**Response to Reviewer Comments on Manuscript #egusphere-2024-1685 "Performance evaluation of Atmotube Pro sensors for air quality measurements" by Shittu et al.**

We thank the reviewers for their constructive comments and feedback. In our responses below, we have reproduced reviewers' comments in black text, followed by our responses in red. Proposed text additions to the manuscript are in blue and added as tracked changes in the manuscript. Our reference to the line numbers in this document is based on the originally submitted manuscript. The major and minor comments from both reviewers are listed using numbers and letters respectively for more clarity.

**# REVIEWER 1**

This paper presents a comparison of 8 Atmotubes to a FIDAS reference monitor in Leeds, UK. It is a comparison of a single sensor type at a single location that is well-monitored, thus the impact and reach of this study is a bit limited. The analysis of the performance is done adequately, but it is a bit shallow. I recommend major revisions.

Here are the two main things that I expected to see in this paper. I think these additions would substantially increase the impact and value of this paper.

1. A comparison of the size distribution of what the AtmoTubes measure versus what the FIDAS measures. It is claimed that the AtmoTubes can do PM1, PM2.5, and PM10. It would be very interesting to see if that is true, or if as recent literature shows, these size distributions are more mathematical constructions. See Molina Rueda et al (https://pubs.acs.org/doi/full/10.1021/acs.estlett.3c00030). This analysis would be especially pertinent given the FIDAS ability to size particles (PM1, PM2.5, PM4, PM10, etc).

Thank you, this is a very valid point. We have added to the manuscript, the performance of $PM_1$, $PM_{2.5}$ and $PM_{10}$ however, the focus of rest of the manuscript was on $PM_{2.5}$ because it is the key standard in air quality by WHO and other regulatory agencies for epidemiological research. $PM_{10}$ recorded by low-cost sensors has been established and reported in earlier literature and more recently by the suggested paper by Molina Rueda et al., (2023) to have poor performance in comparison with reference-grade monitors. We carried out a performance evaluation of other size distributions, and this was not different for Atmotube Pro as the $PM_{10}$ association with FIDAS reference was also quite poor; $R^2$ of 0.48 and error bias (RMSE) 6.2 µg/m$^3$ while $PM_1$ performed very well ($R^2$ >0.9 and RMSE value < 2 µg/m$^3$). The following statement and figure have been added to the manuscript.

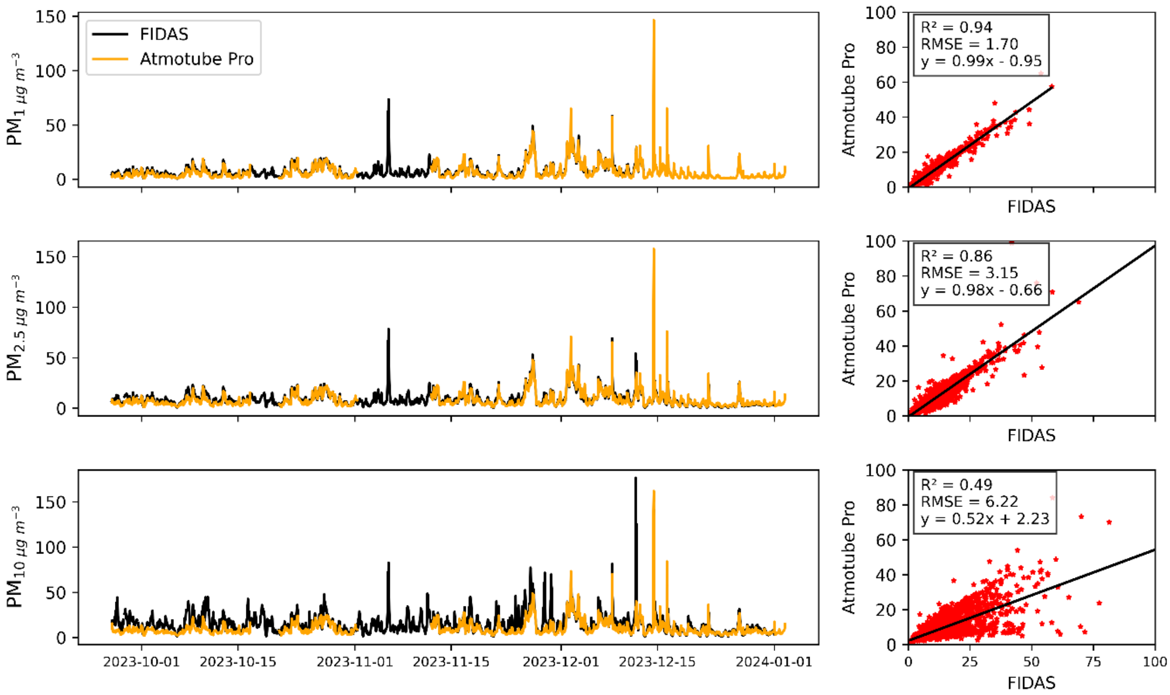



Figure 1: Comparison of Atmotube Pro and FIDAS reference monitor data for size distribution $PM_1$, $PM_{2.5}$ and $PM_{10}$ size (hourly averaged data).

Line185: The performance of Atmotube Pro sensor for $PM_1$, $PM_{2.5}$ and $PM_{10}$ in comparison with the reference FIDAS monitor are shown in Figure 1. The average of all 8 sensors was computed at hourly time resolution. $PM_1$ had a very low error bias of 1.7 µgm$^{-3}$ and a strong $R^2$ of 0.94. $PM_{2.5}$ had a larger error bias of 3.2 µgm$^{-3}$and a decrease in the $R^2$ value to 0.86 in comparison to $PM_1$. The poorest performance was recorded for $PM_{10}$ with a larger error bias of 6.2 µgm$^{-3}$and a further decline in $R^2$ of 0.49. Similar results were reported in the study by Molina et al., 2023 using Plantower, Sensirion and Piera low-cost sensors. The rest of the paper will focus on particle size <2.5 µm ($PM_{2.5}$) as $PM_{2.5}$ is the key standard by WHO and other regulatory agencies for health-related research.

2. Correction factor development. If a sensor evaluation is already being undertaken, it is not much more work to provide correction factors based on these co-locations. There are numerous references in the literature that do this. This would help provide motivation to the paper so that future AtmoTube users may be able to apply corrections to their data in certain environments.

Thank you for the above suggestion. We were able to add a new section 3.4 “Correction factor development” to the manuscript. We showed that adding relative humidity values recorded by the Atmotube Pro sensors using a multiple linear regression model improved the $PM_{2.5}$ data. Similar methods have been used and found useful for other brands of low-cost sensors (Badura et al., 2019; Barkjohn et al., 2021, 2022; Karaoghlanian et al., 2022; Malings et al., 2019, 2020; Raheja et al., 2023). We are aware the correction equation developed may not be applicable across varied environmental conditions since our study is limited to a particular season (September – December) and location. The study was done at a single site in Leeds in an “urban background” environment, and we believe it would be beneficial to edit the title of the manuscript to include this: “Performance evaluation of Atmotube pro sensors for air quality measurements in an urban location”. This will give a guidance to the limitations the suitability of the calibration. Future work may look at using multiple models in a longer-term colocation study and in multiple colocation sites to achieve a more robust calibration.

Line 415 - 3.4: Correction factor development

Many studies have used multiple linear regression (MLR) calibration models that include temperature, relative humidity, and dew point to improve the $PM_{2.5}$ data recorded by low-cost sensors (Badura et al., 2019; Barkjohn et al., 2021, 2022; Karaoghlanian et al., 2022; Malings et al., 2019, 2020; Raheja et al., 2023). In this section we explored the use of MLR using RH and T values to improve Atmotube Pro $PM_{2.5}$ data. We tested using 15 minutes and 1-hour time resolution for the calibration model and assessed the model performance using $R^2$ and RMSE. Given the results of investigating performance of $PM_{2.5}$ concentrations < 100 $\mu g/m^{-3}$ for the 15 minutes average data, we got improved performance as shown in Fig. 5

Table 4: Correction equation forms, the $R^2$ and the RMSE. Best performing calibration equation is indicated as (*) a = slope; i = intercept

| | Equation | $R^2$ | RMSE |
|---|---|---|---|
| | Hourly averaged $PM_{2.5}$ data | | |
| Linear | $S = a_1 \times PM_{2.5} + i$ | 0.86 | 3.38 |
| +RH | $S = a_1 \times PM_{2.5} + a_2 \times RH + i$ | 0.88* | 3.05* |
| +T | $S = a_1 \times PM_{2.5} + a_2 \times T + i$ | 0.87 | 3.20 |
| | 15 minutes average ($PM_{2.5}$ <100 $\mu g/m^3$) | | |
| Linear | $S = a_1 \times PM_{2.5} + i$ | 0.73 | 3.97 |
| +RH | $S = a_1 \times PM_{2.5} + a_2 \times RH + i$ | 0.79* | 3.48* |
| +T | $S = a_1 \times PM_{2.5} + a_2 \times T + i$ | 0.78 | 3.53 |

The addition of RH and temperature values to the model improved the $R^2$ value and decreased the RMSE value. However, the addition of T values only resulted in smaller improvement in $R^2$ and RMSE relative to using RH values. Similar improvement was also gained in higher resolution data at concentrations < 100 $\mu g/m^{-3}$. We note that this result cannot be generalised, since the calibration is done at a single location in an urban background during the winter months. It is possible that warmer seasons or different influences on aerosol composition would require different calibration factors.

We also added the following statement in the conclusion section of the manuscript.

Line 438: Calibration using a multiple linear regression model improved the performance of the Atmotube Pro sensors. $R^2$ improved from 0.86 to 0.88 and RMSE decreased from 3.38 to 3.05 $\mu gm^{-3}$ when accounting for RH values. Future work may look at using multiple models in a longer-term colocation study and in multiple colocation sites to achieve a more robust calibration.

Below are other less major comments:

a) Abstract: quantify "particularly good precision"

The precision metric used in this context is the coefficient of variation which is 18% for hourly $PM_{2.5}$ data and this was mentioned earlier in line 20. In the manuscript this statement has been modified.

Line 20: The result of the $PM_{2.5}$ assessment showed the Atmotube Pro sensors had particularly good precision with a coefficient of variation (CoV) of 28%, 18% and 15% for minutes, hourly and daily $PM_{2.5}$ data averages, respectively.

b) Line 60: Would be good to cite more than one group working on source apportionment with sensors, consider also citing these and potentially others: https://pubs.acs.org/doi/10.1021/acs.estlett.9b00393 (Hagan et al., 2019) https://pubs.acs.org/doi/abs/10.1021/acs.est.1c07005 (Yang et al., 2022) https://pubs.acs.org/doi/10.1021/acsestair.3c00024 (Westervelt et al., 2024) In fact, some of these came before the papers that are cited in the original manuscript.

This has been done and highlighted as tracked changes in the reviewed manuscript.

c) Line 61: The paragraph about LMICs isn't important or relevant, as this paper does colocation in the UK. Sensor colocations should be done locally according to the vast majority of the literature and best practices.

Thank you. We agree the statement is unnecessary and it has been removed as suggested.

d) Line 80: consider citing major sensor intercomparison project publications such as: https://pubs.acs.org/doi/full/10.1021/acs.est.2c09264 (Raheja et al., 2023)

This has been done and highlighted as tracked changes in the reviewed manuscript.

e) Line 94: It is claimed that there are no detailed sensor evaluation studies for Atmotube, but in the next sentence an AQ-SPEC evaluation of the AtmoTube is cited. AQ-SPEC is a global authority on sensor evaluation, so I'm not sure how these two sentences can be consistent.

The statement has been removed in the manuscript.

f) Line 106: "Especially in LMICs". Again, I don't see how this can be a motivation for this paper considering evaluating sensors in Leeds would not be scientifically valid for use in LMICs.

The statement has been removed in the manuscript.

g) Line 118: Purple Air should be capitalized, and the specific details of the device and how long it has been operating should be mentioned.

This has been corrected and highlighted as tracked changes in the reviewed manuscript:

Line 131: The Purple Air sensors contain two Plantower PMS5003 sensors, which record two-minute averaged data. The Purple Air sensor uses a similar principle to the Atmotube Pro sensors described above, based on scattering of laser light. The Plantower sensors also estimate mass of particles with aerodynamic diameters <1 μm, <2.5 μm and <10 μm which are reported as cf_1 and cf_atm which both have channels A and B in the Purple Air dataset. The cf_atm data is displayed on the Purple Air map (BarkJohn, 2021) and this sensor input is the dataset used in this study. The Purple Air sensor was deployed at the colocation site since June 2022.

h) Line 126: Mie Theory, not MIE

This has been corrected and highlighted as tracked changes in the reviewed manuscript.

i) Line 129: Can the Atmotube really size-resolve particles (PM1, PM2.5, PM10), though? Doubtful, unless it is using a true optical particle counter such as the Alphasense. See https://pubs.acs.org/doi/full/10.1021/acs.est.2c09264 (Raheja et al 2023). Related, what kind of "bare sensor" is in the Atmotube? Some candid discussion about the limitations of the device is needed. See also the major comment related to this.

The bare sensor in the Atmotube Pro sensor is a sensirion sensor for PM measurements. The Atmotube device also contains BOSCH BME280 sensors for measuring temperature and relative humidity values. One of the limitations of the Atmotube Pro device is the data retrieval memory with limited history size of 10 days after which data not downloaded would be overwritten.

The additional information regarding the bare PM sensor within the Atmotube Pro device has been modified in the manuscript as follows:

Line 122: Atmotube Pro is a small and lightweight sensor (0.104 kg) classified as a low-cost device ($250) and is commercially available. The Atmotube Pro device has sensirion SPS30 sensors which use a laser scattering principle to radiate and detect suspended particles in an air chamber. A micro fan draws in air through an inlet, and the air passes through the laser beam where the light scattered by the particles is captured by a photodiode. A signal is transmitted to the micro control unit based on Mie theory where a proprietary algorithm processes the data and supplies outputs for the concentration of particulate matter detected ($\mu gm^{-3}$). Atmotube Pro sensors report the estimated mass concentration of particles with an aerodynamic diameter of <1µm ($PM_1$), < 2.5µm ($PM_{2.5}$) and <10µm ($PM_{10}$). In addition to the sensirion sensors for PM measurements, the Atmotube device also contains BOSCH BME280 sensors for measuring temperature and relative humidity values. The sensors also log data every second and store it in memory every minute (Atmotube, 2024). One of the limitations of the Atmotube Pro device is the data retrieval memory with limited history size of 10 days after which data not downloaded is overwritten.

j) Line 173: How was the filter threshold of 50% CoV determined? Is there a possibility to be filtering "real" data, e.g. the extreme events mentioned later (Guy Fawkes night, etc)

Thank you for this feedback. We investigated further and realised the CoV filtering method did eliminate some real data with an overall effect on the performance. Although the filter improved the inter-sensor variability, overall CoV for the raw data was 17.7% while that of the filtered data was 14.8%. The standard deviation however increased from 8.8 to 9.1 for the raw and filtered data respectively. In comparison with the reference, there was indeed an improvement in the accuracy for sensor 6 "S6" after filtering (S6 that had anomalous data with high peaks where other sensors did not record) but the accuracy of other sensors worsened. This is shown in the table below and the comparison of the performance of S6 for the filtered and raw data is highlighted in grey. The decrease in the accuracy of the other sensors when applying filtering was attributed to the possibility of eliminating real data from the other sensors which was less important for sensor 6 which had large anomalous data recorded. The Coefficient of Variation (CoV) filter removes data points when a sensor's reading deviates significantly from others, which can lead to artificially higher averages. Over time, this can distort comparisons across sensors, making it seem like (e.g S1) is consistently reporting higher values when, in fact, the filtering process is amplifying the discrepancy. Thus, CoV filtering can potentially skew results, especially when sensors like S6 already have large anomalies.

Comparison of performance of filtered and raw Atmotube Pro data (hourly average $PM_{2.5}$ data)

| | **Filtered data** | | | | **Raw data** | | | |
|---|---|---|---|---|---|---|---|---|
| | **$R^2$** | **RMSE** | **a** | **b** | **$R^2$** | **RMSE** | **a** | **b** |
| | | | | | | | | |
| S1 | 0.74 | 4.1 | 0.94 | -1.86 | 0.85 | 2.94 | 0.85 | -1.01 |
| S2 | 0.78 | 4.78 | 1.08 | -2.13 | 0.85 | 3.42 | 0.97 | -1.08 |
| S3 | 0.79 | 3.89 | 0.91 | -1.48 | 0.85 | 2.85 | 0.83 | -0.63 |
| S4 | 0.82 | 4.29 | 1.1 | -1.87 | 0.87 | 3.23 | 1.01 | -0.91 |
| S5 | 0.8 | 4.7 | 1.13 | -1.89 | 0.85 | 3.63 | 1.05 | -1.11 |
| S6 | 0.79 | 4.52 | 1.07 | 0.19 | 0.77 | 5.20 | 1.15 | -0.15 |
| S7 | 0.83 | 4.18 | 1.11 | 1.88 | 0.86 | 3.42 | 1.03 | -1.04 |
| S8 | 0.8 | 4.86 | 1.16 | -2.15 | 0.85 | 3.64 | 1.07 | -1.21 |
| Mean | 0.82 | 4.19 | 1.06 | -1.63 | 0.86 | 3.38 | 0.99 | -1.63 |
| PA | 0.85 | 4.86 | 1.37 | -3.49 | 0.85 | 4.79 | 1.37 | -3.47 |

$R^2$ =coefficient of determination, RMSE = root mean square error ($\mu gm^{-3}$), a= slope, b= intercept

Prior to data filtering, there were inconsistencies observed among Atmotube Pro sensors leading to varying readings under same conditions and similar observation was reported by Diez et al., 2024. Our approach of filtering the data however did not improve the performance for all the sensors. We realised this error and we decided to remove this part from the manuscript and conduct the performance without the CoV data filtering method prior to performance analysis. This prompted us to replace Fig.1 in the submitted manuscript with the Performance of $PM_1$, $PM_{2.5}$ and $PM_{10}$ as stated in major comment 1 while Figure 2 is replaced by initial Fig.1 (excluding 1c the filtered data correlation matrix) in the manuscript.

****The entire figures and tables in the corrected manuscript (with tracked changes) were recreated using the raw data. Asides the improved accuracy and reduced error bias, there was very little differences in the plots in the manuscript. All texts and values for the filtered data with the manuscript were also replaced with the unfiltered $PM_{2.5}$ data. The recreated figures and tables are at the end of this document (pages 13-16).**

k) Section 3.1 Data cleaning seems more like methodology than results/discussion.

The “data cleaning” section was highlighting the result of the “Quality assurance filtering”. We agree “data cleaning” is not appropriate as a sub-heading in the results section and this has been removed in the modified manuscript.

l) Line 251: Plantower and PurpleAir should be essentially the same thing. Purple Air reports directly from dual plantower sensors. Also, which version of the Plantower sensor? It is mentioned again in the very next sentence.

Yes, we agree Plantower is the bare PM sensor within the Purple Air device. We made the distinction because the different authors cited in line 251 Zimmerman, (2022) mentioned the use of Purple Air while Badura et al., (2019) mentioned the use of PMS7003 used for analysis. We have made corrections to use Plantower sensors from both authors and this statement has been modified in the manuscript as follows:

Line 251: Previous studies have reported CoV <10% for Plantower sensors (Badura et al.2019; Zimmerman, 2022) while other models of low-cost sensors have also reported a higher CoV > 25% for

Dylos (Carvlin et al., 2017), Plantower and Syhitech (indoor colocation) had CoV > 30% (Zamora et al., 2020).

We used the cf_atm Plantower sensor meant for outdoor data for this study. We averaged the A and B sensors for analysis.

m) Figure 4: It would be better to see these on the same plot, with some smoothing (daily averaging). As the figure stands now it is impossible to glean much information. The accompanying statistics, however, are useful.

Figure 4 has been modified using daily averaged data as shown below.

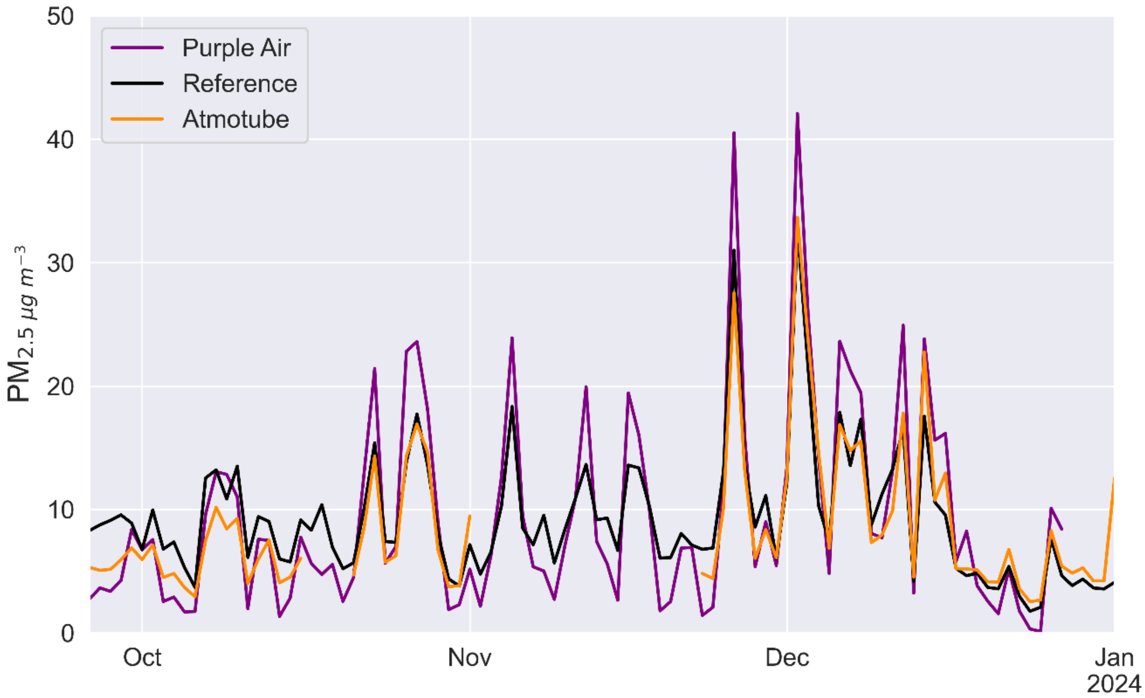
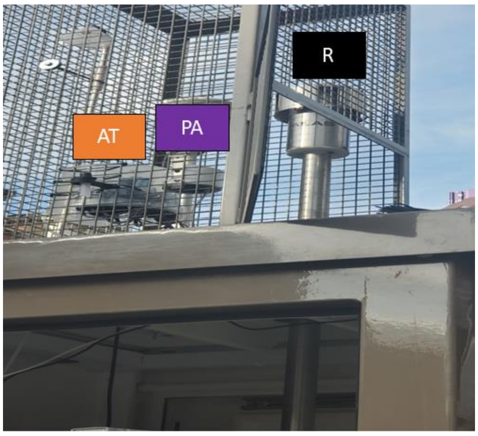



Figure 4 daily averaged $PM_{2.5}$ data

n) Section 3.3: Provide some details about the colocation. Height of deployment, proximity of sensors to reference, and as mentioned previously there is no info about the Purple Air.

Line117: The colocation exercise took place at the Leeds city centre where 8 Atmotube Pro sensors and 1 Purple air sensor were collocated side-by-side (within 3-4 metres) with a Fine Dust Analysis System (FIDAS 200S) reference-grade air quality monitor in an urban background at the Leeds city centre (53°47'51'N, 1°33'8'W) at a height of about 3.5 metres. The duration of the colocation exercise was from September 26, 2023, to January 2, 2024.

We provided an image of the site as a supplementary figure in the manuscript as seen below:




Supplementary Fig.1: AT= Atmotube Pro sensors in a makeshift cover to shield rain; PA= Purple Air sensor; R= FIDAS reference monitor inlet

More information about the Purple Air sensor has been added to the manuscript as suggested in an earlier minor comment (g).

**#REVIEWER 2**

The focused of the study is on the assessment of 8 Atmotube Pro air quality units to ascertain whether they are fit for purpose at different time resolutions. The study collocated these units against a reference-grade instrument at Leeds, UK for a 14-week period. Detailed analysis for the study was well presented to assess the unit's performance which is commendable. However, the paper needs revision.

I expected the following in the paper:

3. The motive of testing the unit's data at different time resolutions. Most of the uncalibrated particulate matter sensors data contains noise below 1-hour resolution. I think testing the level of noise in low-cost sensors data is appropriate for different brands of sensors like Plantower sensor against Alphasense, Sensirion, etc.

Thank you for your feedback. The motive for testing the performance at higher time resolution was because higher resolution data provides more detailed and temporal insights that may capture localized PM variations (such as short time events) that might be missed at lower resolution. However, we found that the accuracy of the Atmotube Pro sensors at higher resolution was much lower probably due to the level of noise as you highlighted. There was an improved performance for the higher resolution (15 minutes) data at concentrations < 100 µg/m$^3$ as shown in Fig 5. We added the following text to the manuscript as our motivation.

Line 106: The aim of the study is to assess the inter-sensor variability and accuracy of Atmotube Pro sensors to provide insight on the reliability and robustness of these sensors $PM_{2.5}$ measurements. By demonstrating a good framework for testing the precision, accuracy, and the reliability of sensors within a sensor network, the results will provide the users a clear understanding of the limitations as well as the confidence in the in-situ $PM_{2.5}$ levels measurements obtained for Atmotube Pro sensors. In addition, we investigated the performance of the sensors at higher time resolution (15 minutes) to test the feasibility of their application in capturing short-time events that may be missed at lower resolution.

4. Description of quality control measures that were taken before the colocation. For example, should the Atmotube units be assembled and charged before deployment, etc. I think adding such information to the study will educate the general public.

The Atmotube Pro came assembled and needs to be charged frequently. The sensor requires a charging time of about 2.5 hours. The battery requires daily charging when set to "always on" mode thus we left the sensor plugged in throughout the entire duration of the study alongside the Purple Air and reference monitor. We added the following text to the manuscript.

Line 156: The Atmotube Pro came assembled and needs to be charged frequently. The sensor requires a charging time of about 2.5 hours. The battery requires daily charging when set to "always on" mode thus we left the sensor plugged in throughout the entire duration of the study alongside the Purple Air and reference monitor. The Atmotube Pro also stores historical data in an onboard flash memory when not connected to a smartphone. This memory can only store 10 days of data after which it is overwritten. The historical data can be transferred to a smartphone during data synchronization each time the sensor is connected to a smartphone. The Atmotube Pro sensors are designed for mobile monitoring, and to protect the sensors from rain at the colocation site, a makeshift cover was used to enclose all the sensors used in the study.

5. Description of the Leeds city centre should be presented to understand the environmental features and sources of pollutants impacting particulate matter measurement by the Atmotube Pro units. Also, description of the study period, was it during winter, fall or summer and some literature review of how meteorological conditions affect low-cost PM sensors.

The site in Leeds city centre is an urban location. Kindly find the image of the study site in response to comment (n) above. This would be added as a supplementary image in the revised manuscript.

Line117: The colocation exercise took place in an ambient environment at Corn Exchange, Leeds city centre (next to a bus stop) where 8 Atmotube Pro sensors and 1 Purple air sensor were collocated side-by-side with a Fine Dust Analysis System (FIDAS 200S) reference-grade air quality monitor in an urban background at the Leeds city centre (53°47'51'N, 1°33'8'W) at a height of about 3.5 metres. The duration of the colocation exercise was done during Autumn from September 26$^{th}$, 2023, to January 1$^{st}$, 2024. The city centre is representative of an ideal urban centre, which included frequent stops from public buses (vehicular emissions).

The comments below can be considered minor:

o) Line 60: LMIC countries need to be removed since the sensors were colocated in Leeds.

Thank you. We agree the statement is unnecessary and it has been removed as suggested.

p) Line 80: Is calibration part of the study?

Calibration was not initially part of the study; however, this was suggested by Reviewer 1 above and it has been included as in the manuscript as suggested by Reviewer 1 in section 3.4 of the corrected manuscript.

q) Line 85: Here, adding some literature concerning work done in the Uk on sensor performance evaluation will be good.

Line 85-88 has been replaced with:

Line 85 Inconsistencies among devices from the same manufacturer might emerge, leading to varying readings under similar conditions. Sensor performance can be highly variable between different devices and end users need to be provided with inter-sensor precision, accuracy, long-term drift and calibration transferability to decide on the right measurement tool for their specific application (Diez et al., 2024).

r) Line 95: Here, you can throw more light on AQ-SPEC standard operating procedure (SOP) which can be found on their website: https://www.aqmd.gov/aq-spec/. Also, (Polidori et al., 2017) and Feenstra et al, 2019 highlighted the AQ-SPEC SOP.

Line 95: AQ-SPEC program is a testing centre for low-cost air monitoring sensors to establish performance standards by which low-cost sensors are evaluated both in the field under ambient conditions and laboratory testing under controlled environmental conditions for sensors measuring criteria pollutants (Feenstra et al., 2019; Polidori et al., 2017).

s) Line 105: Same here too; LMIC needs to be removed.

This has been removed as suggested.

t) Line 120: Here, are you referring to the sensor itself (e.g, Plantower - PMS or Alphasense sensors) or the unit itself?

We were referring the reference FIDAS monitor in Line 120 highlighting where the colocation of the Atmotube, Purple Air sensors and the reference monitor was done.

u) Line 125: Here, you can throw more light on the type of PM sensor in the Atmotube unit.

The Atmotube Pro sensor contains Sensirion SPS30 and the explanation of how the Sensirion sensor functions has been added as tracked changes to the manuscript as in minor comment (i).

v) Line 155: Section 2.2 needs more information. For example, the method that was used to calculate data completeness. Did the units achieve the data completeness threshold before the analysis?

The data completeness for the overall sensors used for the analysis was calculated based on minute-wise data, with data completeness ranging between 73%-84%. The data completeness for this was calculated as shown in Eq. (1). This was done prior to data resampling to lower resolution data averages. This is added to the manuscript as follows.

Line 157. Data completeness as shown in equation 1 for Atmotube Pro sensors is the percentage ratio of minute-wise data available for each sensor and the total number of minutes expected for the study period (Polidori, et al., 2017). This ranged from 73-84% for $PM_1$, $PM_{2.5}$, $PM_{10}$ data.

$$\text{Data recovery} = \frac{N\ valid\ data}{N\ study\ period} * 100 \qquad (1)$$

Where: N valid data = number of valid sensor data points during test period

N study period= total number of data points for the study period

The equation (1) above replaces the eliminated “data filtering equation using CoV” in the initial manuscript as explained in minor comment (j).

w) Line 270: The statement here needs to be rewritten because adding ‘‘it is possible that some sensors were not calibrated as precisely as others’’ sound a bit confusing.

This has been reframed to;

“There were inconsistencies observed among measurements from these sensors leading to varying readings under similar conditions thus contributing to high CoV”.

x) Line 325: Here, does it mean that other studies have focus on calibration of the Atmotube Pro units? You need to clarify this.

This statement has been removed as calibration has been included in the scope of the manuscript as mentioned in major comment (2).

y) Line 390: Here, citation is needed for the statement: "Data were collected from a local weather station rather than from the Atmotube Pros because the RH and temperature sensors in the Atmotube Pro sensors can be influenced by sensor heating when connected to power". Many studies have utilized the internal temperature and relative humidity to calibrate the particulate matter data from the units.

A quote from Zimmerman, 2022 “temperature and relative humidity sensors inside low-cost sensor units are often influenced by their internal environment and thus for these analyses external meteorological data should be used”. This is also observed for Atmotube pro sensors which become slightly heated when plugged in for charging, which may influence the temperature readings of the sensor. The Atmotube Pro sensors were connected to power throughout the period of study. The aim of section 3.4 was to understand how the temperature and relative humidity values influence the ratio of the average of all 8 Atmotube Pro sensors and the Reference $PM_{2.5}$ data for hourly averages ($PM_{2.5}$ ratio) as shown in Eq. (3). We believe the weather station data is more robust, but the sensors temperature and relative humidity data are required for calibration as shown in comment (2). In section 3.4 of the corrected manuscript, it is beneficial to base the calibration on the internal temperature and relative humidity data because it is most readily available to users of the sensors, particularly when deployed for mobile monitoring. The following text was added to the manuscript.

Line 391: Data were collected from a local weather station rather than from the Atmotube Pro sensors themselves, since Zimmerman, (2022) reported that RH and temperature sensors in the low-cost devices can be influenced by sensor heating when connected to power.

**References cited in the author responses.**


Atmotube. (2024, January). *Atmotube* [Website]. Https://Atmotube.Com/Atmotube-Support/Atmotube-Technical-Specification. https://atmotube.com/atmotube-support/atmotube-technical-specification

Diez, S., Lacy, S., Coe, H., Urquiza, J., Priestman, M., Flynn, M., Marsden, N., Martin, N. A., Gillott, S., Bannan, T., & Edwards, P. M. (2024). Long-term evaluation of commercial air quality sensors: An overview from the QUANT (Quantification of Utility of Atmospheric Network Technologies) study. *Atmospheric Measurement Techniques*, *17*(12), 3809–3827. https://doi.org/10.5194/amt-17-3809-2024

Feenstra, B., Papapostolou, V., Hasheminassab, S., Zhang, H., Boghossian, B. D., Cocker, D., & Polidori, A. (2019). Performance evaluation of twelve low-cost PM2.5 sensors at an ambient air monitoring site. *Atmospheric Environment*, *216*, 116946. https://doi.org/10.1016/j.atmosenv.2019.116946

Hagan, D. H., Gani, S., Bhandari, S., Patel, K., Habib, G., Apte, J. S., Hildebrandt Ruiz, L., & Kroll, J. H. (2019). Inferring Aerosol Sources from Low-Cost Air Quality Sensor Measurements: A Case Study in Delhi, India. *Environmental Science & Technology Letters*, *6*(8), 467–472. https://doi.org/10.1021/acs.estlett.9b00393

Molina Rueda, E., Carter, E., L'Orange, C., Quinn, C., & Volckens, J. (2023). Size-Resolved Field Performance of Low-Cost Sensors for Particulate Matter Air Pollution. *Environmental Science & Technology Letters*, *10*(3), 247–253. https://doi.org/10.1021/acs.estlett.3c00030

Polidori, A., Papapostolou, V., Feenstra, B., & Zhang, H. (2017). *Field Evaluation of Low-Cost Air Quality Sensors*. https://www.aqmd.gov/docs/default-source/aq-spec/protocols/sensors-field-testing-protocol.pdf?sfvrsn=0

Raheja, G., Nimo, J., Appoh, E. K.-E., Essien, B., Sunu, M., Nyante, J., Amegah, M., Quansah, R., Arku, R. E., Penn, S. L., Giordano, M. R., Zheng, Z., Jack, D., Chillrud, S., Amegah, K., Subramanian, R., Pinder, R., Appah-Sampong, E., Tetteh, E. N., … Westervelt, D. M. (2023). Low-Cost Sensor Performance Intercomparison, Correction Factor Development, and 2+ Years of Ambient $PM_{2.5}$ Monitoring in Accra, Ghana. *Environmental Science & Technology*, *57*(29), 10708–10720. https://doi.org/10.1021/acs.est.2c09264

Westervelt, D. M., Isevulambire, P. K., Yombo Phaka, R., Yang, L. H., Raheja, G., Milly, G., Selenge, J.-L. B., Mulumba, J. P. M., Bousiotis, D., Djibi, B. L., McNeill, V. F., Ng, N. L., Pope, F., Mbela, G. K., & Konde, J. N. (2024). Low-Cost Investigation into Sources of PM $_{2.5}$ in Kinshasa, Democratic Republic of the Congo. *ACS ES&T Air*, *1*(1), 43–51. https://doi.org/10.1021/acsestair.3c00024

Yang, L. H., Hagan, D. H., Rivera-Rios, J. C., Kelp, M. M., Cross, E. S., Peng, Y., Kaiser, J., Williams, L. R., Croteau, P. L., Jayne, J. T., & Ng, N. L. (2022). Investigating the Sources of Urban Air Pollution Using Low-Cost Air Quality Sensors at an Urban Atlanta Site. *Environmental Science & Technology*, *56*(11), 7063–7073. https://doi.org/10.1021/acs.est.1c07005

Zimmerman, N. (2022). Tutorial: Guidelines for implementing low-cost sensor networks for aerosol monitoring. *Journal of Aerosol Science*, *159*, 105872. https://doi.org/10.1016/j.jaerosci.2021.105872

*****The recreated plots and tables to replace the filtered data plots as explained in minor comment (j) are listed below:**

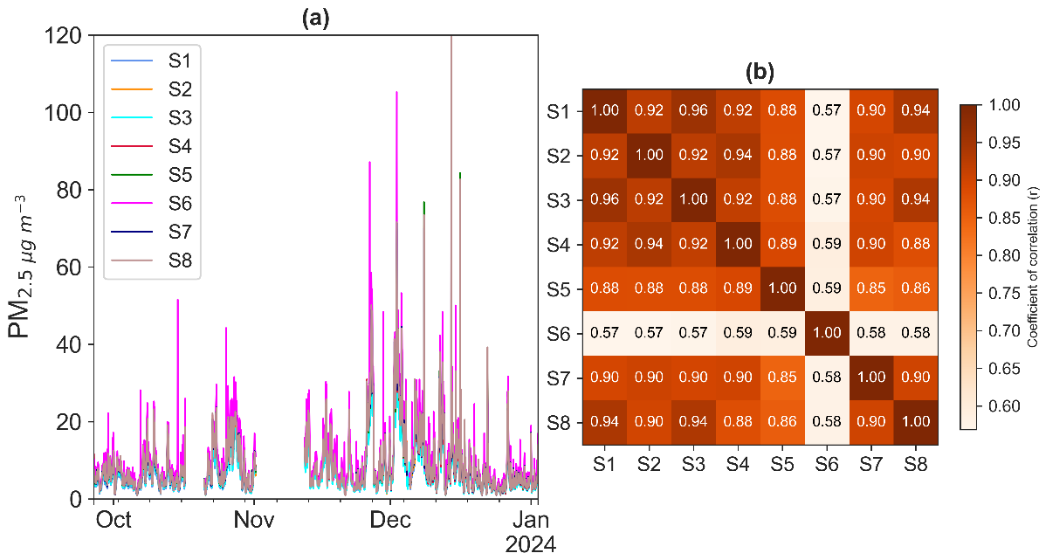



Figure 1: $PM_{2.5}$ data of 8 Atmotube Pro sensors (a) showing the time series of each sensor (hourly average), (b) coefficient of correlation (r) (minute-wise data)

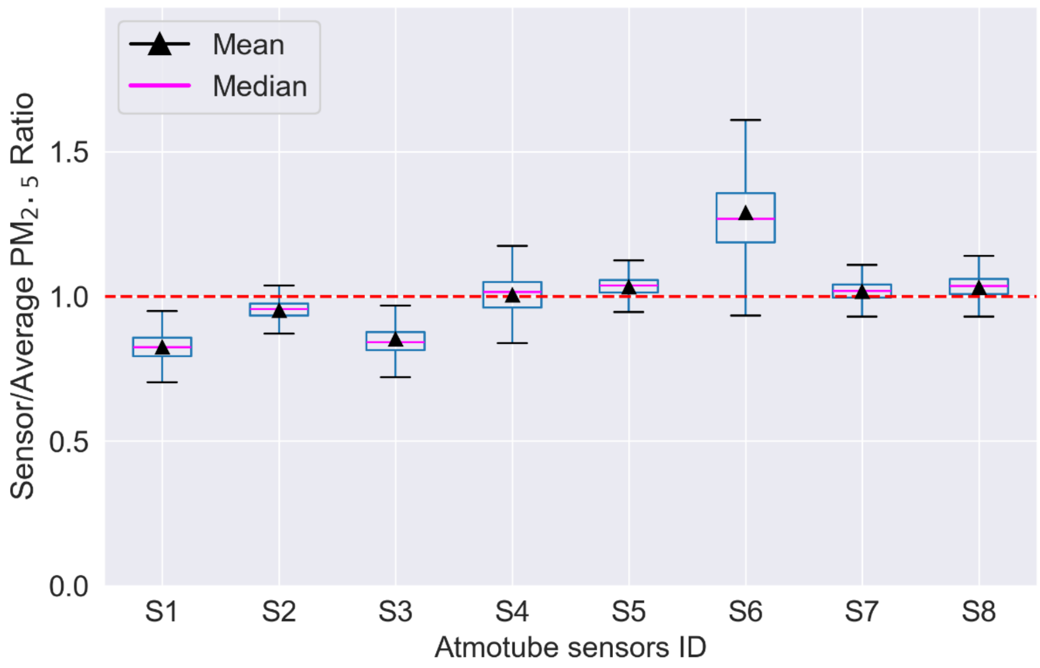



Figure 3: Sensor-sensor precision comparing the ratio of sensor hourly $PM_{2.5}$ values to the 8 Atmotube Pro multi-sensor average as a reference. The whiskers represent the 5th and the 95th percentile. Red dashed line indicates sensor/average ratio of 1 where <1 represents low bias and vice versa.

Table 2: Accuracy metrics using Atmotube Pro and Purple Air sensors in comparison with reference data at 15-minutes and hourly averaging time. $R^2$ = correlation of determination, "RMSE" = root mean square error; "NRMSE" = normalized root mean square error, "a" = slope; "b" = intercept, "S1-S8" = Atmotube Pro sensors, "Mean" = Atmotube Pro sensor average, "PA" = Purple Air sensor.

| | 15-minutes Average $PM_{2.5}$ µgm$^{-3}$ | | | | | Hourly Averaged $PM_{2.5}$ µgm$^{-3}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **$R^2$** | **RMSE µgm$^{-3}$** | **NRMSE (%)** | **a** | **b** | **$R^2$** | **RMSE µgm$^{-3}$** | **NRMSE (%)** | **a** | **b** |
| S1 | 0.51 | 6.05 | 0.02 | 0.65 | 0.92 | 0.85 | 2.94 | 0.02 | 0.85 | -1.01 |
| S2 | 0.51 | 7.02 | 0.02 | 0.75 | 1.13 | 0.85 | 3.42 | 0.02 | 0.97 | -1.08 |
| S3 | 0.50 | 5.91 | 0.02 | 0.64 | 1.23 | 0.85 | 2.85 | 0.02 | 0.83 | -0.63 |
| S4 | 0.53 | 6.86 | 0.02 | 0.78 | 1.35 | 0.87 | 3.23 | 0.02 | 1.01 | -0.91 |
| S5 | 0.51 | 7.45 | 0.02 | 0.81 | 1.27 | 0.85 | 3.63 | 0.02 | 1.05 | -1.11 |
| S6 | 0.48 | 9.73 | 0.02 | 0.99 | 1.42 | 0.77 | 5.20 | 0.03 | 1.15 | -0.15 |
| S7 | 0.52 | 7.11 | 0.02 | 0.79 | 1.27 | 0.86 | 3.42 | 0.02 | 1.03 | -1.04 |
| S8 | 0.51 | 7.56 | 0.02 | 0.82 | 1.22 | 0.85 | 3.64 | 0.02 | 1.07 | -1.21 |
| Mean | 0.54 | 6.74 | 0.02 | 0.78 | 1.23 | 0.86 | 3.38 | 0.02 | 0.99 | -1.63 |
| PA | 0.58 | 8.46 | 0.03 | 1.07 | -0.45 | 0.85 | 4.79 | 0.03 | 1.37 | -3.47 |

Table 3: Overview Performance Summary of reproducibility and accuracy among identical Atmotube Pro sensors using US-EPA guidelines.

| **Performance metrics (US-EPA)** | | **Target values** | **Atmotube Pro sensors ($PM_{2.5}$ values)** | |
|---|---|---|---|---|
| Base Testing | | | 15-minutes | 1-hour average |
| Precision | SD | <5 µgm$^{-3}$ | **Failed** | **Failed** |
| | CoV | <30% | **Passed** | **Passed** |
| Bias | Slope | 1 ± 0.35 | **Passed** | **Passed** |
| | Intercept | -5 ≤ b ≤ +5 | **Passed** | **Passed** |
| Linearity | $R^2$ | ≥ 0.7 | **Failed** using full dataset ($R^2$ 0.48-0.53)<br>**Passed** at $PM_{2.5}$ values <100 µgm$^{-3}$ ($R^2$ 0.72 - 0.75) | **Passed** |
| Error | RMSE | ≤7 µgm$^{-3}$ | **Failed** using full dataset (RMSE 7.6-9.2 µgm$^{-3}$) (RMSE 5.9-9.7 µgm$^{-3}$)<br>**Passed** at $PM_{2.5}$ values <100 µgm$^{-3}$ (RMSE 3.3 - 4.6 µgm$^{-3}$ | **Passed** |
| | NRMSE | ≤ 30% | **Passed** | **Passed** |

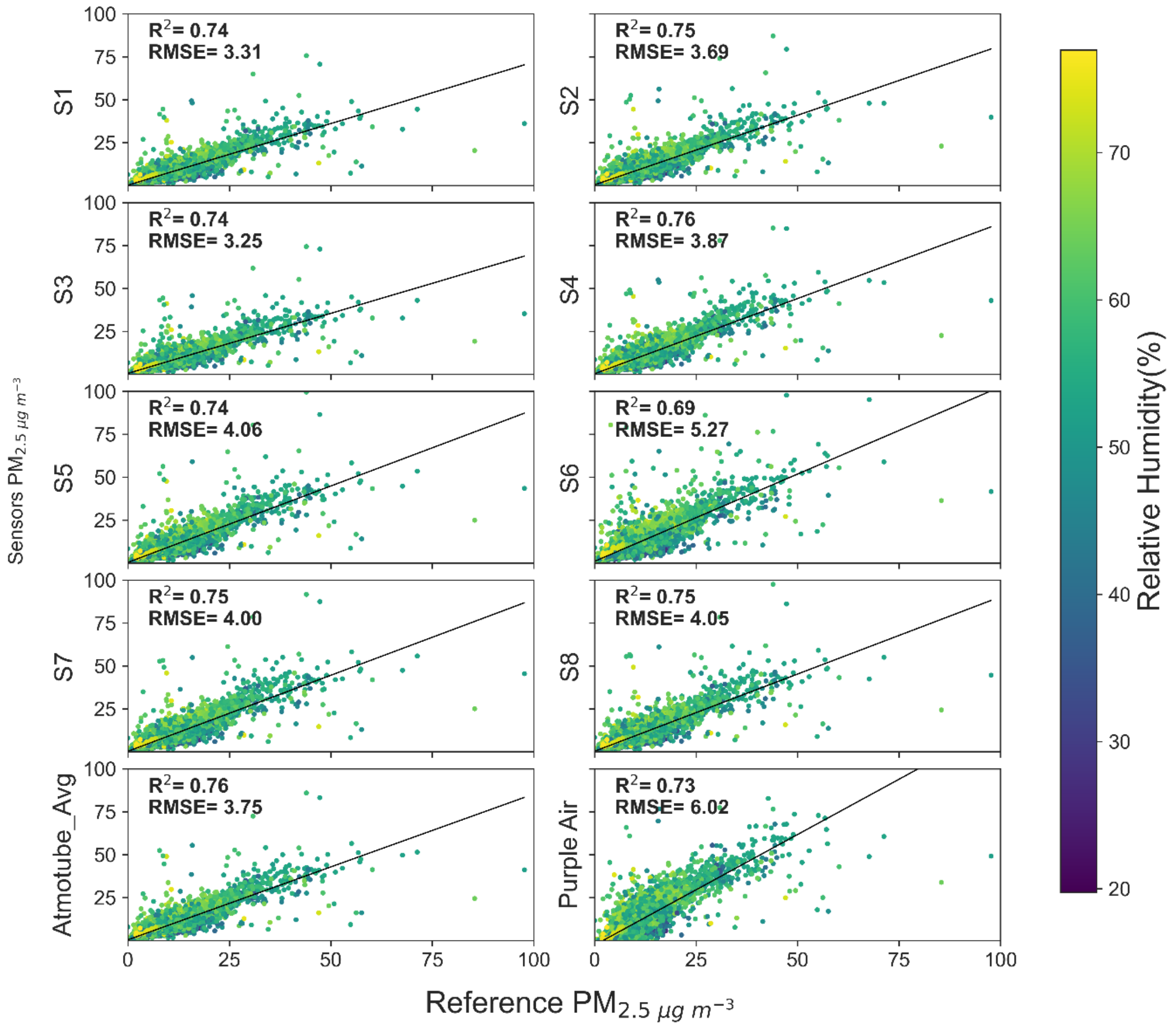



Figure 5: Summary of comparison metrics of each Atmotube Pro sensor, Atmotube Pro Average, Purple Air sensor and reference (15 minutes averaged data) showing $PM_{2.5}$ concentration below 100 μgm$^{-3}$.

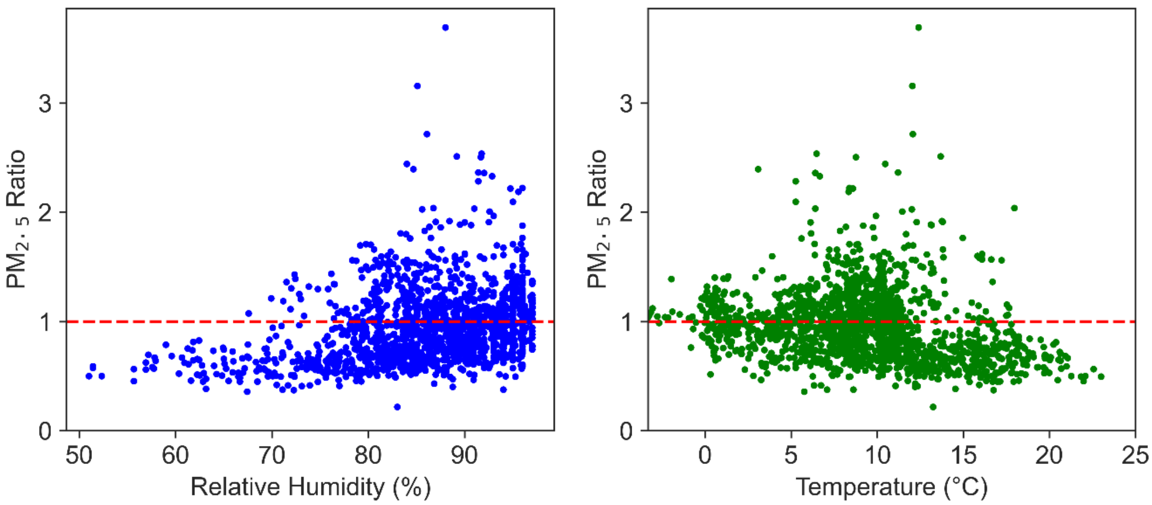



Figure 6: Relationships between (a) relative humidity (RH) and (b) temperature (T) and average Atmotube Pro Sensor/Reference $PM_{2.5}$ ratio.