Response to Comments from Reviewer #2 AMT-2018-111

The authors would like to first and foremost thank the reviewer #2 for the careful perusal of the manuscript and the insightful comments which helped improve the manuscript. The reviewer's comments are in italics, the summaries of our responses are in plain font, and the changes in the manuscript are in red text. Page and line numbers refer to the original document.

Reviewer #2

- How the flow rate of low-cost PM sensors of this study were calibrated before, during, and after field deployments? Especially, I am wondering how well the flow rate was maintained in high PM concentrations (i.e., Kanpur)

Response: Thank you for bringing up this question. The configuration of the PMS3003 sensors suggests that their detection approach is volume scattering of the particles rather than light scattering at the single particle level. This volume scattering detection approach determines that the PMS3003 sensors' PM measurements are roughly independent of flow rate. We have added two sentences in the 1st paragraph of Section 2.1 to clarify this issue.

Modified text in the 1st paragraph of Section 2.1 (additions in bold):

"The low-cost sensors evaluated in the present study are Plantower particulate matter sensors (model PMS3003). The Plantower PMS3003 sensors were chosen because 1) they are priced at a small fraction of the cost of reference monitors (approximately USD 30) and 2) their manufacturer reported maximum errors are relatively low (±10 μg m^{-3} in the 0–100 μg m^{-3} range, and ±10% in the 100–500 μg m^{-3} range). The sensors employ a light-scattering approach to measure PM_1, $PM_{2.5}$, and PM_{10} mass concentrations in real-time. Ambient air laden with different-sized particles is drawn into the sensor measurement volume where the particles are illuminated with a laser beam, and the resulting scattered light is measured perpendicularly by a recipient photo-diode detector. These raw light signals are filtered and amplified via electronic filters and circuitry before being converted to mass concentrations. The manufacturer datasheet indicates that the measurement range of this specific sensor model spans from 0.3 μm to 10 μm. **The configuration of the PMS3003 sensors suggests that their detection approach is volume scattering of the particle population rather than light scattering at the single particle level. This volume scattering detection approach results in PM measurements that are independent of flow rate.** PM mass concentration measurements either with or without a manufacturer "atmospheric" calibration are available from the Plantower sensor outputs. Nevertheless, the manufacturer did not provide any documentation to elaborate on how the calibration algorithm was derived. The influence of meteorological factors (e.g., RH, temperature) was likely not accounted for in the manufacturer calibrations. Therefore, we used the sensor reported PM concentration estimates without an "atmospheric" calibration in the current study. Prior to field deployment, no attempt was made to calibrate these sensors under laboratory conditions due to a potentially marked discrepancy in particle size, composition, and optical properties of field and laboratory conditions."

- More specific descriptions on (1) how the instrument segregate the particles into the different size, especially for PM2.5, and (2) its efficiency and accuracy.

Response: Thank you for raising this question. We apologize for not mentioning in the text that the Plantower PMS3003 PM sensors (unlike their PMS1003 and PMS5003 counterparts) are not designed as single particle counters (i.e., the Plantower PMS3003 PM sensors report only PM mass concentration but not count). We believe the allocation of light scattering to PM_1, $PM_{2.5}$, and PM_{10} is based on Plantower's proprietary algorithm, which the manufacturer does not disclose to the public. Therefore, comparing the PMS3003's particle size distribution with that of regulatory-grade instrument (e.g., GRIMM) is impractical and beyond the scope of the current manuscript. Since PMS's particle size distribution is obtained based on a theoretical model rather than an actual measurement (a physical segregation), we would tend to conclude that the efficiency of the segregation is not applicable to the PMS series devices. The primary focus of the current manuscript is how accurately the PMS3003 PM sensors can measure $PM_{2.5}$ mass concentration (rather than particle count) after calibration by co-location with research-grade instruments and after adjustment to meteorology interferences (if available). The evaluation of mass concentration instead of particle count takes priority because only the PM mass concentration (but not particle count) is regulated and monitored under the National Ambient Air Quality Standards (NAAQS) at this stage. We regret that we are unable to comment on the accuracy of PMS3003's theoretical segregation. However, Kelly et al. (2017) thoroughly compared the size distribution provided by the PMS1003 sensors with those provided by a GRIMM Model 1.109 (a portable aerosol spectrometer). The results can be considered as a rough representative estimate of the segregation accuracy of the PMS series devices. In particular, Kelly et al. (2017) demonstrated that "the PMS1003 sensors overestimate total daily average PM counts by a factor of 1.5–2.4 compared to the GRIMM" and "the PMS overestimates particle counts by a factor of (1.1–1.9) for the 0.3 μm bin and increasingly overestimates particle counts as particle size increases". We have added two sentences in the 1st paragraph of Section 2.1 to be more upfront about the characteristics of the PMS3003 sensors.

Modified text in the 1st paragraph of Section 2.1 (additions in bold):

"The low-cost sensors evaluated in the present study are Plantower particulate matter sensors (model PMS3003). The Plantower PMS3003 sensors were chosen because 1) they are priced at a small fraction of the cost of reference monitors (approximately USD 30) and 2) their manufacturer reported maximum errors are relatively low (±10 μg m^{-3} in the 0–100 μg m^{-3} range, and ±10% in the 100–500 μg m^{-3} range). **Unlike their PMS1003 and PMS5003 counterparts, the PMS3003s are not designed as single particle counters.** The sensors employ a light-scattering approach to measure PM_1, $PM_{2.5}$, and PM_{10} mass concentrations in real-time **and are believed to apportion light scattering to PM_1, $PM_{2.5}$, and PM_{10} based on a confidential proprietary algorithm (Kelly et al., 2017)**. Ambient air laden with different-sized particles is drawn into the sensor measurement volume where the particles are illuminated with a laser beam, and the resulting scattered light is measured perpendicularly by a recipient photo-diode detector. These raw light signals are filtered and amplified via electronic filters and circuitry before being converted to mass concentrations. The manufacturer datasheet indicates that the measurement range of this specific sensor model spans from 0.3 μm to 10 μm. PM mass concentration measurements either

with or without a manufacturer "atmospheric" calibration are available from the Plantower sensor outputs. Nevertheless, the manufacturer did not provide any documentation to elaborate on how the calibration algorithm was derived. The influence of meteorological factors (e.g., RH, temperature) was likely not accounted for in the manufacturer calibrations. Therefore, we used the sensor reported PM concentration estimates without an "atmospheric" calibration in the current study. Prior to field deployment, no attempt was made to calibrate these sensors under laboratory conditions due to a potentially marked discrepancy in particle size, composition, and optical properties of field and laboratory conditions."

- Fig 2: what "Raw Sensor" means?

Response: In Fig. 2, "raw sensor $PM_{2.5}$ measurements" means uncalibrated sensor $PM_{2.5}$ measurements. We have added one sentence within the caption of Fig. 2 on page 29 to clarify this.

Modified text within the caption of Fig. 2 on page 29 (additions in bold):

"Figure 1: Flow path for sensor calibrations. **Note raw sensor $PM_{2.5}$ measurements are uncalibrated sensor $PM_{2.5}$ measurements.**"

- Fig. 4 & other scatter plots: (1) indicate the number of used data in color scale to figure out the distribution of PM concentration, (2) add bias and root mean square difference in each figure.

Response: Thank you for suggesting these changes to the scatter plots in the current manuscript.

1) We have shown the distribution of $PM_{2.5}$ concentration in all the scatter plots by adding marginal rugs on both x- and y-axis. Within all the figure captions for the scatter plots, we have also described explicitly that marginal rugs were added to better visualize the distribution of data on each axis. Thus, we feel that these scatter plots do not warrant any additional visualization to illustrate the distribution of PM concentration.

2) We thank the reviewer for suggesting these two performance metrics for low-cost sensor evaluation. First of all, mean bias error (MBE) defined as $\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)$, where $\hat{y}_i$ is the calibrated PMS3003 $PM_{2.5}$ mass concentrations and y_i is the reference monitor $PM_{2.5}$ mass concentrations, is equivalent to the mean of ratios presented in the current paper (defined as the average of ratios of the calibrated PMS3003 $PM_{2.5}$ mass concentrations to reference monitor values). Although not obvious, calibrated PMS3003 $PM_{2.5}$ values always have an MBE of 0 (i.e., $\hat{y}_{i_{mean}} = y_{i_{mean}}$) using a simple linear regression calibration equation and an MBE roughly close to 0 (i.e., $\hat{y}_{i_{mean}} \cong y_{i_{mean}}$) using a quadratic calibration equation. This is equivalent to say calibrated PMS3003 $PM_{2.5}$ values should always have a mean of ratios close to 1 (i.e., $\hat{y}_{i_{mean}} \cong y_{i_{mean}}$). The only difference between MBE and mean of ratios is that the former one is expressed in a subtraction form while the latter one in a division form. Second, we agree that RMSE is a standard and widely used performance score. Yet, unlike percent error used in the current manuscript, RMSE is generally unfavourable for comparison across different data sets (particularly data sets from different field campaigns with drastically different PM concentrations). Specifically, calibrated PMS3003 $PM_{2.5}$ values at a 1 h time resolution had an average RMSE of 10.4±0.2 μg m^{-3}, 12.7±0.1 μg m^{-3}, and 31.0±0.9 μg m^{-3} at the Duke site, IIT Kanpur site during monsoon, and post-monsoon season, respectively. However, given that the hourly averaged ambient $PM_{2.5}$ levels at IIT Kanpur during post-monsoon (116±57 μg m^{-3}) were roughly 13 times

higher than that at the Duke site (9±9 μg m^{-3}), it is unfair to conclude that the sensors had the best performance at the Duke site while the worst performance at the IIT Kanpur site during post-monsoon based on RMSEs. We tend to believe that a normalized metrics such as the percent error in this study is more straightforward and more suited for observing the trend in measurement accuracy with PM concentrations. We tend to hold that RMSE is more appropriate to be used in model selection given a same data set as described in Section 3.3.3 and 3.3.4 (see Table S2 and S3). Overall, mean of ratios defined in this study are equivalent to MBE and the error defined in this study is a legitimate quantitative performance criterion to more intuitively validate that PMS3003 sensors can have credible data and good accuracy after proper calibrations based on reference monitors and corrections for meteorological interferences. Therefore, we feel that these scatter plots do not warrant additional performance metrics such as bias and RMSE. However, during our calculation of the RMSEs prompted by this question, we found that we have incorrectly defined RMSE as the standard deviation of differences between calibrated and raw PMS3003 $PM_{2.5}$ mass concentrations (it should be the difference between calibrated PMS3003 and reference monitor $PM_{2.5}$ mass concentrations) in Section 2.3.3 on page 11 in the original manuscript and consequently incorrectly calculated RMSEs for Table S2 and S3 on page 10 in the original supplementary document. This is a major oversight on our side and we deeply regret for this mistake and have made the corresponding corrections. In particular, the correct recalculation has significantly lowered the RMSEs by up to 15 μg m^{-3} for 1 h quadratic method in Table S2 and for 1 h linear method in Table S3. The revision did not change the conclusion of the manuscript.

Modified text in Section 2.3.3 on page 11, lines 2–3 (changes in bold):

"where n is the number of observations, $\hat{y}_i$ is the calibrated PMS3003 $PM_{2.5}$ mass concentrations, and y_i is the **reference monitor** $PM_{2.5}$ mass concentrations."

Modified text in Section 3.3.3 on page 20, lines 16–19 (changes in bold):

"Even when the nonlinearity was not strong enough to make the simple linear fit statistically different from the quadratic fit (i.e., the quadratic term a_2 in the quadratic fit (Eq. (7)) not significantly different from 0 with p>0.1) at 24 h integration time, the quadratic fit can still reduce the mean error and **the range of RMSEs** by 2% (Table 4)**,** and **3** μg m^{-3} (Table S2), respectively."

Modified Table S2 in supplementary document on page 10 (changes in bold):

Table S1: Summary of AIC and RMSE (goodness of fit and accuracy estimates) for the two E-BAM calibrated PMS3003 $PM_{2.5}$ responses during the post-monsoon season at IIT Kanpur, using the simple linear and quadratic calibration methods as a function of time averaging intervals. The results are displayed in mean (range) format. Note the mean statistics were obtained by fitting the models to the PMS3003 $PM_{2.5}$ measurements averaged across the two sensor package units at each point in time. The model that had the best goodness of fit and accuracy estimates at each averaging time interval is indicated with shading.

Timescales	1 h		6 h		12 h		24 h	
Method	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic[1]
AIC	5731 (5731–5778)	5670 (5670–5720)	932 (932–949)	916 (916–933)	462 (462–474)	454 (454–469)	214 (214–229)	210 (210–225)
RMSE	**31 (31–33)**	**27 (27–28)**	**21 (21–23)**	**19 (19–21)**	**19 (19–21)**	**18 (18–21)**	**13 (13–17)**	**13 (13–14)**

[1]The quadratic term (a_2) in the quadratic fit (Eq. (7)) for the PMS3003-6 was not significantly different from 0 (p>0.1).

Modified Table S3 in supplementary document on page 10 (changes in bold):

Table S2: Summary of AIC and RMSE (goodness of fit and accuracy estimates) for the E-BAM calibrated PMS3003 $PM_{2.5}$ results of the pooled Duke University and IIT Kanpur data sets, using the simple linear and quadratic calibration methods as a function of time averaging intervals. Note the values are statistics for the averaged sensor models, which were obtained by fitting the models to the means of PMS3003 $PM_{2.5}$ measurements averaged across all sensor package units during each sampling period. The model that had the best goodness of fit and accuracy estimates at each averaging time interval is indicated with shading.

Timescales	1 h		6 h		12 h		24 h	
Method	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
AIC	21005	20638	3376	3225	1697	1590	836	764
RMSE	**18**	**17**	**12**	**11**	**12**	**10**	**10**	**8**

[*]All the models' coefficients were statistically significant (p<0.1).

- Why the slope is higher than 1 for most cases (e.g., Figs 4 and 7)? Discussion on the potential factors effecting on low-cost PM sensor measurements would be helpful in future research. - It would be nice to provide in supplement how much the total cost, including the sensor, was.

Response:

1) Thank you for bringing up this question. The slope is higher than 1 for most cases suggests that Plantower PMS3003s mostly overestimate ambient $PM_{2.5}$ mass concentrations prior to calibration based on reference monitors. We regret that we are unable to comment on the underlying reasons for this overestimation since Plantower PMS3003s allocate light scattering to PM_1, $PM_{2.5}$, and PM_{10} mass concentrations based on Plantower's confidential proprietary algorithm (as described in our response to your second comment). Plantower is not clear on whether, and how the sensors may be calibrated prior to shipping them to customers. We would like to emphasize that the important question is how well the $PM_{2.5}$ mass concentrations made by these low-cost sensors after calibration by collocation and adjustments to meteorological interferences can compare to the true ambient $PM_{2.5}$ values reported by reference monitors. And we have validated in the current manuscript that these PMS3003 sensors can have credible $PM_{2.5}$ data and good accuracy after proper calibrations based on ideal reference monitors and corrections for meteorological biases.

2) Thank you for the suggestion. We have added a sentence in Conclusion on page 22, line 1 to make our discussion more prominent.

Modified text in Conclusion on page 22, line 1 (additions in bold):

"Even though the RH correction factor models might be highly location-specific, it is striking to see that they were capable of explaining up to nearly 30% of the variance in 1 min, 1 h and 6 h aggregated sensor measurements and reducing mean errors down from ~22–27% to roughly 10% even at the finest 1 min time resolution. Compared to the RH corrections, temperature corrections were found to be relatively small and can only scale uncertainties down by 7% at most; however, in addition to the other corrections this may help to achieve the highest possible accuracy level. It is important to note that the success of both RH and temperature corrections relies on the precision of reference instruments. **Properly accounting for these systematic meteorology-induced influences is helpful in making high quality $PM_{2.5}$ measurements at a low cost.**"

3) Thank you for the suggestion. The cost of Plantower PMS3003 sensor itself has already been mentioned in Section 2.1 on page 4, lines 22–23, which reads "…1) they are priced at a small fraction of the cost of reference monitors (approximately USD 30) …". The total cost of a full Duke PM air quality monitoring package has also been mentioned in Section 2.1 on page 5, lines 14–16, which reads "The total material costs for one PM monitoring package including the Plantower PMS3003 sensor, the supporting circuitry, the enclosure, and additional power cords are approximately USD 200.". Prompted by this comment, we have added the costs of the Plantower PMS3003 sensor, the supporting circuitry, the enclosure, and the power cords to this sentence in Section 2.1 on page 5, lines 14–16 to clarify the breakdown of the total costs.

Modified text in Section 2.1 on page 5, lines 14–16 (additions in bold):

"The total material costs for one PM monitoring package including the Plantower PMS3003 sensor **(~ USD 30)**, the supporting circuitry **(~ USD 140 including PCB with almost all components)**, the enclosure **(~ USD 20)**, and additional power cords **(~ USD 20)** are approximately USD **210**."