Response to Comments from Reviewer #1 AMT-2018-111

The authors would like to first and foremost thank the reviewer #1 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 #1

Section 2.1: Although the dimensions of the electrical box housing the instruments is provided, it would be useful to also provide the dimensions and weight of the instruments themselves.

Response: Thank you for pointing this out, we agree that the weights of the enclosure and the instrument itself are important as the weight is one of the major design criteria that low-cost air quality packages are required to meet, and the lightweight characteristic gives emerging low-cost devices an advantage over traditional instrumentation. We have added the weight information about both the enclosure and the instrument itself (the PMS3003 PM sensor) in Section 2.1 along with the dimension of the instrument.

Modified text in Section 2.1 (additions and changes in bold):

"The Plantower PMS3003 sensor **(dimension: 5.0 cm L × 4.3 cm W × 2.1 cm H; weight: 40 g)** along with a Sparkfun SHT15 RH and temperature sensor, a Teensy 3.2 USB-based microcontroller, a ChronoDot V2.1 high precision real-time clock, a microSD card adapter, a Pololu 5V S7V7F5 voltage regulator, a DC barrel jack connector, and a basic 5 mm LED was connected to a custom designed printed circuit board (PCB), shown in Fig. 1a. We programmed the Teensy 3.2 microcontroller to measure PM mass concentrations (μg m^{-3}) every second and to store the time-stamped 1 min averaged measurements to text files on a microSD card. To protect sensors from rain and direct sunlight, all components were housed in a 20.50 cm L × 9.95 cm W × 6.70 cm H**, 363 g lightweight** NEMA (National Electrical Manufacturers Association) electrical box (Bud Industries NBF32306) as shown in Fig. 1b. The inlet of the Plantower sensor was aligned with a hole drilled in the electrical box to ensure unrestricted airflow into the sensor. **Each Duke PM air quality monitoring package is estimated to weigh ~430 g in total and was** continuously powered up by **a** 5V 1A USB wall **charger**. 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. More detailed instructions on how to assemble the sensor packages and information on how to use their data can be found on our webpage (http://dukearc.com)."

Page 8, line 20: change to ". . .AGREED quite well".

Response: Thank you, we have made the suggested change to wording.

Modified text in Page 8, line 20 (changes in bold):

"They found that the model **agreed quite well with the** field data **both** collected from their study and **from** a previous study (Day and Malm, 2000)."

Page 8, line 26: change to ". . .WAS THE RH correction factor. . ."

Response: Thank you, we have corrected the grammatical error.

Modified text in Page 8, line 26 (change in bold):

"Ordinary least squares (OLS) regressions were conducted to obtain the empirical regression parameters *a* and *b* in Eq. (1), where the dependent variable was **the** RH correction factors calculated as the ratio of…"

Page 8, line 32: change to ". . .THE RH correction factor. . ."

Response: Thank you, we have corrected the grammatical error.

Modified text in Page 8, line 32 (change in bold):

"The empirical equations derived were used to compute **the** RH correction factor for a given RH…"

Page 9, line 1: change to ". . .were compensated FOR by. . ."

Response: Thank you, we have corrected the grammatical error.

Modified text in Page 9, line 1 (change in bold):

"The RH interferences were compensated **for** by dividing…"

Page 9, line 5: Why were the RH adjustments only made for R2 values greater than 0.4? Please add an explanation to the text.

Response: A relatively high correlation value (i.e., $R^2 = 0.4$) of the empirical equation for computing the RH correction factors (i.e., Eq. (1)) was chosen as the cut-off point in this study because we want to ensure that the RH corrections can indeed lower the error of the low-cost sensor $PM_{2.5}$ measurements. Given the poor precision of the E-BAM, lower correlation values may lead to marginal (if any) improvements in the accuracy of the low-cost sensor measurements. This theory can be corroborated by the temperature correction results from the current study: An AIC difference of 2 is a standard threshold for model selection. However, even when the AIC indicated that the temperature predictor was statistically significant in the calibration model, the temperature correction still resulted in marginal (Kanpur monsoon 6 h results, see Table 4) or no (Kanpur monsoon 1 h results, see Table 4) or negative (Duke 1 h results, see Table 2) improvements. Furthermore, the highest correlation of the empirical RH correction factor equation obtained at sites using an E-BAM as the reference monitor was 0.13 (Kanpur monsoon 1 h results, see Fig. S7). This value is too low to warrant conducting the RH correction (even if we lowered the cut-off point to a non-ideal 0.20). Other users who have access to more precise regulatory-grade instruments can choose to lower this threshold as they see fit. We agree that the original text lacks the corresponding justification and is therefore unclear. We have added an explanation to the original text.

Modified text in the 3rd paragraph of Section 2.3.1 (additions in bold):

"We only performed the RH adjustments when the fitted models for any of the sampling locations over any time averaging interval had at least a moderate coefficient of determination ($R^2 \geq 0.40$). **The slightly high correlation cut-off value was implemented in this study to ensure that the RH corrections can effectively lower the error of the low-cost sensor**

$PM_{2.5}$ measurements. Despite the similarity of the general shape of correction factor curves in different studies, the detailed behaviors of aerosols diverged greatly due to considerable difference in particles' chemical composition and diameter (Waggoner et al., 1981; Zhang et al., 1994; Day and Malm, 2000; Chakrabarti et al., 2004; Soneja et al., 2014). In a previous study (Day and Malm, 2000), aerosols mass at some locations began to increase continuously above a relatively low RH (such as 20%), whereas at other locations it exhibited a distinct deliquescent behavior (i.e., aerosols water uptake occurred at a relatively high RH). Even for aerosols showing deliquescent behavior, the observed deliquescence RH (RH threshold) varies from study to study. Soneja et al. (2014) also found underestimation of PM concentrations (correction factors less than 1) below 40% RH. Because of these uncertainties, we conducted RH adjustments across the entire range of recorded RH without incorporating an RH threshold. Additionally, the RH adjustments in this study were always performed separately from and prior to either temperature adjustments or reference monitor adjustments."

Page 9, line 8: change to ". . .PARTICLE chemical composition and . . ."

Response: Thank you, we have made the suggested change to wording.

Modified text in Page 9, line 8 (change in bold):

"…the detailed behaviors of aerosols diverged greatly due to considerable difference in **particle** chemical composition and diameter…"

Page 9, lines 9 – 10: change to ". . .AEROSOL mass at some locations INCREASED continuously. . ."

Response: Thank you, we have made the suggested change to wording.

Modified text in Page 9, lines 9–10 (changes in bold):

"…**aerosol** mass at some locations **increased** continuously above a relatively low RH…"

Page 11, line 11: R2 is not a measured parameter. Please change the sentence.

Response: Thank you for pointing this out. We agree that all the performance metrics including R^2, RMSE, MAE, and MBE are calculated rather than measured parameters. We have revised the language.

Modified text in Page 11, line 11 (change in bold):

"To date, only a few studies have attempted to **compute** parameters other than R^2 to gauge the overall performance of low-cost sensor technologies."

Page 12, Lines 14 – 16: For the statement that "correlations among the five uncalibrated PMS3003 units were high. . .", please provide the timescale.

Response: Thank you, the timescale has been added.

Modified text in Page 12, lines 14–16 (additions in bold):

"Correlations among the five uncalibrated PMS3003 units were high (R^2 = 0.98–1.00) **on 1 h timescale** even under low ambient $PM_{2.5}$ concentrations with slopes averaging 1 ± 0.1 and negligible intercepts averaging 0.3 ± 0.3 (Fig. S1), suggesting excellent intra-PMS3003 precision."

Page 13: lines 32 – 33: change to . . ."through -3) from the Duke University site to the US EPA. . ."

Response: Thank you, we have made the suggested change.

Modified text in Page 13, lines 32–33 (changes in bold):

"…we moved three PMS3003 units (labeled PMS3003-1 through -3**)** from the Duke University study site to the US EPA…"

Page 15, line 14: Figure 6 is referenced in the text before Figure 5b.

Response: Thank you for pointing this out. We have revised Section 3.2.1 and 3.2.2 by moving the descriptions of both Figure 5a and 5b to the beginning of Section 3.2.1, before the first reference to Figure 6 (at the beginning of Section 3.2.2).

Modified text in Section 3.2.1 and 3.2.2 (additions and changes in bold):

"3.2.1 $PM_{2.5}$ concentration, RH, and temperature on 1 h scale

Fig. 5a shows 1 h time series data from all the reference monitors including the SHARP's embedded nephelometer and Fig. 5b juxtaposes the T640_Roof and the three uncalibrated PMS3003 units $PM_{2.5}$ measurements at 1 h time resolution. Table 1 indicates that the 1 h averaged ambient $PM_{2.5}$ levels at the US EPA RTP (9–10 µg m^{-3}) matched those at Duke University (9 µg m^{-3}). However, Fig. 5a ~~**shows 1 h time series data from all the reference monitors including the SHARP's embedded nephelometer and**~~ depicts smaller ranges of ambient $PM_{2.5}$ concentrations than were measured at Duke University. Table 1 indicates that the Std.Dev (less than 4 µg m^{-3}) and maximum $PM_{2.5}$ concentration (less than 20 µg m^{-3}) at the EPA RTP were significantly lower than at Duke University (9 µg m^{-3} and 62 µg m^{-3} for Std.Dev and maximum, respectively). These comparisons imply that the RTP sampling location had overall lower ambient $PM_{2.5}$ concentrations and was consequently more challenging for low-cost sensors than the Duke University sampling site. During the measurement period, the mean RH and temperature were 64 ± 22% and 30 ± 7°C, respectively. The higher average RH level at the EPA RTP than at Duke University (45 ± 19%) accentuated the RH interference in the PMS3003 $PM_{2.5}$ measurements, as seen in Sect. 3.2.3.

3.2.2 PMS3003 performance characteristics on various timescales prior to adjustment for meteorological parameters

Figures 6a–b summarize graphically and statistically the pairwise correlations between all the instruments' 1 min aggregated and 1 h aggregated $PM_{2.5}$ mass concentrations, respectively. The R^2 and calibration factors between all the instruments on 1 min and 1 h scale were similar. The PMS3003 sensors were well correlated with one another (R^2 = 0.97), the two T640s (R^2 ≥ 0.63) and the SHARP's embedded nephelometer (R^2 ≥ 0.49) even for 1 min aggregated data at exceptionally low ambient $PM_{2.5}$ levels. In contrast, the 1 min or 1 h PMS3003–SHARP correlations (R^2 ≥ 0.25) were poor and worse than the 1 h PMS3003–E-BAM correlations (R^2 ≥ 0.36) at the Duke site. Additionally, the SHARP had only moderate correlations with the two T640s (R^2 ≤ 0.58) or the SHARP's embedded nephelometer (R^2 = 0.59) even though both the SHARP and T640 are US-designated $PM_{2.5}$ FEMs and the SHARP readings take into account its raw nephelometer values.

While the common optical-based principles of operation shared by T640 (and nephelometer) and PMS3003 could partially explain the stark performance contrast between the SHARP and T640 (and nephelometer), the lower reported precision of

the beta-attenuation-based approach with a 24 h average of ±2 μg m^{-3} for SHARP than the T640 with an 1 h average of ±0.5 μg m^{-3} in low ambient $PM_{2.5}$ concentration environments appears to be the root cause (Thermo Fisher Scientific, 2007; Teledyne Advanced Pollution Instrumentation, 2016). A previous study by Holstius et al. (2014) demonstrated the poor performance of BAM-1020 in a comparably low concentration environment in Oakland, CA. They have used both statistical simulation based on the true ambient $PM_{2.5}$ distribution and the measurement uncertainty of BAM-1020 (1 h average: ±2.0–2.4 μg m^{-3}) provided by the manufacturer (Met One Instruments) and field test results to show that an R^2 of ~0.59 is as correlated as one would expect from the 1 h measurements of a pair of collocated BAM-1020s. In contrast to the moderate intra-BAM-1020 correlation (~0.59) reported by Holstius et al. (2014), the two collocated T640s yielded an ideal R^2 of 0.95 (Fig. 6), which suggests a significantly smaller measurement error in the T640 than in the BAM-1020. The SHARP is known to derive its reported values by dynamically adjusting its embedded nephelometer readings based on its BAM measurements. In other words, the SHARP performance was adversely affected by the low precision of its embedded BAM at low ambient $PM_{2.5}$ levels. All these observations seem to imply that beta-attenuation-based monitors might be unfavorable for low-cost particle sensor evaluation at the low concentrations typically present in the US. US EPA FEMs are valid for 24 h $PM_{2.5}$ measurements rather than for 1 h measurements (Jiao et al., 2016). An inappropriate selection of reference monitors might prejudice the overall performance of low-cost sensors particularly for time resolutions finer than 24 h.

The T640 sitting on the roof (T640_Roof) was chosen over the SHARP and the other T640 unit (T640_Shelter) as the reference monitor because 1) the T640 as a US-designated $PM_{2.5}$ FEM is better for sensor evaluation at low concentrations than a SHARP; 2) the T640_Roof had slightly lower correlations with the sensors than the T640_Shelter, therefore giving conservative estimates of PMS3003 performance. ~~**Figure 5b juxtaposes the T640_Roof and the three uncalibrated PMS3003 units $PM_{2.5}$ measurements at 1 h time resolution.**~~ Similar to the Duke University results, comparisons of the data using regression between the same set of instruments in Figs. 7a–d present similar calibration factors across the sensors on the same timescale, therefore indicating the excellent precision of the PMS3003 model. Unlike the analysis of the Duke University data, the calibration factors (prior to adjustments for meteorological parameters) varied little from one averaging timescale to another (Table 3). Despite an appreciable improvement in R^2 compared to the Duke University site being found only on the 1 h scale, the accuracy of the T640 calibrated PMS3003 units substantially outperformed their E-BAM calibrated counterparts across the entire averaging time spectrum (Table 3) with the most pronounced difference on 1 h scale (27% vs. 201%). A less dramatic mean error drop from 1 h to 24 h scale at the EPA RTP (27% to 9%) compared to what was seen at the Duke University site (201% to 15%) highlights the inferior precision of the E-BAM and further undermines its credibility as a reference sensor at low $PM_{2.5}$ concentrations. It should be noted that the non-normally distributed residuals on 1 min, 1 h and 6 h scales in Figs. 7a–c indicate that the true ambient $PM_{2.5}$ concentration term alone was not sufficient to explain the variation of PMS3003 measurements, therefore revealing the likely existence of RH or temperature impacts."

Page 16, line 11: Figure 8 is referenced in the text before Figure 7e-g.

Response: Thank you for pointing this out. We have revised Section 3.2.3 to ensure that Figures 7e–g are referenced in the text before Figure 8.

Modified text in Section 3.2.3 (additions and changes in bold):

"3.2.3 RH adjustment to sensor $PM_{2.5}$ measurements

~~**As shown in Fig. 8, the empirical RH adjustment equation (i.e., Eq. (1)) fitted well with the 1 min, 1 h, and 6 h aggregated data ($R^2 \geq 0.48$). The regression fit statistics degraded when evaluating 12 h and 24 h aggregated data, likely because of an insufficient number of observations and stronger smoothing effects at longer averaging time intervals.**~~ **Figures 7e–g display the regressions of $PM_{2.5}$ measurements from the RH adjusted PMS3003 units versus the T640_Roof on 1 min to 6 h timescales. The empirical equations of the RH correction factors (i.e., Eq. (1)) on the corresponding timescales are shown in Fig. 8 and they fitted well with the 1 min to 6 h aggregated data ($R^2 \geq 0.48$). The RH adjustment was not implemented to the 12 h and 24 h aggregated data because the equation regression fit statistics degraded when evaluating these data, likely because of an insufficient number of observations and stronger smoothing effects at longer averaging time intervals.** Aerosols at the EPA RTP generally exhibited smooth and continuous growth above the lowest collected RH rather than distinct deliquescence behavior (Fig. 8). The RH correction factors were roughly 20 to 30% above 1 even at the lowest RH (below 30%), which justifies the decision of conducting RH adjustments across the entire range of recorded RH without incorporating an RH threshold. Despite the promising descriptions of correction factors as a function of RH, wide divergence in the magnitude of correction factors for a given RH exists. This divergence is likely the result of substantial day-to-day variation in the chemical composition of the aerosols (Day and Malm, 2000). A higher fraction of soluble inorganic compounds can contribute to a larger magnitude of RH correction factors (Day and Malm, 2000).

~~**Figures 7e–g display the regressions of $PM_{2.5}$ measurements from the RH adjusted PMS3003 units versus the T640_Roof.**~~ The RH corrections brought the PMS–T640 correlations to above 0.90 for all 1 min, 1 h, and 6 h aggregated data **(see Figs. 7e–g)**. This significant improvement in R^2 implies a major RH influence that can explain up to nearly 30% of the variance in 1 min and 1 h PMS3003 $PM_{2.5}$ measurements in addition to the true ambient $PM_{2.5}$ concentration variable. Figure S3 demonstrates that the PMS3003-to-T640 ratios after the RH corrections were also considerably closer to a strict normal distribution than those with only the FEM corrections (Fig. S4). However, Figs. 7e–g suggest that the PMS3003 $PM_{2.5}$ measurements were still not in complete agreement with the T640 readings even after the RH adjustments. This discrepancy might stem from variations in aerosol composition described previously or impacts of particle size biases (Chakrabarti et al., 2004), therefore warranting a further step of FEM conversion (adjustment). According to Table 3, the combination of RH and FEM corrections were able to substantially improve the accuracy of PMS3003 $PM_{2.5}$ measurements by reducing the mean errors to within 12% even for data at 1 min time resolution. The ideal normal distribution of PMS3003-to-T640 ratios in combination with the high accuracy and precision of the finest-grained data proves especially

beneficial for minimization of exposure measurement errors in short-term $PM_{2.5}$ health effect studies (Breen et al., 2015) or mapping of intra-urban $PM_{2.5}$ exposure gradients (Zimmerman et al., 2018)."

Page 16, line 30: Omit "achieved".

Response: Thank you, we have made the suggested change.

Modified text in Page 16, line 30 (change in bold):

"This marginal improvement ~~**achieved**~~ stands in marked contrast to that brought about by the RH corrections (up to 17%)…"

Page 17, line 24: change to ". . .RH values MEASURED in. . ."

Response: Thank you, we have made the suggested change to wording.

Modified text in Page 17, line 24 (change in bold):

"These RH values **measured** in Kanpur were also similar to those at the EPA RTP site (64 ± 22%)."

Page 18, lines 5 – 10: Is it possible to clean the sensors and see if that changes the instrument performance?

Response: Thank you for suggesting this possibility. Unfortunately, we have not attempted to clean the sensors throughout the current Kanpur field test. We acknowledge that the effect of PM deposition on the low-cost PM sensor performance and calibration particularly in areas of high ambient PM concentrations (e.g., Kanpur) is understudied. Considering the substantial implications of this research topic for the development and maintenance of future low-cost PM sensors networks in environments such as polluted urban areas, we believe a separate, specialized, and well-designed field campaign is required for a rigorous evaluation. Also given the present long length of the manuscript, we also inclined not to expand on this complicated issue. However, we are planning to address this issue by determining if routine cleaning (e.g., gently blowing through the low-cost sensor with canned air) will be helpful for maintaining or improving the sensor performance in a forthcoming publication. We have added additional text to the 1st paragraph of Section 3.3.2 to clarify our points.

Modified text in Section 3.3.2 (additions in bold):

"As with the two field tests in the low concentration region, the two PMS3003 units were highly correlated with each other during both the monsoon ($R^2 = 0.99$) and post-monsoon seasons ($R^2 = 0.93$) in Kanpur (Fig. S6). This good agreement is also reflected in Fig. 9, which displays that the two sensors were in sync and tracked reasonably well with the E-BAM. However, there was a minor decrease in the intra-sensor correlation from the monsoon to post-monsoon seasons that might signal a performance change of the two PMS3003 sensors either due to minor deterioration or a change in the pollutant source. Figure S6 illustrates that the magnitude of the deviation from the regression line during the monsoon season was likely irrelevant to the deployment time (measured by the number of hours past the beginning of the Kanpur study, i.e., 2017 June 08 00:00). In contrast, the extent of the divergence was somewhat larger for the longer deployment time near the high end of the $PM_{2.5}$ range over the post-monsoon period. One plausible explanation for the distinguishable post-monsoon (but not monsoon season) change is the routine exposure (for nearly a month) of the sensors to high concentrations of accumulation mode aerosols. This may be especially detrimental to PM sensors; all the more so because the foggy condition during post-monsoon and winter over Kanpur may further exacerbate the accumulation of aerosol particles at lower surfaces and therefore the deposition of particles within the sensors (Li et al., 2015; Bran and Srivastava, 2017). This constant exposure

possibly caused disproportionately large detection errors primarily near the upper end of the $PM_{2.5}$ range. **The effect of PM deposition on the low-cost PM sensor performance and calibration particularly in areas of high ambient PM concentrations (e.g., Kanpur) was not evaluated as part of this work. Future studies will present how preventive maintenance of low-cost sensors including periodic cleaning can benefit their performance.** Another possible explanation is the change of dominant pollutant source from the early stage of monsoon (long-range transport of mineral dust from Iran, Afghanistan, Pakistan, and the Thar Desert) to post-monsoon (local impact of biomass burning emissions) season (Ram et al., 2010). Sensors are likely to respond differently to different varieties of aerosols and the change in sensor responses might be most pronounced near the upper end of the $PM_{2.5}$ range. Figure 9b substantiates the potential change by showing that the two uncalibrated PMS3003s were unable to match the troughs of the E-BAM (even troughs below 40 µg m^{-3}) throughout the post-monsoon season, as they were during the monsoon season in Fig. 9a."

Page 18, line 25: What is meant by ". . .reaching the troughs of true ambient PM2.5 concentrations"?

Response: The troughs mean the local minima of the true ambient PM2.5 concentrations. We have changed the terminology from troughs to local minima throughout the manuscript for clarity.

Modified text in Page 18, lines 14–16 (changes in bold):

"Figure 9b substantiates the potential change by showing that the two uncalibrated PMS3003s were unable to match the **local minima** of the E-BAM (even **local minima** below 40 µg m^{-3}) throughout the post-monsoon season, as they were during the monsoon season in Fig. 9a."

Modified text in Page 18, lines 24–25 (change in bold):

"This appreciable drift in baseline agreed with the sensors being incapable of reaching the **local minima** of true ambient $PM_{2.5}$ concentrations."

Modified text in Page 20, lines 7–9 (change in bold):

"Figure S8 shows that the quadratic model might suit the post-monsoon 1 h aggregated data better than the simple linear model as the simple linear model failed to capture the **local minima** of the E-BAM throughout the post-monsoon period."

Page 19, line 14: change to ". . .PARTICLE chemical,. . ."

Response: Thank you, we have made the suggested change to wording.

Modified text in Page 19, line 14 (change in bold):

"…great reliance on **particle** chemical, microphysical, and optical properties (Laulainen, 1993)."