. These sections were merged and discussed in detail (with updated Figure 2) in the revised manuscript. The changes can be found in Section 3.2.2.

L339: What is meant by "large offsets in the light absorption measurements."? I find this to be unclear.

Author's Response (MC): We have removed this line in the revised manuscript to improve readability, as it was not a key point.

Fig. 4: I find myself somewhat confused regarding these data and the slopes shown. These slopes should, presumably, be related to the slopes shown in Fig. 2. Yet they are not the same. This should be clarified.

Author's Response (MC): In Figure 4, we aim present the effect of loading different loading correction method adopted in this study and compare them reference measurement (AE33). First, the raw data (blue line in Figure 4) from MA300 which has no correction and found to be significantly overestimating ($MA300_{IR,raw} = 2.7 \times AE33_{IR}$). The slope comes closer to 1 when correction methods were applied, instrument correction (in Green) or modified Drinovec correction (in Yellow). Here we do not segregate the data by periods, as we have done in Figure 2, hence slopes will not remain same. For clarity, we present a simple calculation on the attribution of slopes used in Figure 4 and 1.

<u>How the slope of Figures 4 and 1 are related?</u>

From Equation 5, we know, $b_{abs}(\lambda) = eBC_{\lambda} \times MAC_{\lambda}$

Hence, $\left[\frac{b_{abs,MA300}}{b_{abs,AE33}}\right]_{slope} = \left[\frac{eBC_{MA300}}{eBC_{AE33}}\right]_{slope} \times \left[\frac{MAC_{IR,MA300}}{MAC_{IR,AE33}}\right]$

Now, from Table S1, MAC ratio = $\left[\frac{MAC_{IR,MA300}}{MAC_{IR,AE33}}\right] = \frac{10.12}{7.77} = 1.30244$

For MA300A, $\left[\frac{eBC_{MA300}}{eBC_{AE33}}\right]_{slope}$ = 0.869 --- From Figure 1(a)

Hence, $\left[\frac{eBC_{MA300}}{eBC_{AE33}}\right]_{slope} \times \left[\frac{MAC_{IR,MA300}}{MAC_{IR,AE33}}\right] = 0.869 \times 1.30244 = 1.13$ (slope observed in Figure 4 for $b_{abs,IR}$)

We mention this direct relationship in the main text. (**L464**)

The slope from MA300A's $b_{abs,IR}$ (1.13) is directly related to the slopes presented in Figure 1(a).

L365: The authors here refer to aging affecting wildfire smoke and seemingly exclude aging of fossil fuel-derived particles. Yet, fossil fuel-derived particles also age and can become more hygroscopic over time, so I do not follow this argument. I think that the argument needs to be strengthened or removed.

Author's Response (MC): We thank the reviewer for highlighting this fact. In the updated manuscript, in Section 3.3 (**L441-L446**) we discuss the presence of aged fossil fuel derived particles with the variations of α values.

The peak of AE33's α distribution was found to be 1.69, which is very close to previously recommended as an optimal αbb value by Zotter et al. (2017) for SA calculations. In Figure 3, we present each device's hourly mapped α values estimated by UV-IR and Blue-IR pairs. Day-time α values for both wavelength pairs during the Regular period were found to be lowest and closer to unity, representing aerosol sources from traffic sources Healy et al. (2019); Bernardoni et al. (2021). In contrast, nighttime α values were found to be highest during the regular period, which could be attributed to the local wood-burning sources (Healy et al., 2019).

Further in **Section 3.5(L486-L495)**, we updated the explanation on the ageing of fossil fuel derived particles and its relationship with hygroscopicity.

In our study, all three MA300s were influenced by strong filter loading (Table S6) in addition to RH changes (45% to 95%) during the sampling periods. Being a near-road emission measurement site, our measurements captured complex aerosol mixtures of various mixing states. During the regular period, local traffic during the daytime contributes to fresh BC-enriched aerosols, which can be hydrophobic in nature (Sarangi et al., 2019; Wang et al., 2020) and by night-time, these fresh BC-enriched aerosols can evolve by aging and change their morphological and optical properties. In contrast, during the wildfire smoke-affected days, the measurement site experienced increased quantities of aged aerosols through long-range transport from the Pacific Northwest. These claims align with our calculated α values, as shown in Fig. 3. With the abundance of organic aerosol components during wildfire days, coated BC particles have been found to dominate and often enhance light absorption in lower wavelengths due to the presence of BrC (Healy et al., 2015). This wildfire smoke-affected BC particles can be mixed with a significant fraction of secondary organics, which can be hygroscopic in nature (Wang et al., 2020).

L368: The authors here state that measurements are not useful when they are too noisy. This is a very general statement appropriate for any measurement and so I don't see the value in including this here.

Author's Response (MC): Upon reflection, we agree with the reviewer that this statement is too general to be useful. We have now amended the sentence to transition to discussion of the Blue-IR pairing.

Given these challenges with the UV channel, the Blue-IR channel pairing can be considered as an alternative for source apportionment (Zotter et al. 2017; Deng et al., 2020).

L370: Again, I do not see where the authors have demonstrated the greater accuracy of the blue measurements relative to the UV. This statement is not consistent with Fig. 2.

Author's Response (MC): We thank the reviewer for identifying this point. In the revised manuscript, in Section 3.2.2, and in Figure 2, we have demonstrated the performance of individual channels' b_{abs} measurements in MA300 units. Our results show that average slope (during Wildfire period) for the Blue channel was 0.96 and for the UV channel was 0.78. Lower slope means underestimation with respect to the reference instrument. Also, we calculate the unit-to-unit variability in Blue channel was 4% whereas for UV channel it was 8%. Based on these observations, we considered developing an alternative source apportionment study using Blue channel.

Fig. 5 caption and Section 3.4.1: To reiterate an above point, I do not think it is appropriate to state that the method measures a fraction of eBC mass from biomass burning. That is not what is apportioned. The method does not differentiate between BC sources. It looks at the influence of brown carbon on the eBC determination and ultimately apportions the absorption between BC and BrC.

Author's Response (MC): We understand reviewer's point of view of BC-BrC source apportionment instead of eBC_{bb} and eBC_{ff}. As noted in our response to Major Comment #2, we have added text to clarify the distinction when the labels are introduced, though we have kept the existing labels for consistency with previous literature using this method.

L390: I would go further here to say that not only "may" an assumption of the constant AAE for the biomass component not accurately separate the components to state that it definitely "does not." This links to my earlier point that the AAE values are not fixed for biomass burning and any method that assumes it is a constant is inherently flawed. This goes to the authors finding that the blue-IR apportionment method gives different values than the UV-IR apportionment. The authors used distinct AAE values for each wavelength pair but they have not done any robust assessment of the reasonableness or robustness of these numbers. There is an epistemic uncertainty associated with these values making it impossible to actually know what they "should" be in any given situation. This comes back to my point above that there needs to be a more rigorous assessment of known and unknowable uncertainties.

Author's Response (MC): We hope that this comment has been mostly addressed in Major Comment #5, where we discuss and present the results from AAE (α) measurements and sensitivity analysis done on the data.

L405 and Fig. 5: The authors focus in this manuscript is on assessment of the MA300 units. Thus, I find it a surprising choice to show the AE33 apportionment results in the main text (Fig. 5) while putting the MA300 apportionment results in the supplemental. To me, this seems backwards. I also do not fully see the need for the discussion in Section 3.4.1 as this is all from the AE33 and thus not a core focus of this manuscript.

Author's Response (MC): We appreciate the reviewer's suggestion in light of the focus of the manuscript. We have chosen to keep the AE33 apportionment results in the main text (Figure 5) as we see value in presenting the reference instrument results as a baseline, to which the MA300 results are compared. However, we have also moved Figure S6 (MA300 Source Apportionment Result from Onboard Correction) into the manuscript as Figure 6.