Dear Editor,

Thank you for agreeing to consider a revision of our manuscript, "Evaluation and Calibration of Clarity Node S Low-Cost Sensors in Lubbock, Texas". We modified and revised the manuscript to address the reviewers' comments as well as to clarify points that they found confusing or unclear.

We would like to thank the two reviewers, Dr. Brayden Nilson and the anonymous reviewer for their helpful comments and suggestions, and many thanks to you for your time and efforts with this revision. In line with the comments and suggestions, we revised the manuscript and made the requested additions and changes. Below are all the comments (in bold) followed by the replies. The parts that are in italics are corrections that are included in the revised version of the paper:

Sincerely,
Karin Ardon-Dryer

Response to Reviewer 1

RC1: 'Comment on egusphere-2025-4300', Brayden Nilson, 23 Oct 2025
This study assess the performance of 4 Clarity Node S sensors collocated with/near an EDM and BAM FEM monitor for PM1, PM2.5, and PM10. They provide a good summary of their study site and instrumentation used, and a detailed description of their correction process. However, I fear their correction model is heavily overfit to the EDM data and needs to be re-evaluated. There is also a concern over the seemingly lack of pre-treatment (QA/QC) for the LCS data – it is well established that these monitors need to have erroneous values removed prior to correcting, typically through comparing the data from replicated internal sensors (ie. https://doi.org/10.5194/amt-15-3315-2022), and this likely explains the relatively low correlation of the raw PM2.5 data.

We thank the reviewer for this detailed and constructive comment. We acknowledge that the original manuscript did not sufficiently describe the QA/QC procedures applied to the low-cost sensor (LCS) data, and we appreciate the opportunity to clarify this in the revised version.

In response, we have substantially expanded the manuscript to explicitly document the full QA/QC workflow applied prior to model development. A new section has been added that describes the identification and removal of erroneous measurements specific to the Clarity Node S sensors, including sensor-specific behaviors and failure modes. Because QA/QC procedures are sensor-dependent, the approach used here differs from those commonly applied to other LCS platforms; this distinction is now clearly explained. We have also incorporated the reference suggested by the reviewer and discussed how its methodology informed our revised QA/QC approach.

To address concerns regarding potential overfitting, we have added detailed information on the temporal separation of training and testing datasets, along with additional justification for the correction model structure. The intent of this study is not to produce a universally transferable correction, but to evaluate performance under a well-defined calibration framework, which is now more clearly articulated.

Regarding the relatively low correlation observed in the raw PM2.5 data, we believe this is primarily attributable to site-specific conditions and particle composition rather than deficiencies in data pre-treatment. The calibration site is dominated by natural dust rather than anthropogenic emissions, which presents known challenges for optical LCS performance. This context has now been explicitly discussed, and we have added a new section outlining the limitations of the method, including implications for deployment in dust-dominated environments.

Overall, the revised manuscript provides substantially greater transparency in QA/QC procedures, model development, and study limitations, directly addressing the concerns raised. Some of the new information was added to the revised manuscript:

2.3 Quality Assurance and Control, Training and Testing
To ensure the validity of the data measured, all parameters measured pass through Quality Assurance (QA) and Quality Control (QC) steps. As part of the LEAP project, a Quality Assurance Project Plan (QAPP) was written and approved, where every month the reference sensor passes through cleaning and evaluation steps that include checking the flow, leak test, and 0 filter count test to verify the instrument performance. For the HI, the units flow was measured before and after collection to ensure the system measures an accurate flow. All filter weight was based on three consecutive weights, with the assumption that the difference between filter weights is less than 5 µg. If the measurement were above this threshold, another weight was taken. To assure the quality of the LEAP units, the raw hourly values measured by each of the four units were examined for measurement errors, hourly values > 1000 µg m^{-3}, which are outside the sensor operational range. No hourly value during the measurement period was above that threshold. Unlike other sensors that might have two Plantower sub-units, allowing for additional QC or QA of the measured values, the Clarity sensor has only one unit in each device. Therefore, such QC or QA could not be performed. Since measurements for the TCEQ BAM-1022 or meteorological data (ASOS) were downloaded from online, they should pass QA and QC before being posted on the website.

This project had one training period and three testing periods. The first test was a comparison between the four sensors. A collocation period of all four Clarity sensors, which took place at AEROS, a research station located on the rooftop of the Electrical Engineering building at Texas Tech University (TTU). Additional information on AEROS can be found in Ardon-Dryer et al. (2022b) and Ardon-Dryer and Kelley (2022). The Clarity Node S sensors were placed on the AEROS filter sampler unit on March 4, 2024, at 13:00 Central Daylight Time (CDT), defined as Local Time (LT). By 14:00 LT, all sensors were deployed at AEROS. Sensors were active at AEROS until May 21, 2024, at 23:00 LT, when the units were removed and prepared for deployment around the city. During these 78 days, the sensors reported the measurements at a default rate of once every 15 minutes. Hourly averages of PM concentrations were retrieved from the Clarity cloud, and then sensors were compared to each other to determine overall sensor network behavior and identify outliers among the Clarity units. On May 23, 2024, according to a *conversation with the company, all sensors' reporting rates were changed to the fastest reporting-time intervals, of 3 to 4-minute intervals.*

Next, from May 24 to June 30, 2024, one Clarity unit (named LEAPS01) was left in AEROS near the reference unit EDM-180, and corrections and calibration for the Clarity units were developed (training time). One Clarity unit (LEAPS02) was placed near the TCEQ (BAM-1022) unit (on May 24, 2024), while the remaining Clarity units (LEAPS41, and LEAPS42) were placed back on AEROS (only on July 2, 2024), near LEAPS01 (as shown in Fig. 1). Comparison between each LEAPS unit to the collocated reference unit was then performed to evaluate the calibration and correction developed.

The first testing period post-calibration took place from July 3 to July 14, 2024, after all three LEAPS units were back in AEROS. During this time, a comparison of the three LEAPS units in AEROS (based on raw and corrected daily $PM_{2.5}$ and PM_{10} values) was made to the EDM-180 and the Harvard impactor, an FRM unit. The second testing period post-calibration took place from July 2, 2024, until February 28, 2025. During this long testing period at each collocation site (AEROS and TCEQ), the reference unit was compared to the $PM_{2.5}$ (raw and corrected) measured by the collocated LEAPS units. Analysis was performed for the entire period, specific months, and for specific events (with high and low $PM_{2.5}$ periods).

4 Limitation

There are several limitations that arise from this work that should be mentioned. This correction may be only effective in locations impacted by dust events or storms (or pollution with large particles) and may not be effective or useful for other pollution types. The correction depended on having measurements of PM_{10}. To develop this correction in other locations, the reference unit should contain both $PM_{2.5}$ and PM_{10} measurements. This means that locations without a reference unit that contain both $PM_{2.5}$ and PM_{10} measurements would not be able to follow this correction. And across the USA, the number of locations that contain both PM sizes is very limited (Ardon-Dryer et al. 2023). After the correction is developed, it is recommended to use the LCS unit PM_{10} values. Unfortunately, in the case of the Clarity Node S sensors, that was not an option, as Clarity Node S sensors were unable to detect these particle size concentrations accurately. Ideally, if the LCS cannot allow the usage of its PM_{10} values to correct the $PM_{2.5}$, correction should be made to the closest reference unit with PM_{10} values. If only one reference unit with PM_{10} available, the correction might be effective only during synoptic dust events that have an impact on a large area, meaning only small differences will be found between neighborhoods (Sandhu et al., 2024; Robinson and Ardon-Dryer, 2024). But it may not be effective during convective dust events when the impact might be localized at a neighborhood level, as shown in Ardon-Dryer (2025). To overcome this issue, as in the case of Phoenix, Arizona, which has multiple PM_{10} sensors, the LCS should be corrected based on the nearest reference sensor. Another limitation is the fact that since the correction depends on the reference unit PM_{10} measurements, times when the reference unit is not active cannot be used, and no calibration will be produced. This could be the case when the reference sensors are down for calibration or maintenance.

The authors do not mention splitting their data into testing/training sets like what Clarity Co. did for the correction they provided– without this there is a very high risk of overfitting, and the presented statistics will be biased. In addition, the EMD observations were included in the correction model, which will result in overfitting and a risk of the EDM observations dominating the corrected value. This is especially visible in Figure 7, where the corrected data nearly perfectly follows the EDM timeseries, in contrast to Figure 9 where the BAM monitor was used instead of the EDM. Due to the overfitting resulting from a lack of train/test splitting and the incestuous inclusion of the EDM data in the correction, I have serious concerns over the efficacy and transferability of the correction presented. The authors must split their data properly and seriously reconsider the inclusion of EDM data within the regression terms for this to be statistically sound.

Thank you for this thorough and important comment. We acknowledge that the original manuscript did not clearly describe how the data were divided into training and testing periods, and we agree that this information is critical for evaluating potential overfitting. In the revised manuscript, we have clarified the temporal separation used to define the training and testing datasets and now explicitly describe this process in the Methods section. In addition, to address concerns regarding overfitting, we conducted additional statistical evaluations to assess model generalization performance. The results of these analyses are now included and demonstrate that the correction does not rely on overfitting to the reference data.

With respect to the inclusion of EDM observations in the correction model and the apparent differences between Figures, we have expanded the discussion to clarify the role of each reference instrument in the calibration framework. Additional analyses were performed using the BAM-1022 as the reference to evaluate the sensitivity of the correction to the choice of monitor. These results have been added to the revised manuscript and show that differences in corrected performance reflect instrument-specific characteristics rather than dominance of the EDM data within the regression. Finally, in response to this comment and related concerns raised by the reviewer, we have added a new section explicitly discussing the limitations of the correction approach, including implications for transferability and reference-monitor dependence.

We believe these revisions substantially improve the transparency and statistical rigor of the correction methodology and directly address the concerns regarding overfitting and model validity.

Some of the new information was added to the revised manuscript:
2.3 Quality Assurance and Control, Training and Testing
To ensure the validity of the data measured, all parameters measured pass through Quality Assurance (QA) and Quality Control (QC) steps. As part of the LEAP project, a Quality Assurance Project Plan (QAPP) was written and approved, where every month the reference sensor passes through cleaning and evaluation steps that include checking the flow, leak test, and 0 filter count test to verify the instrument performance. For the HI, the units flow was measured before and after collection to ensure the system measures an accurate flow. All filter weight was based on three consecutive weights, with the assumption that the difference between filter weights is less than 5 µg. If the measurement were above this threshold,

another weight was taken. To assure the quality of the LEAP units, the raw hourly values measured by each of the four units were examined for measurement errors, hourly values > 1000 µg m^{-3}, which are outside the sensor operational range. No hourly value during the measurement period was above that threshold. Unlike other sensors that might have two Plantower sub-units, allowing for additional QC or QA of the measured values, the Clarity sensor has only one unit in each device. Therefore, such QC or QA could not be performed. Since measurements for the TCEQ BAM-1022 or meteorological data (ASOS) were downloaded from online, they should pass QA and QC before being posted on the website.

This project had one training period and three testing periods. The first test was a comparison between the four sensors. A collocation period of all four Clarity sensors, which took place at AEROS, a research station located on the rooftop of the Electrical Engineering building at Texas Tech University (TTU). Additional information on AEROS can be found in Ardon-Dryer et al. (2022b) and Ardon-Dryer and Kelley (2022). The Clarity Node S sensors were placed on the AEROS filter sampler unit on March 4, 2024, at 13:00 Central Daylight Time (CDT), defined as Local Time (LT). By 14:00 LT, all sensors were deployed at AEROS. Sensors were active at AEROS until May 21, 2024, at 23:00 LT, when the units were removed and prepared for deployment around the city. During these 78 days, the sensors reported the measurements at a default rate of once every 15 minutes. Hourly averages of PM concentrations were retrieved from the Clarity cloud, and then sensors were compared to each other to determine overall sensor network behavior and identify outliers among the Clarity units. On May 23, 2024, according to a conversation with the company, all sensors' reporting rates were changed to the fastest reporting-time intervals, of 3 to 4-minute intervals.

Next, from May 24 to June 30, 2024, one Clarity unit (named LEAPS01) was left in AEROS near the reference unit EDM-180, and corrections and calibration for the Clarity units were developed (training time). One Clarity unit (LEAPS02) was placed near the TCEQ (BAM-1022) unit (on May 24, 2024), while the remaining Clarity units (LEAPS41, and LEAPS42) were placed back on AEROS (only on July 2, 2024), near LEAPS01 (as shown in Fig. 1). Comparison between each LEAPS unit to the collocated reference unit was then performed to evaluate the calibration and correction developed.

The first testing period post-calibration took place from July 3 to July 14, 2024, after all three LEAPS units were back in AEROS. During this time, a comparison of the three LEAPS units in AEROS (based on raw and corrected daily $PM_{2.5}$ and PM_{10} values) was made to the EDM-180 and the Harvard impactor, an FRM unit. The second testing period post-calibration took place from July 2, 2024, until February 28, 2025. During this long testing period at each collocation site (AEROS and TCEQ), the reference unit was compared to the $PM_{2.5}$ (raw and corrected) measured by the collocated LEAPS units. Analysis was performed for the entire period, specific months, and for specific events (with high and low $PM_{2.5}$ periods).

Although overall good agreement was found between the LEASP units and the EDM-180 at AEROS, one might suspect that the usage of the PM_{10} values from the EDM-180 and the fact that the correction was based on the EDM-180 measurements will lead to overfitting of the correction model for $PM_{2.5}$ concentrations. According to Montesinos López et al. (2022), overfitting is defined as a case when the *predicted values match the true observed values in the training period too well, causing what is known as overfitting. In many overfitting cases, the correction model developed will perform very well during the training period, but will perform poorly on new, unseen data (testing period) because it fails to capture the general underlying patterns or changes in $PM_{2.5}$ values (Lever et al., 2016). To examine if the model developed in this study (TTU-calibration) could be defined as overfitting, different statistical tests were performed on the data, including Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE). These values were first calculated for the training period (May 24 to June 30, 2024) using the exact measurements that were used to develop the correction (TTU-calibration, Eq. 1) using the LEAPS01 and the EDM-180. Next, for each month from September 2024 to February 2025 (excluding July and August due to the issue presented above), each raw hourly $PM_{2.5}$ value was examined, compared to the corrected (predicted) and MAD, MAPE, and SMPAPE were calculated for that month. Every month examined during the testing period had low and similar (MAD, MAPE, and SMPAPE) values as those found during the training period, indicating a good fit and not overfitting.*

A comparison between the two reference units (BAM-1022 and the EDM-180) was made for both hourly and daily values that were measured throughout the studied period from May 2024 to February 2025 (Fig. 9). Observation of hourly values indicates that overall, both sites experience similar air quality conditions, except for two very short and localized events, when BAM-1022 (on October 23 at 21:00 LT, and December 01 at 16:00 LT) had an increase of hourly $PM_{2.5}$ values, but these increases were only for one hour. Observation of daily values shows a much better agreement between the two units, as both have similar fluctuations of $PM_{2.5}$ values. However, for both hourly and daily values, it is clear from the box-and-whisker plot that the BAM-1022 measures lower concentrations than the EDM-180 measures. Three times lower compared to the EDM-180 (for both hourly and daily values). These overall low concentrations of the BAM-1022 could explain the low comparison made between the two reference units. Comparison based on hourly values (6580 hours) had R^2 of 0.44, RMSE and MAE of 4.0 and 2.7 $\mu g\ m^{-3}$, respectively, and a slope of 0.7. Daily values (based on 275 days) demonstrated a slight improvement in the comparison, yielding an R^2 of 0.71, RMSE, and MAE of 2.0 and 1.5 $\mu g\ m^{-3}$ (respectively), and a slope of 0.92. Even a comparison between the HI and the BAM-1022 daily values for July 3 - 14, 2024 (shown in Fig. S9), highlights that the BAM-1022 reported overall lower $PM_{2.5}$ concentrations. These lower $PM_{2.5}$ values from the BAM-1022 have been reported in the literature when compared to FRM (Khan et al., 2024). Long et al. (2023) reported that the BAM-1022 underestimates the $PM_{2.5}$ concentrations compared to the FRM unit by ~15%.

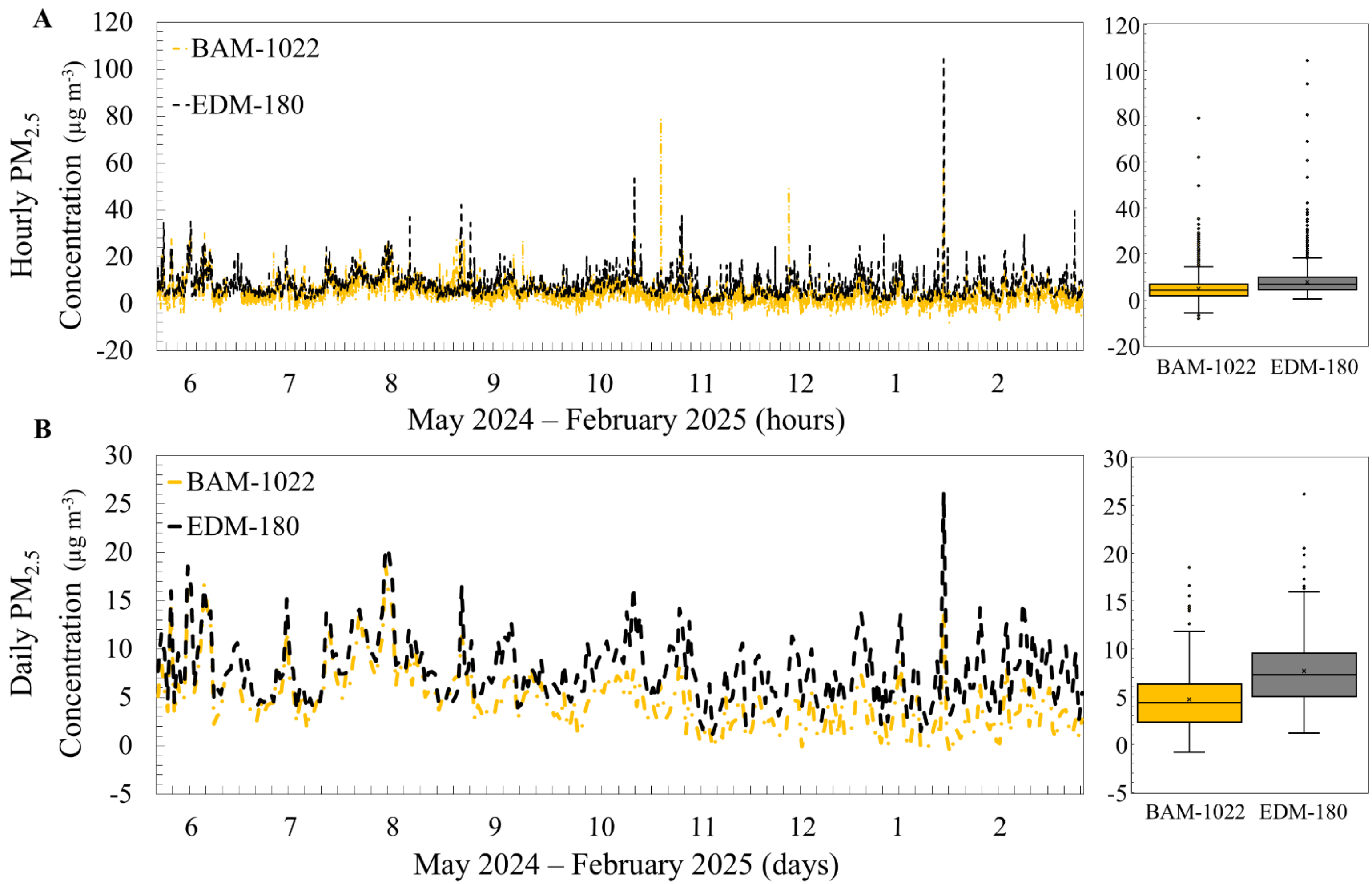

Figure 9: Comparison of $PM_{2.5}$ concentrations from BAM-1022 at TCEQ site (in light orange) and EDM-180 (black) with Box-and-whisker plots comparing the two reference units (A) for hourly and (B) daily values measured for May 24, 2024, to Feb 28, 2025.

Perhaps to find a better agreement between the LEAPS02 to the BAM-1022 unit, the LEAPS02 should be calibrated based on the BAM-1022 and not based on the EDM-180, which is ~8 km away. The same correction period used to correct the LEAPS unit in AEROS was used here (May 24 to June 30, 2024). Three different correction attempts were made to correct LEAP02 based on the BAM-1022. The first was based only on $PM_{2.5}$ values from BAM-1022, the second added the RH measured from the LEAP02, and the third also included the T and RH from the LEAPS02. The coefficients found for each correction, as well as the results of the comparison based on this correction presented in Table S4. Comparison between the corrections of the LEAPS02 $PM_{2.5}$ values just based on the BAM-1022 improved the comparisons, as lower RMSE and MAE values were found, although similar R^2 values were found. Improvement of the correlation between the LEAP02 and BAM-1022 was found when the RH and T with RH were added to the correction. R^2 values improved to 0.72, and both RMSE and MAE values decreased. Next, each of these corrections was implemented on the LEAPS02 data from July 2, 2024, to the end of February 2025, and a comparison between the BAM-1022 and the corrected LEAPS02 unit was made. Very low correlation values ($R^2 \leq 0.3$) and high MAE and RMSE values were found when the LEAPS02 was corrected based on BAM-1022, regardless of the correction type, and even with or without July and August (as shown in Table S5).

To examine the corrections based on the BAM-1022 future, observation of the most severe pollution event during the study period was made. This severe pollution was the dust event of January 17, 2025. Hourly $PM_{2.5}$ values of LEAPS02 based on the different corrections were made to those measured by BAM-1022 as well as the EDM-180 (Fig. S10). It is clear to see that none of the corrections for LEAP02 based on the BAM-1022 units were able to capture the dust event. During the dust (11:00 – 19:00 LT), the BAM-1022 measured on average 2.4 times higher hourly $PM_{2.5}$ values than the LEAPS02 corrected (based on BAM-1022) values. Observations of the peak of the dust (15:00 LT) highlight the low quality of these corrections, as all three corrections based on BAM-1022 detected 2.5 times lower hourly $PM_{2.5}$ values than the BAM-1022 measurements, and 5.7 times lower than those detected by the EDM-180. None of these models developed based on the BAM-1022 were able to detect the dust, indicating that without proper correction, the LEAPS would not have been able to detect the dust particles, which are the main source of pollution in this area (Robinson and Ardon-Dryer, 2024; Ardon-Dryer, 2025). It can be concluded that the correction of LEAPS02 based on the BAM-1022 did not produce sufficient or accurate $PM_{2.5}$ values and should not be used. Perhaps using other corrections developed in the literature for Clarity sensors will result in better comparisons between LEAPS02 and the BAM-1022.

Three different corrections for Clarity Node S units were found in the literature that used FEM hourly $PM_{2.5}$ values, with T and RH measured from the Clarity Node S units (Liu et al., 2022; Raheja et al., 2023; Nobell et al., 2023). Both LEAPS02 and LEAPS01 were corrected based on these corrections for the training period of May 24 to June 30, 2024. Each LEAPS unit was corrected based on the reference sensor it was collocated with. The results of these corrections and comparisons can be found in Table S6. Between BAM-1022 and LEAPS02, there were 788 hours of comparison. Liu et al. (2022) was the only one that produced a reasonable R^2 value of 0.5. The other two corrections had $R^2 \leq 0.3$ and much higher RMSE and MAE values. Between the EDM-180 and LEAPS01, none of these corrections were able to produce a good comparison, and much lower R^2 values were found ($R^2 \leq 0.31$). Even with these low corrections, *these coefficients (in Table S6) were used to correct the entire data set (July 2024 to February 2025). Hourly values from each LEAPS unit were then compared to the reference sensor it was collocated with. Regardless of the period examined (with or without July-August), R^2 values for both cases were ≤ 0.28. Highlighting the issue of using corrections made in a different location that had different meteorological and pollution types. Nilson et al. (2022) also stated that correction models do not perform the same at different locations and should be examined and/or developed per location.*

Since the correcting of LEAP02 based on BAM-1022, and correction from the literature, did not produce any good corrections or agreement between the LEAP02 and BAM-1022, it was decided to examine the LEAP02 from a different perspective. Since the comparison between the BAM-1022 and EDM-180 shows that the areas experience similar air quality conditions (except for very few cases, shown in Fig. 9). A comparison was made between the EDM-180 in AEROS to the corrected LEAPS02 (based on TTU-calibration) in the TCEQ site from September to February (since July and August were problematic). High comparison between the EDM-180 to the corrected LEAPS02 was found, with R^2 of 0.89, RMSE and MAE of 1.7 and 1.1 µg m^{-3}, and a slope of 0.9 (based on 4332 hours of comparison). Even with July and August, high R^2 values were found (R^2 of 0.85; based on 5792 hours of comparison). Observations comparison based on daily $PM_{2.5}$ values between the EDM-180 and the corrected LEAPS02 (based on TTU-calibration) from September to February had even better correlation values (R^2 of 0.93, RMSE and MAE of 0.9 and 0.7 µg m^{-3}; 181 days). Even the three corrected LEAPS units in AEROS (LEAPS01, LEAPS41, and LEAPS42) had a good agreement with the corrected LEAPS02 based on hourly $PM_{2.5}$ values. LEAPS01 and LEAPS41 had a better agreement (R^2 of 0.94, RMSE of 1.3, MAE of 0.68) with LEAPS02 than LEAPS42 had with LEAPS02 (R^2 of 0.79, RMSE of 2.3, MAE of 1.16). If a good agreement is found between the LEAPS units and the EDM-180 across the two locations, perhaps the issue is with the BAM-1022 unit. It should be noted that we did not have any control over the BAM-1022 unit, as it was operated and calibrated by the TCEQ.

4 Limitation

There are several limitations that arise from this work that should be mentioned. This correction may be only effective in locations impacted by dust events or storms (or pollution with large particles) and may not be effective or useful for other pollution types. The correction depended on having measurements of PM_{10}. To develop this correction in other locations, the reference unit should contain both $PM_{2.5}$ and PM_{10} measurements. This means that locations without a reference unit that contain both $PM_{2.5}$ and PM_{10} measurements would not be able to follow this correction. And across the USA, the number of locations that contain both PM sizes is very limited (Ardon-Dryer et al. 2023). After the correction is developed, it is recommended to use the LCS unit PM_{10} values. Unfortunately, in the case of the Clarity Node S sensors, that was not an option, as Clarity Node S sensors were unable to detect these particle size concentrations accurately. Ideally, if the LCS cannot allow the usage of its PM_{10} values to correct the $PM_{2.5}$, correction should be made to the closest reference unit with PM_{10} values. If only one reference unit with PM_{10} available, the correction might be effective only during synoptic dust events that have an impact on a large area, meaning only small differences will be found between neighborhoods (Sandhu et al., 2024; Robinson and Ardon-Dryer, 2024). But it may not be effective during convective dust events when the impact might be localized at a neighborhood level, as shown in Ardon-Dryer (2025). To overcome this *issue, as in the case of Phoenix, Arizona, which has multiple PM_{10} sensors, the LCS should be corrected based on the nearest reference sensor. Another limitation is the fact that since the correction depends on the reference unit PM_{10} measurements, times when the reference unit is not active cannot be used, and no calibration will be produced. This could be the case when the reference sensors are down for calibration or maintenance.*

I was expecting a discussion or conclusions section, but following the results there is just a summary section that repeats the methods and key findings. It could be helpful for the reader to have the large results section parsed into results/discussion/conclusion as is normally done.

We would like to thank the reviewer for notifying us that we forgot to add the word discussion to section 3 of the results. We recognize that the organization of the Results section was not sufficiently clear in the original submission.

In the revised manuscript, we have clarified the section headings to more explicitly reflect that interpretive discussion is included alongside the presentation of results, particularly where findings are compared with previous studies and contextualized. In addition, we have revised the Summary section to serve as a combined Conclusions section, focusing on the main implications and takeaways rather than reiterating methods.

We chose to retain a combined Results and Discussion format, as this structure is commonly used in *Atmospheric Measurement Techniques* and allows for clearer interpretation of complex analyses by discussing results immediately after they are presented. We have, however, revised the text to ensure this organization is clear and intuitive for the reader. We believe these changes improve the readability of the manuscript while aligning with both the reviewer's suggestion and the journal's conventions.

See below for specific line-by-line comments and suggestions.
Specific comments:

1. **Line 9: Use "LCS" acronym for "Low-Cost Sensors" once defined (i.e. "Although [LCS] allow for […]")**

Changes were made per the reviewer's suggestion

2. **Line 16: what is a "LEAP unit" – does this refer to the Clarity Node S sensors?**

Changes were made per the reviewer's suggestion. We changed the name to LEAPS, so it will be easy to distinguish between the project and the sensors themselves.

Changes made to the revised manuscript:
Next, during the training period, one Clarity unit (named the Lubbock Environmental Action Plan Sensor – LEAPS01) was collocated at AEROS with a reference unit, and different calibration tests were

performed for the three PM concentrations measured by the Clarity units (PM_1, $PM_{2.5}$, and PM_{10}, particles with diameters <1, 2.5, and 10μm, respectively).

3. Line 20-21: specific numbers would be useful here, "were very different" is subjective

Information was added to the revised manuscript:
Next, two different testing periods examine the selected calibration developed. At first, over 12 days, a comparison of the corrected LEAPS units was made to a Federal Reference Method unit. During the second testing period, which lasted eight months (July 2024 to February 2025), the calibrated LEAPS units were collocated at two different sites with reference units (EDM-180 and BAM-1022), and a comparison was performed. While one reference unit (EDM-180) showed a good agreement ($R^2 \geq 0.75$ for one LEAPS unit and ≥ 0.91 for the other two LEAPS units), the other reference unit (BAM-1022) had a significantly lower agreement ($R^2 \leq 0.3$). None of the additional attempts to correct the collected LEAPS unit with the BAM-1022 unit were successful, although that LEAPS unit had a very good agreement with the other LEAPS units, as well as with the EDM-180 units that were ~8 km away.

4. Line 25: suggested edit – "High concentrations of PM reduce air quality **and produce negative impacts on** human health**"**

Changes were made per the reviewer's suggestion.

5. Line 26: "Economy" should not be capitalized

Changes were made per the reviewer's suggestion.

6. Lines 33-62: suggest splitting this paragraph into two at essentially line 45; as a reader the transition from FRM to FEM felt unexpected, and readability would be improved with shorter paragraphs. You may need to add a third initial paragraph that introduces the terms FRM and FEM (basically expand on your first sentence of this paragraph) and potentially LCS as well, which would then flow nicely into the next 3 paragraphs, 1 for each of FRM, FEM, LCS.

Per the reviewer's suggestion, the paragraph was split into several smaller paragraphs, and more information was added.

7. Line 47: "and an optical monitor" feels like an after though – suggest expanding on to the same level of detail as the others

Per the reviewer's suggestion, changes were made.
Changes made to the revised manuscript:

FEM methods, including the Beta Attenuation Mass Monitors (BAMs), Tapered Element Oscillating Microbalance (TEOM), and an optical monitor that uses scattered light to measure the PM.

8. Line 62: capitalize "Low-Cost Sensors" to be consistent with the Abstract

Changes were made per the reviewer's suggestion.

9. Line 70: suggest finding a more up to date publication than 2015/16 given that the citations are used to say how there still are many uncertainties

New publications were added and replaced the old ones we had originally

…there are still many uncertainties regarding the reliability and quality of the collected data (Giordano et al., 2021; Nalakurthi et al., 2024).

10. Line 73: suggest adding a transition such as "In addition, LCSs can produce […]"

Changes were made per the reviewer's suggestion.

11. Lines 73-75: The "LCS" acronym is already plural, remove the "s" from "LCSs"

Changes were made to the revised manuscript per the reviewer's suggestion.

12. Line 76: For PM2.5, my experience has been very high correlation (>80%) with collocated FEMs post data cleaning: see table 1 of https://doi.org/10.5194/amt-15-3315-2022

We agree that PurpleAir sensors can show high correlation with collocated FEMs after data cleaning, as demonstrated in the cited study. However, our original statement was intended to be more general, reflecting observations across multiple low-cost sensor types beyond PurpleAir. To clarify this, we have revised the sentence to indicate that while some LCS platforms achieve high correlations, this is not universal across all units.

Changes made to the revised manuscript:
While most LCS show high correlations between sensors (of the same type). Many of the uncalibrated or uncorrected LCS types measurements normally have low correlation and high error when compared to a reference FEM or FRM monitor (Ardon-Dryer et al., 2020; Zaidan et al., 2020; Raheja et al., 2023).

13. Line 81: suggest adding the above publication as a citation as it builds on the Crilley et al method and compares with the Barkjohn et al correction

The reference was added as suggested by the reviewer.

14. Line 96: suggested edit to remove duplicate usage of "climate": "The [area has] a semi-arid climate, […]"

Changes were made to the revised manuscript per the reviewer's suggestion.

15. **Line 124: edit for specificity: "[…] which are then converted to hourly and daily [mean averages] using MATLAB code"**

Changes were made to the revised manuscript per the reviewer's suggestion.

16. **Line 150: suggest removing "At the first step," or providing a paragraph before this that gives a basic outline of each "step" so the reader has context**

Changes were made to the revised manuscript per the reviewer's suggestion.

17. **Line 155: clarify if the measurement every 15 minutes is an instantaneous sample or an integration over the 15 minute period**

Per the reviewer comments, we realize the time interval was not clear; therefore, more information was added to the method section to explain the sampling time

New information added to the revised manuscript per the reviewer's comment:
By default, the Clarity Node S sensor samples every 60 seconds. Once the sampling is complete, the unit sends the collected data to the cloud. The upload period typically lasts one to two minutes. After uploading the data, the unit enters a low-power state to conserve energy. It remains asleep for a set duration before waking up to start the next sampling period. By default, the sleeping period lasts 15 minutes. Per communication with the company, this sleeping period can be reduced to 1 minute, but given the needed time to transmit and upload the data into the cloud, measurements are recorded once every 3-4 minutes (Clarity, 2025).

During these 78 days, the sensors reported the measurements at a default rate of once every 15 minutes. Hourly averages of PM concentrations were retrieved from the Clarity cloud, and then sensors were compared to each other to determine overall sensor network behavior and identify outliers among the Clarity units. On May 23, 2024, according to a conversation with the company, all sensors' reporting rates were changed to the fastest reporting-time intervals, of 3 to 4-minute intervals.

18. **Line 170: combine these first two sentences into one, "Different calculations" is vague, and this sentence is essentially repeated on lines 170-171**

Changes made to the revised manuscript:
Different calculations, including hourly (and daily) average ± standard deviation (SD) values, were made using Excel and MATLAB codes. Daily values were calculated only when >12 hours of measurements were available.

19. Line 170-173: equations for metrics should be defined or cited

We do not think this is necessary, as these are basic statistical tools commonly used by the communities, yet we added references for these calculations to the revised manuscript.

This sentence was added to the revised manuscript:
Additional information on these statistical tools (R^2, RMSE, and MAE) with equations can be found in the literature (Chai and Draxler, 2014; Chicco et al., 2021; Hayward et al., 2024).

20. Line 173: remove capitalization of "Intercept"

Per the reviewer's suggestion, changes were made.

21. Line 173: "best-fit information" is vague, clarify the regression method used (presumably linear regression based on the slop and intercept mentioned)

Changes were made to the revised manuscript per the reviewer's suggestion.
Changes made to the revised manuscript:
To evaluate the similarities and differences between each sensor, different calculations and comparisons were performed using MATLAB and Excel. These include R-squared (R^2), root-mean-square error (RMSE), mean absolute error (MAE) values, as well as the slope and intercept from the linear regression.

22. Line 176-195: suggest adding a table to display this and only referring to key numbers that help with the discussion. It is difficult as a reader to glean useful information from a long list of statistics.

Based on this comment and comment 23, we reduced the number of numbers provided in the text. We added a table to the supplement section that includes all the information. We only kept critical values that could not be seen directly from the table as averages.

These are the changes that were made in the revised manuscript:
Comparisons of the four Clarity units with each other were based on linear regression, where R^2, RMSE, MAE, and the slope values between the sensors were used. Overall, the Clarity units demonstrated good agreement with each other for each of the examined PM sizes (raw concentrations for PM_1, $PM_{2.5}$, and PM_{10}) when 1881 hours were used, as shown in Fig. 2 and Table S1. Comparison between these four Clarity units for PM_1 values resulted in R^2 values that ranged from 0.99 to 1.0, with an average of 0.99 ± 0.003. The average RMSE value was 0.57 ± 0.1 μg m^{-3}, and the average MAE was 0.37 ± 0.08 μg m^{-3}. The slope ranged from 0.93 to 1.19, while the average slope was 1.06 ± 0.11. $PM_{2.5}$ concentration among these four units had an average R^2 value of 0.99 ± 0.002, with an average RMSE value of 0.87 ± 0.12 μg m^{-3}, and an average MAE of 0.54 ± 0.09 μg m^{-3}. The average slope was 1.04 ± 0.11. Comparison of PM_{10} concentration resulted in an average R^2 of 0.98 ± 0.004, an average RMSE value of 1.4 ± 0.26 μg m^{-3},

and an average MAE value of 0.86 ± 0.14 µg m^{-3}. The average Slope value was 0.98 ± 0.15. Comparisons were also performed between the units for T and RH measurements, where >1880 hours were used for each comparison. The average R^2 value for the T measurements between all units was 1.0 ± 0.003. The average RMSE value was 0.49 ± 0.24°C, the average MAE was 0.34 ± 0.18°C, and the average slope was 1.0 ± 0.01. Similar findings were found for the comparison of RH, where the average R^2 value was 1.0 ± 0.000. The average RMSE was 0.94 ± 0.32%. The average MAE was 0.65 ± 0.23%, and the average slope was 0.99 ± 0.01. These results highlighted the fact that the Clarity units were similar to each other. Additional studies found high comparability between Clarity units, with good agreement and high correlation values (Ramiro et al., 2019; Zaidan et al., 2020; Byrne et al., 2024). Based on this agreement, LEAPS01 remained on AEROS, LEAPS02 was moved to the TCEQ site, and LEAPS41 and LEAPS42 returned to AEROS in early July.

23. Figure 2: font is small on the stats presented in each panel, requires zooming to 150%+ to be able to read – including a table like suggested previously could allow excluding this text and referencing the table instead

We apologize for the inconvenience. We knew there was a lot of information in this figure; we were hopeful that all the information could be in a figure to allow readers an easy comparison between the figures and comparisons. Per the reviewer's comment, we removed most of the information and only kept the R^2 values in the figure. All the statistical comparisons were placed in Table S1, according to the reviewer's suggestions.

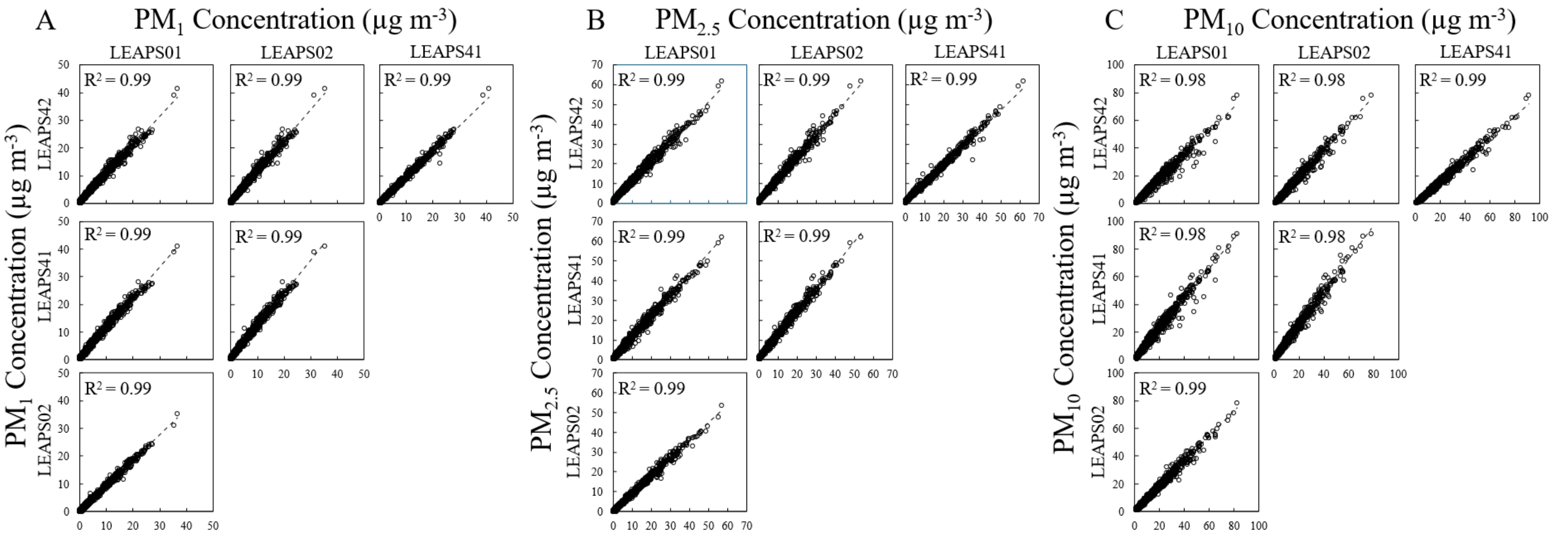

Figure 2: Comparison of RAW PM values between the four clarity units at AEROS, for PM_1 (A), $PM_{2.5}$ (B), and PM_{10} (C). Full statistical comparisons for each figure can be found in Table S1.

Table S1: Details of linear regression output, including R^2, RMSE, MAE, slope, and the number of hours used in each comparison (N) for each comparison. RMSE and MAE units are in µg m^{-3}

Units compared		N (h)	PM_1				$PM_{2.5}$				PM_{10}				T				RH			
			R^2	RMSE	MAE	Slope	R^2	RMSE	MAE	Slope	R^2	RMSE	MAE	Slope	R^2	RMSE	MAE	Slope	R^2	RMSE	MAE	Slope
LEAPS01	LEAPS02	1881	0.99	0.43	0.28	0.93	0.99	0.69	0.43	0.90	0.99	1.14	0.73	0.91	1.00	0.15	0.10	1.00	1.00	0.51	0.36	1.00
LEAPS01	LEAPS41	1881	0.99	0.64	0.44	1.11	0.99	1.00	0.64	1.07	0.98	1.75	1.03	1.11	0.99	0.61	0.43	1.01	1.00	1.10	0.77	0.99
LEAPS01	LEAPS42	1881	0.99	0.63	0.41	1.04	0.99	0.93	0.60	1.01	0.98	1.53	0.91	0.88	0.99	0.71	0.49	1.01	1.00	1.24	0.85	0.98
LEAPS02	LEAPS41	1881	0.99	0.63	0.42	1.19	0.99	0.92	0.59	1.19	0.98	1.58	0.99	1.21	0.99	0.58	0.42	1.01	1.00	1.05	0.75	0.99
LEAPS02	LEAPS42	1881	0.99	0.63	0.40	1.12	0.99	0.90	0.53	1.11	0.98	1.30	0.79	0.97	0.99	0.67	0.47	1.01	1.00	1.18	0.82	0.98
LEAPS41	LEAPS42	1881	0.99	0.45	0.27	0.94	0.99	0.76	0.44	0.93	0.99	1.11	0.71	0.80	1.00	0.22	0.12	1.00	1.00	0.55	0.35	0.99

24. Lines 220-233: it makes sense that the correlation remained 66% (or slightly higher) after applying the LR and MLR given the nature of LR/MLR acting mainly to reduce bias. I have found from working with PurpleAir LCS that the best way to improve correlation is through QA/QC, not bias correction. QA/QC pretreatment was not mentioned in the methods, and could potentially significantly improve these results. Plantower sensors can report unrealistic (>2000 ug/m3) concentrations when sensors fail, or can have reduced sensitivity. These outliers can significantly impact correlation and bias/error. (you elude to this on line 263, but it could be worthwhile to expand the discussion of this). This also may explain the very low correlation observed for PM10 data on lines 366-370.

Thank you for this detailed comment. In response, we have added a new section to the Methods describing the full QA/QC procedures applied to the data, including the identification and removal of erroneous measurements, outliers, and sensor-specific anomalies. We also revised the relevant sentences in the Results to more clearly convey the intended points. Regarding the specific values mentioned, these were not considered errors but rather reflect the inability of the Clarity sensors to detect large coarse-mode dust particles, which is a known limitation of many low-cost sensors, including PurpleAir. During the training period, additional dust events were captured and used for calibration development, ensuring that the correction accounts for these conditions.

Based on this comment, we expanded the discussion to highlight that low-cost optical sensors generally have limited sensitivity to coarse particles, which can reduce correlations with reference instruments during dust events. This expanded explanation now appears in the revised manuscript and provides context for both the observed $PM_{2.5}$ correlation and the lower PM_{10} correlation.

The following information was added to the revised manuscript to address several of the comments that came up by the reviewer for this point.

The Harvard Impactor (HI) unit is an FRM method that collects $PM_{2.5}$ and PM_{10} on filter substrates over 24-hour cycles (midnight to midnight; Marple et al. 1987). The HI samples $PM_{2.5}$ and PM_{10} at a flow rate of 16.7 and 10 $Lmin^{-1}$, respectively, using impactor stages in series with polyurethane foam (PUF) impaction substrates (Lee et al., 2011). The filter samplers contain six HI units that can operate two concurrent setups with three HI units in each; two daily HI's for $PM_{2.5}$ and one for PM_{10}. A 37-mm filter is pre- and post-weighed using a microbalance (XRP2U Microbalance). The HI has been used by many studies to measure gravimetric measurements for $PM_{2.5}$ and PM_{10} (Ayers et al. 1999; Cyrys et al. 2001; Vanderpool et al. 2018). The filters were stored in a climate-controlled room set to EPA standards with a temperature set between 20°C and 23 ℃ and RH between 30% and 40% before and after weight.

2.3 Quality Assurance and Control, Training and Testing

To ensure the validity of the data measured, all parameters measured pass through Quality Assurance (QA) and Quality Control (QC) steps. As part of the LEAP project, a Quality Assurance Project Plan

(QAPP) was written and approved, where every month the reference sensor passes through cleaning and evaluation steps that include checking the flow, leak test, and 0 filter count test to verify the instrument performance. For the HI, the units flow was measured before and after collection to ensure the system measures an accurate flow. All filter weight was based on three consecutive weights, with the assumption that the difference between filter weights is less than 5 µg. If the measurement were above this threshold, another weight was taken. To assure the quality of the LEAP units, the raw hourly values measured by each of the four units were examined for measurement errors, hourly values > 1000 µg m^{-3}, which are outside the sensor operational range. No hourly value during the measurement period was above that threshold. Unlike other sensors that might have two Plantower sub-units, allowing for additional QC or QA of the measured values, the Clarity sensor has only one unit in each device. Therefore, such QC or QA could not be performed. Since measurements for the TCEQ BAM-1022 or meteorological data (ASOS) were downloaded from online, they should pass QA and QC before being posted on the website.

This project had one training period and three testing periods. The first test was a comparison between the four sensors. A collocation period of all four Clarity sensors, which took place at AEROS, a research station located on the rooftop of the Electrical Engineering building at Texas Tech University (TTU). Additional information on AEROS can be found in Ardon-Dryer et al. (2022b) and Ardon-Dryer and Kelley (2022). The Clarity Node S sensors were placed on the AEROS filter sampler unit on March 4, 2024, at 13:00 Central Daylight Time (CDT), defined as Local Time (LT). By 14:00 LT, all sensors were deployed at AEROS. Sensors were active at AEROS until May 21, 2024, at 23:00 LT, when the units were removed and prepared for deployment around the city. During these 78 days, the sensors reported the measurements at a default rate of once every 15 minutes. Hourly averages of PM concentrations were retrieved from the Clarity cloud, and then sensors were compared to each other to determine overall sensor network behavior and identify outliers among the Clarity units. On May 23, 2024, according to a conversation with the company, all sensors' reporting rates were changed to the fastest reporting-time intervals, of 3 to 4-minute intervals.

Next, from May 24 to June 30, 2024, one Clarity unit (named LEAPS01) was left in AEROS near the reference unit EDM-180, and corrections and calibration for the Clarity units were developed (training time). One Clarity unit (LEAPS02) was placed near the TCEQ (BAM-1022) unit (on May 24, 2024), while the remaining Clarity units (LEAPS41, and LEAPS42) were placed back on AEROS (only on July 2, 2024), near LEAPS01 (as shown in Fig. 1). Comparison between each LEAPS unit to the collocated reference unit was then performed to evaluate the calibration and correction developed.

The first testing period post-calibration took place from July 3 to July 14, 2024, after all three LEAPS units were back in AEROS. During this time, a comparison of the three LEAPS units in AEROS (based on raw and corrected daily $PM_{2.5}$ and PM_{10} values) was made to the EDM-180 and the Harvard impactor, an FRM unit. The second testing period post-calibration took place from July 2, 2024, until February 28, 2025. During this long testing period at each collocation site (AEROS and TCEQ), the reference unit was compared to the $PM_{2.5}$ (raw and corrected) measured by the collocated LEAPS units. Analysis was performed for the entire period, specific months, and for specific events (with high and low $PM_{2.5}$ periods).

It was found that one of the dust events (on May 25-26) had the biggest difference in $PM_{2.5}$ concentrations between the LEAPS01 and the EDM-180 (of 0.96 μg m^{-3} on average). This specific event lasted for 15 hours (May 25 at 15:00 LT until May 26 at 5:00 LT) and contained high PM values (PM_{10} and $PM_{2.5}$ reached concentrations of 159 and 35 μg m^{-3}). It should be noted that there were additional dust events during this period, with PM_{10} and $PM_{2.5}$ concentrations up to 132 and 25.3 μg m^{-3} that were included in the analysis. The 15 hours of May 25-26 were removed from the data, and the MLR was run again, now with 897 hours instead of 912. Removal of these 15 hours improved the regression values. R^2 increased to 0.95, RMSE reduced to 1.26 μg m^{-3}, and MAE dropped to 0.86 μg m^{-3} (Fig. 3M). The difference between corrected $PM_{2.5}$ and those measured by the EDM-180 also improved to an average of 0.86 μg m^{-3}.

The first testing period of the correction was an intercomparison for daily values between the HI, EDM-180, and the three LEAPS units on AEROS from July 3 to July 14, 2024. Daily values were calculated based on the hourly values measured by the EDM-180, and the three LEAPS units (LEAPS01, LEAPS41, and LEAPS42) for raw, TTU-calibration, and CC calibration. Since there were two daily filters for $PM_{2.5}$ concentrations, an average between the two filters was performed as long as the difference between the two filters was less than 1.5 μg m^{-3}. Since July 4th had higher differences between the two filters, this date was removed from the analysis, leaving us with 11 days of comparison. The daily average for $PM_{2.5}$ concentrations from the HI ranged from 3.1 ± 0.4 up to 9.9 ± 0.4 μg m^{-3}, EDM-180 daily values for the same time range from 3.6 ± 1.2 up to 13.8 ± 5.5 μg m^{-3}. The raw daily $PM_{2.5}$ average concentrations for LEAPS01 ranged from 1.6 ± 0.9 up to 5.4 ± 5.3 μg m^{-3}, for LEAPS41, and LEAPS42, the raw values ranged from 1.9 ± 0.9 up to 6.5 ± 5.8 μg m^{-3}, and 2.5 ± 1 up to 6.9 ± 5.7 μg m^{-3}, respectively. Higher daily values were measured after the TTU-calibration and the CC calibration. After the CC calibration, daily $PM_{2.5}$ average concentrations for LEAPS01, LEAPS41, and LEAPS42 ranged from 5.3 ± 0.8 up to 12.1 ± 4.7 μg m^{-3}, from 5.5 ± 0.9 up to 12.8 ± 4.8 μg m^{-3}, and from 2.2 ± 0.7 up to 6.0 ± 2.7 μg m^{-3}, respectively. After the TTU-calibration, daily $PM_{2.5}$ average concentrations for LEAPS01, LEAPS41, and LEAPS42 ranged from 3.5 ± 1.4 up to 7.9 ± 2.9 μg m^{-3}, from 3.8 ± 1.5 up to 8.2 ± 2.8 μg m^{-3}, and from 3.9 ± 1.4 up to 8.5 ± 2.9 μg m^{-3}, respectively.

Observations of daily values (average ± SD) show similar values for most days between the HI, the EDM-180, and the different LEAPS units across the different corrections (raw, TTU-calibration, and CC calibration), as shown in Fig. S2. In order to get a better understanding of the similarities and differences between the calibration, a comparison was performed between these daily values. Overall, the EDM-180 had a good agreement with the HI units. FEM/FRM ratio, as described in Khan et al. (2024), had an average ratio of 1.0 ± 0.2, indicating a good correlation between the units. Statistical comparison between the units (HI and EDM-180) had an R^2 of 0.81, with RMSE and MAE values of 1.2 and 1.1 μg m^{-3}, respectively. It is possible that a larger number of days would lead to a better comparison between the units. Both EDM-180 and HI daily values were then compared to the LEAPS daily values. A summary of this comparison, including R^2, RMSE, MAE, and slope, is presented in Table S2. Overall, all three raw LEAPS units had very low R^2 values and high RMSE and MAE values. Correction based on CC calibration had a good agreement with the EDM-180 (average R^2 of 0.85, RMSE and MAE of 0.6 and 0.5 μg m^{-3}, respectively), but very low compared to the HI unit (average R^2 of 0.51, RMSE and MAE of 1.1 *and 1.0 μg m^{-3}, respectively). On the other hand, similar and overall good agreement was found between the TTU-calibration LEAPS units to the HI (average R^2 of 0.7, RMSE and MAE of 0.6 and 0.5 μg m^{-3}, respectively) and to the EDM-180 (average R^2 of 0.79, RMSE and MAE of 0.5 and 0.5 μg m^{-3}, respectively) measurements. Combining the comparison between the two reference units to the LEAPS units indicates a better comparison for the TTU-calibration (average R^2 of 0.74, RMSE and MAE of 0.6 and 0.5 μg m^{-3}, respectively) compared to the CC calibration (average R^2 of 0.68, RMSE and MAE of 0.8 and 0.7 μg m^{-3}, respectively). Highlighting the efficiency of the TTU-calibration. It should be noted that these days were part of the development of the CC calibration, which could explain the good comparison to the EDM-180.*

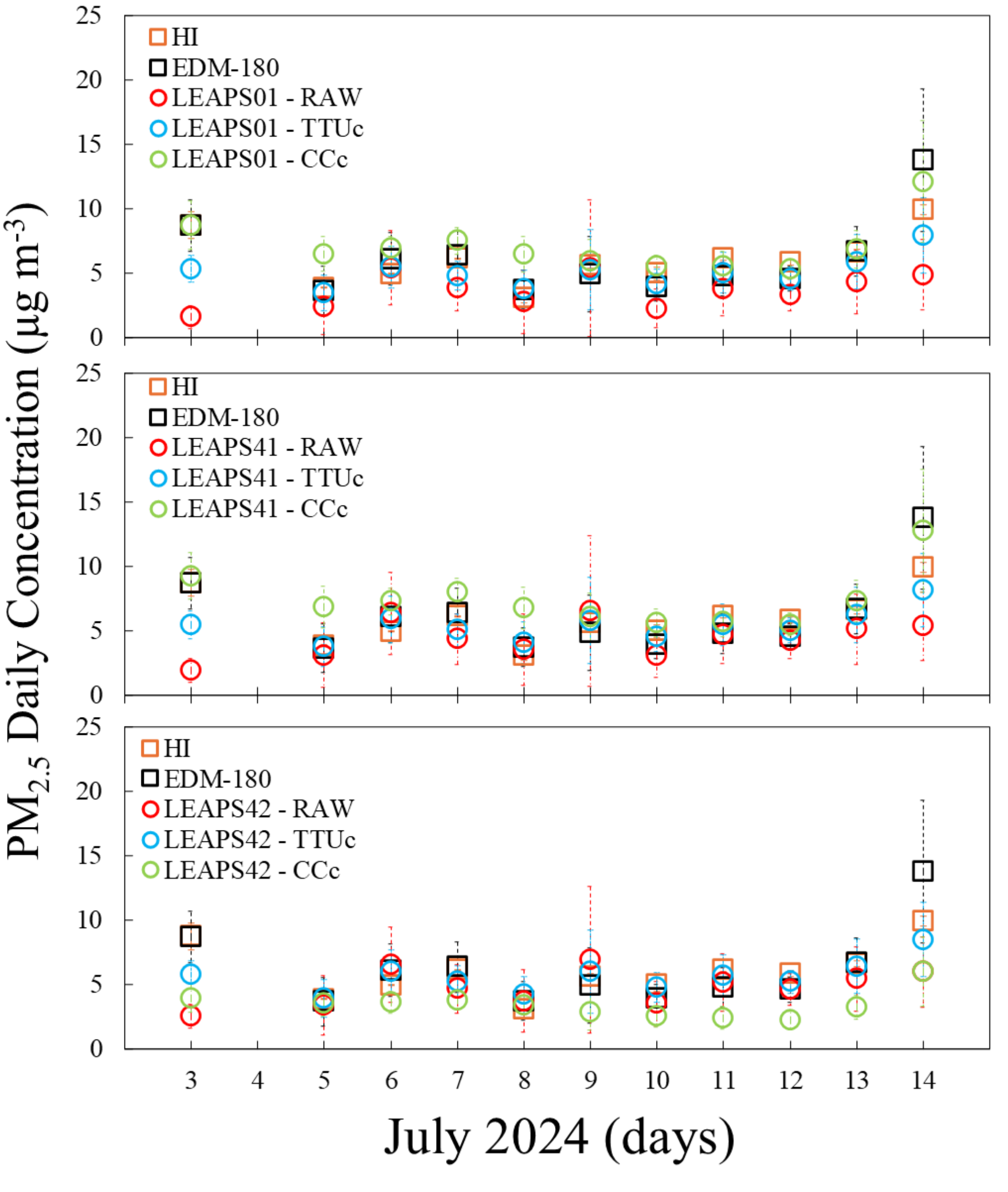

Figure S2: Comparison of daily $PM_{2.5}$ concentrations between Harvard impactor (orange), EDM-180 (black), and the three LEAPS units for raw (red), TTU-calibration (TTUc; light blue), and Clarity Co. calibration (CCc; light green) for July 2024.

Since there were daily filter measurements from the Harvard impactor for PM_{10}, a comparison of daily PM_{10} values was made between the HI, EDM-180, and raw LEAPS units that were active on AEROS during July 3-14, 2024. It should be noted that since only one PM_{10} filter was collected each day, no SD values could be calculated. Results of this comparison are presented in Fig. S3. HI and EDM-180 had a good agreement with each other (for these 12 days), with an average difference of ~0.7μg m^{-3}. Comparison between the two had an R^2 value of 0.91, an RMSE of 2.24 μg m^{-3}, an MAE of 1.75 μg m^{-3}, and a slope of 1.08. Next, a comparison was made between the HI and EDM-180 to the three LEAPS units (LEAPS01, LEAPS41, and LEAPS42). Overall, the LEAPS unit measured much lower PM_{10} daily values than those measured by the HI and EDM-180 (on average, lower by 2.2 and 2.4 times, respectively). Comparison of these 12 days between the three LEAPS units to the two reference units resulted in a very low R^2 value (average R^2 value of 0.32 for HI and 0.27 for EDM-180). Highlighting the inability of the Clarity sensor to detect the PM_{10} concentrations. This comes as no surprise, as previous studies have found that the Clarity sensor or the PMS5003 did not respond to variations in PM_{10} concentrations, regardless of high or low PM_{10} concentrations, producing very high uncertainties, with a combination of bias and noise (Demanega et al., 2021; Molina Rueda et al., 2023).

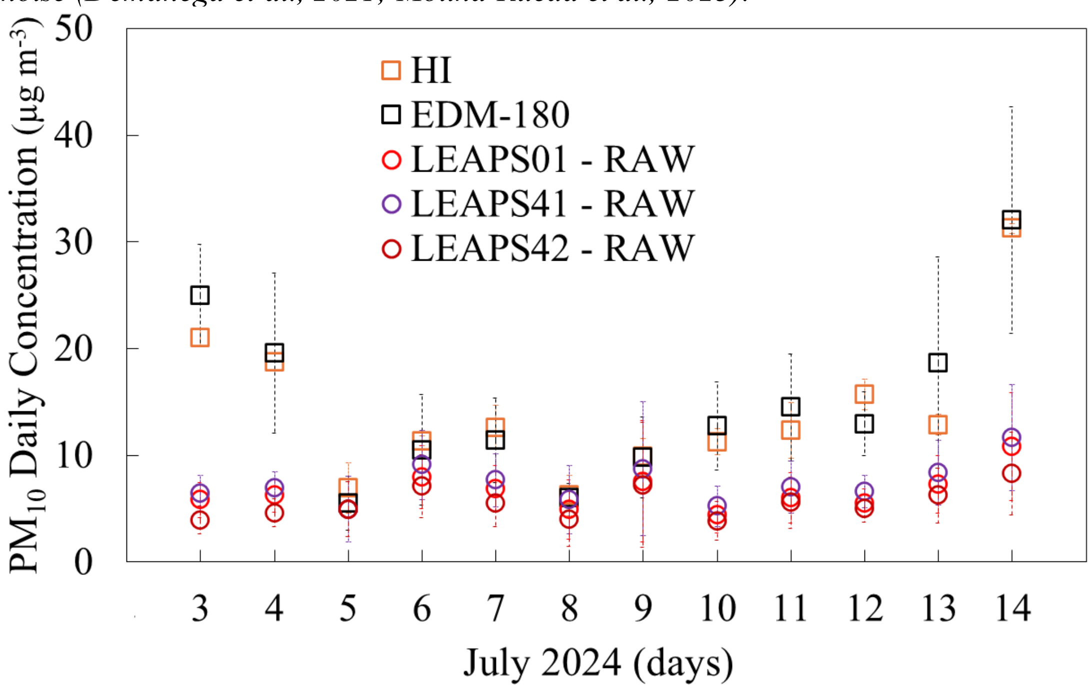

Figure S3: Comparison of daily PM_{10} concentrations between Harvard impactor (orange square), EDM-180 (black square), and the three raw LEAPS units (different colored circles) for July 2024.

25. Line 245: RMSE is not a normalized metric, so it makes sense that a dramatic difference between studies can exist. It is possible that the Nobell et al (2023) study just had higher concentrations on average than this study.

Based on the comment, the sentence was removed from the revised manuscript. Yet studies have used the RMSE as a method to compare between studies (e.g., https://doi.org/10.3390/atmos15121523)

26. Lines 248 and 256: "wondering" is something to be done in the discussion, to me this feels out of place in the results

As mentioned earlier, the result section also contains the discussion part. Regardless, both sentences were changed, and the word wondering was removed from the revised manuscript. Per this comment and comments from the other reviewer, we rewrote any section that had similar wording.

Changes made to the revised manuscript:
Since this study was performed in a semiarid area, which experiences dust events, it is possible to assume that it could impact on the calibration.
It was speculated that the usage of PM_{10} values measured by EDM-180 would improve the correction of $PM_{2.5}$ for Clarity units.

27. Line 267: unclear what this means: "The MLR of this calibration was corrected using the following equation"

The statement was changed to reflect the uncertainty. These are the changes made in the revised manuscript:
The final correction (named TTU-calibration) was selected to correct the LEAPS $PM_{2.5}$ values units:

28. 1 - 3: given that your calibration depends on observation data from the EDM monitor, how will this be transferable to other Clarity sensors? They would need to be colocated with an EDM to be able to apply this in real time, and if you have an EDM why would you setup a Clarity at the same location operationally? I have concerns about overfitting as a result of this as well, the improved performance relative to EDM could just be the result of the regression relying on the EDM data itself.

Thank you for raising this important question regarding transferability and the role of the EDM data in the calibration framework. We agree that these points require careful clarification. As noted in the manuscript, this study represents the initial calibration phase of a larger deployment involving 42 Clarity Node S sensors. Because these sensors had not previously been evaluated in a dry, dust-dominated environment, collocation with a reference-grade instrument was a necessary first step to assess performance and develop an appropriate correction. The inclusion of EDM observations was used for model development and evaluation purposes, rather than as a requirement for operational, real-time applications. The intent of the correction is not that each Clarity sensor must be continuously collocated with an EDM. Instead, the calibration is designed to be transferable to other Clarity units deployed within the same airshed, particularly during regionally driven synoptic dust events. In such cases, spatial variability in particle composition and size distribution is relatively limited, making a single-site reference calibration applicable across the network.

The improved performance relative to the EDM does not arise from the corrected signal relying directly on contemporaneous EDM observations, but from leveraging PM10 information to better represent coarse-mode dust particles that are poorly captured by many low-cost sensors in arid environments. As demonstrated in the manuscript and supported by previous studies, corrections that exclude coarse-mode particle information tend to perform poorly under dust-dominated conditions. The role of PM10 in this context is therefore a physical, not statistical, necessity, and similar approaches have been explored in recent AMT literature (https://doi.org/10.5194/amt-16-2455-2023).

We acknowledge that this approach has limitations, including reduced applicability outside dust-dominated regimes or in locations with substantially different source characteristics. These limitations are now explicitly discussed in a new section of the revised manuscript. While the method is not intended to be universally transferable, our results indicate that it provides practical and effective correction for Clarity sensors operating in arid regions where natural dust is the dominant PM source.

29. Line 287: “represents the interception” – this is vague, interception of what?

The sentence was changed to make it clearer. We also made similar changes in other locations in the text. Since reviewer 2 had a similar comment, we ended up removing this part and adding the information into the equation.

30. Figure 4: the colours used are difficult to differentiate especially with the dashed lines. Suggest splitting into 3 panels, each one comparing one of the LEAP timeseries with the EDM

We were hoping to use a similar color scheme in the paper throughout all the figures, but per the reviewer's comment, the color of the lines in Figure 4 was changed. We also had to change the colors in all the other figures in the revised manuscript. This is the new figure 4:

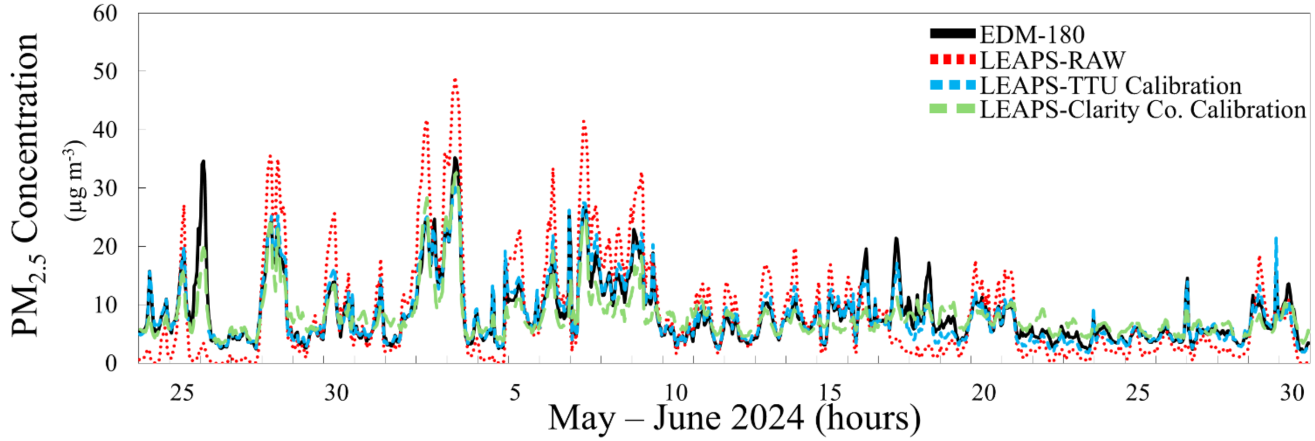

Figure 4: Comparison of hourly averaged $PM_{2.5}$ values for EDM-180 (black) and LEAPS01 raw (red), TTU calibration (light blue), and Clarity Co. calibration (light green) for May to June 2024.

31. Line 369: "the slope improved to 1" is not necessary to say, given that that is what LR does

Yes, we agree with the reviewer, yet there have been cases with other sensor types (and in the literature) where the slope did not improve to 1; Per the reviewer's comment, we removed this part from the sentence.

32. Line 421: if the event was only for 1 hour, how were multiple EDM hourly mesaurements taken to produce the mean+/- SD?

We thank the reviewers for this comment, as it made us understand that we did not provide the information that the EDM-180 provides measurements every minute; these were averaged every hour.
This information was added to the methods section:
The unit provides measurements every minute, operates at a flow rate of 1.2 L min^{-1} and can count up to 3 million particles per Liter.

33. Line 426: the fact that correcting the LCS made them detect the dust event when they did not before points to the overfitting resulting from including the EDM data in the correction

Thank you for this comment. We agree that this behavior warrants careful interpretation, and we appreciate the opportunity to clarify the underlying mechanism. The improved ability of the corrected LCS data to capture the dust event does not result from the regression directly reproducing the EDM signal, but rather from addressing a known physical limitation of many optical low-cost sensors in arid, dust-dominated environments. These sensors often exhibit reduced sensitivity to coarse-mode particles and, under certain conditions, even to fine-mode mass during dust events, which can lead to an apparent failure to detect such episodes in the raw data.

The correction developed in this study explicitly accounts for this limitation by incorporating information relevant to coarse particle loading, allowing the Clarity sensors to respond to dust events in a manner that is physically consistent with reference observations. To further evaluate whether this behavior reflects overfitting, we added additional analyses and comparisons against an independent FRM reference instrument, which are now included in the revised manuscript. These results demonstrate that the corrected LCS response during dust events is not solely driven by the EDM data. We also recognize that the use of reference-based corrections has inherent limitations, and we have therefore added a dedicated section discussing these constraints and the conditions under which the correction is expected to perform reliably. We believe these additions address the concern while providing a clearer interpretation of the observed behavior.

34. Lines 453-454: similar to the previous comment, the fact that after correcting the concentrations went from ~10 ug/m3 to ~100 ug/m3 shows how the EDM observations are dominating the corrected values, which is clearly visible to me in Figure 7

We understand the concern that the magnitude of the change after correction could be interpreted as the EDM observations dominating the corrected values, particularly as illustrated in Figure 7. However, we believe this increase reflects the inability of the raw low-cost sensor measurements to adequately represent dust-dominated particulate matter rather than undue influence from the EDM data. In arid environments, coarse particles can contribute substantially to total PM mass, while optical LCS often under-respond to these particles, resulting in artificially low raw concentrations during dust events. The correction is designed to address this limitation, which can lead to large adjustments when dust is present.

As shown in the manuscript, the corrected data capture both high-concentration dust events and low-concentration clean periods, with consistent performance across multiple figures (Figures 7, S2, and S3). This behavior indicates that the correction is not simply reproducing the EDM time series but is responding appropriately across a range of conditions. We acknowledge that this approach has limitations and that large corrections highlight the importance of careful application in dust-dominated regions. These considerations are now explicitly discussed in the limitations section added to the revised manuscript.

35. Lines 489-491: this reads a bit colloquially – specifically “puzzled” and “seem to be off”

Per the reviewer's comment, the word was replaced. Changes made to the revised manuscript:
This observation was unclear and there was no explanation for what caused the July and August deviations.

36. Section 3.3.2: when switching to the BAM-1022 unit for comparison, was the same regression that was fit to the EDM monitor used? IF so, that would explain the poorer than expected performance given the over-fitting with the EDM data I have previously mentioned and the poor correlation between the BAM and EDM monitors you note on Line 562

We have revised this section of the manuscript to explicitly describe this distinction and to clarify the calibration and evaluation steps used for each reference monitor. The reduced performance observed when compared against BAM-1022 is influenced by multiple factors. As noted in the manuscript, the correlation between the BAM and EDM monitors is lower during dust-influenced periods (and non-dust time), reflecting differences in instrument response and sensitivity under coarse-particle–dominated conditions. To further investigate this, we added a new comparison between the BAM and an independent FRM monitor (not collocated), which provides additional context for the observed discrepancies, particularly during non-dust (clean) periods when concentrations are low. We also performed an additional correction development using the LEAPS02 sensor collocated with BAM. The limited improvement achieved in this case highlights the challenges of correcting LCS data in dust-dominated environments when coarse-mode particle information (PM10) is not adequately represented. This finding is now discussed more explicitly in the revised manuscript.

In response to this and related comments, we have substantially revised Section 3.3.2 to improve transparency regarding the regression methodology, reference-monitor dependence, and the limitations associated with each comparison.

The following information was added to the revised manuscript to address this comment:

The second reference unit is the BAM-1022 unit operated and hosted by TCEQ. $PM_{2.5}$ concentrations measured by LEAPS02 were compared to the BAM-1022 unit. LEAPS02 had 5800 overlapping hours of measurements with the BAM-1022 unit (July 2, 2024, to February 28, 2025). For the measured period, $PM_{2.5}$ concentrations from the BAM-1022 unit ranged from -8.2 to 79.3 $\mu g\ m^{-3}$. 12.5% of the values of the BAM-1022 were < 0 $\mu g\ m^{-3}$. LEAPS02 $PM_{2.5}$ concentrations ranged from 0 to 51.6 $\mu g\ m^{-3}$ for the raw values and from 1.8 to 106.3 $\mu g\ m^{-3}$ for the calibrated values. During these overlapping times, the average hourly $PM_{2.5}$ concentration by the BAM-1022 was 4.4 ± 4.6 $\mu g\ m^{-3}$, while the average $PM_{2.5}$ concentrations from LEAPS02 reported were 6.95 ± 7.3 and 8.2 ± 4.96 $\mu g\ m^{-3}$ for the raw and calibrated values, respectively. A time series plot (Fig. 8A) as well as the box-and-whisker plot (Fig. 8B) comparing the BAM-1022 unit to LEAPS02 values shows that BAM-1022 measured much lower concentrations overall compared to the raw and calibrated LEAPS02. A full comparison between the BAM-1022 and LEAPS02 was made (Fig. S8), with and without July and August. This comparison did not yield a good agreement, as observed for AEROS. The R^2 value between the BAM-1022 and LEAPS02 raw measurements was 0.22, while the corrected LEAPS02 measurements produced a slight increase in R^2 values, but still low (R^2 of 0.34). RMSE and MAE with the corrected values, although improved from the raw values, yet were still high (4.0 and 2.8 $\mu g\ m^{-3}$, respectively). Even when the measurements from July and August were removed, the correlations between the units remained low (R^2 was 0.42, RMSE was 3.9 $\mu g\ m^{-3}$, MAE was 2.7, while the slope was 0.81). Overall, neither the raw nor calibrated data from LEAPS02 (with or without July and August) provided a good correlation with the data from the TCEQ BAM-1022. Even examining the data as daily values did not significantly improve the comparison between the corrected LEAPS02 to the BAM-1022 (R^2 was 0.49 for July-February and 0.72 for September-February).

All three events examined in section 3.3.1 for AEROS were examined here; these include: October 15, 2024, and January 17, 2025, dust events, as well as the 10 clean days for December 9 - 19, 2024 (Fig.

8C). These events show mixed results, with some showing good agreement between the corrected LEAPS02 and the BAM-1022, while others show no agreement at all. For the first dust event of October 15, 2024, at the peak of the dust, the BAM-1022 detected hourly $PM_{2.5}$ concentrations of 32.8 µg m^{-3} (at 18:00 LT) while the corrected LEAPS02 detected 57.9 µg m^{-3} (at 19:00 LT). This disagreement with the time of peak seems like an issue only for this specific event. Both the corrected LEAPS02 and the BAM-1022 detected the dust peak of January 17, 2025, at the same time, but the difference in concentration was significant. While the BAM-1022 reported hourly $PM_{2.5}$ concentrations of 62.2 µg m^{-3}, the corrected LEAPS02 detected concentrations of 106.3 µg m^{-3}. The EDM-180 measured a $PM_{2.5}$ concentration of 104.3 ± 19 µg m^{-3} at the same time. When the correlations between these hours were observed, it was found that for the October dust event, the BAM-1022 had a low correlation value (R^2 of 0.07) to LEAPS02, but for the 72 hours examined during the January dust event, this correlation value was high (R^2 was 0.93). A close look at the difference between the $PM_{2.5}$ concentrations showed that the corrected LEAPS02 values were higher (96% and 99% of the time for October and January examples). Even for the clean days of December 9 to 19, 2024, the corrected LEAPS02 unit measured a higher hourly $PM_{2.5}$ concentration (95% of the 264 hours) than the BAM-1022. A comparison of the hourly $PM_{2.5}$ concentrations measured during these clean days was low ($R^2 \leq 0.3$). It is possible to assume that the low agreement between BAM-1022 and LEAPS02 was because LEAPS02 was corrected based on the EDM-180, which is 8.2 km away. Perhaps the two sites (AEROS and TCEQ) experience different air quality levels, which would impact the comparison of BAM-1022 and LEAPS02. A comparison between the two reference units will help understand that aspect.

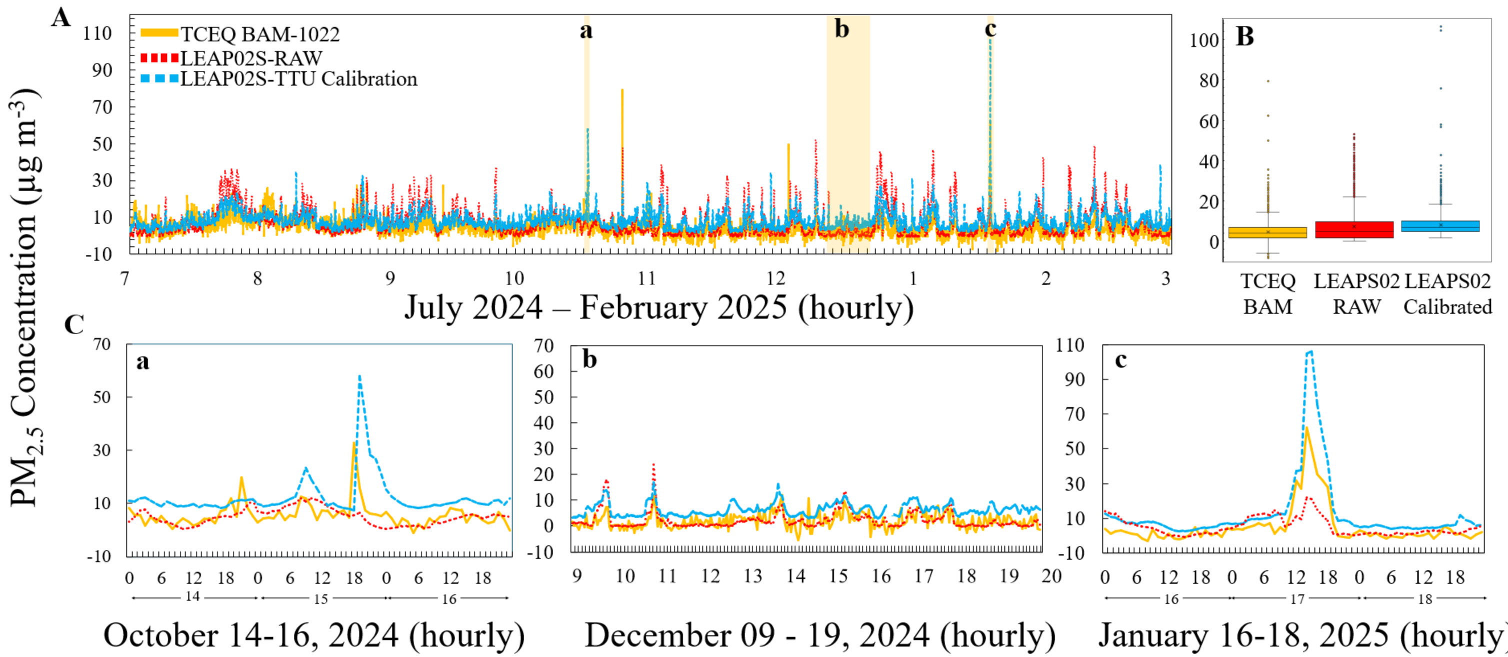

Figure 8: Comparison of $PM_{2.5}$ concentrations from LEAPS02, raw (blue) and calibrated (green), with reference units BAM-1022 at TCEQ site (in orange) for time series plots from July 1, 2024, to Feb 28, 2025. Box-and-whisker plots comparing the LEAPS (raw and calibrated values) with collocated reference units (B). (C) Example of three times of the time series plots of three events (a) October 15, 2024, dust event, (b) 10 days of clean days from December 9 to 19, 2024, and (c) a dust event on January 17, 2025.

A comparison between the two reference units (BAM-1022 and the EDM-180) was made for both hourly and daily values that were measured throughout the studied period from May 2024 to February 2025 (Fig. 9). Observation of hourly values indicates that overall, both sites experience similar air quality conditions, except for two very short and localized events, when BAM-1022 (on October 23 at 21:00 LT, and December 01 at 16:00 LT) had an increase of hourly $PM_{2.5}$ values, but these increases were only for one hour. Observation of daily values shows a much better agreement between the two units, as both have similar fluctuations of $PM_{2.5}$ values. However, for both hourly and daily values, it is clear from the box-and-whisker plot that the BAM-1022 measures lower concentrations than the EDM-180 measures. Three times lower compared to the EDM-180 (for both hourly and daily values). These overall low concentrations of the BAM-1022 could explain the low comparison made between the two reference units. Comparison based on hourly values (6580 hours) had R^2 of 0.44, RMSE and MAE of 4.0 and 2.7 µg m^{-3}, respectively, and a slope of 0.7. Daily values (based on 275 days) demonstrated a slight improvement in the comparison, yielding an R^2 of 0.71, RMSE, and MAE of 2.0 and 1.5 µg m^{-3} (respectively), and a slope of 0.92. Even a comparison between the HI and the BAM-1022 daily values for July 3 - 14, 2024 (shown in Fig. S9), highlights that the BAM-1022 reported overall lower $PM_{2.5}$ concentrations. These lower $PM_{2.5}$ values from the BAM-1022 have been reported in the literature when compared to FRM (Khan et al., 2024). Long et al. (2023) reported that the BAM-1022 underestimates the $PM_{2.5}$ concentrations compared to the FRM unit by ~15%.

Perhaps to find a better agreement between the LEAPS02 to the BAM-1022 unit, the LEAPS02 should be calibrated based on the BAM-1022 and not based on the EDM-180, which is ~8 km away. The same correction period used to correct the LEAPS unit in AEROS was used here (May 24 to June 30, 2024). Three different correction attempts were made to correct LEAP02 based on the BAM-1022. The first was based only on $PM_{2.5}$ values from BAM-1022, the second added the RH measured from the LEAP02, and the third also included the T and RH from the LEAPS02. The coefficients found for each correction, as well as the results of the comparison based on this correction presented in Table S4. Comparison between the corrections of the LEAPS02 $PM_{2.5}$ values just based on the BAM-1022 improved the comparisons, as lower RMSE and MAE values were found, although similar R^2 values were found. Improvement of the correlation between the LEAP02 and BAM-1022 was found when the RH and T with RH were added to the correction. R^2 values improved to 0.72, and both RMSE and MAE values decreased. Next, each of these corrections was implemented on the LEAPS02 data from July 2, 2024, to the end of February 2025, and a comparison between the BAM-1022 and the corrected LEAPS02 unit was made. Very low correlation values ($R^2 \leq 0.3$) and high MAE and RMSE values were found when the LEAPS02 was corrected based on BAM-1022, regardless of the correction type, and even with or without July and August (as shown in Table S5).

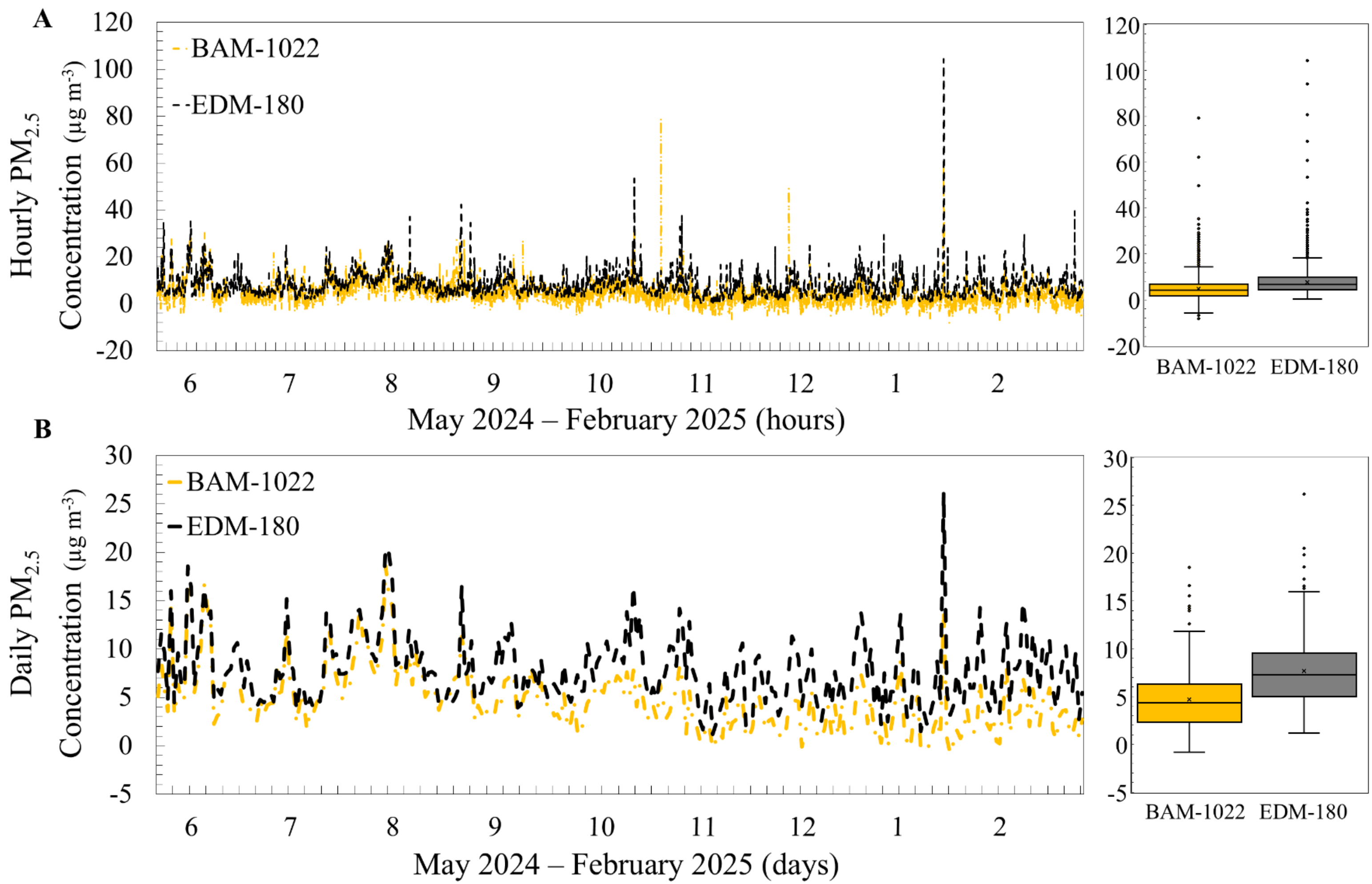

Figure 9: Comparison of $PM_{2.5}$ concentrations from BAM-1022 at TCEQ site (in light orange) and EDM-180 (black) with Box-and-whisker plots comparing the two reference units (A) for hourly and (B) daily values measured for May 24, 2024, to Feb 28, 2025.

To examine the corrections based on the BAM-1022 future, observation of the most severe pollution event during the study period was made. This severe pollution was the dust event of January 17, 2025. Hourly $PM_{2.5}$ values of LEAPS02 based on the different corrections were made to those measured by BAM-1022 as well as the EDM-180 (Fig. S10). It is clear to see that none of the corrections for LEAP02 based on the BAM-1022 units were able to capture the dust event. During the dust (11:00 – 19:00 LT), the BAM-1022 measured on average 2.4 times higher hourly $PM_{2.5}$ values than the LEAPS02 corrected (based on BAM-1022) values. Observations of the peak of the dust (15:00 LT) highlight the low quality of these corrections, as all three corrections based on BAM-1022 detected 2.5 times lower hourly $PM_{2.5}$ values than the BAM-1022 measurements, and 5.7 times lower than those detected by the EDM-180. None of these models developed based on the BAM-1022 were able to detect the dust, indicating that without proper correction, the LEAPS would not have been able to detect the dust particles, which are the main source of pollution in this area (Robinson and Ardon-Dryer, 2024; Ardon-Dryer, 2025). It can be concluded that the correction of LEAPS02 based on the BAM-1022 did not produce sufficient or accurate $PM_{2.5}$ values and should not be used. Perhaps using other corrections developed in the literature for Clarity sensors will result in better comparisons between LEAPS02 and the BAM-1022.

Three different corrections for Clarity Node S units were found in the literature that used FEM hourly $PM_{2.5}$ values, with T and RH measured from the Clarity Node S units (Liu et al., 2022; Raheja et al., 2023; Nobell et al., 2023). Both LEAPS02 and LEAPS01 were corrected based on these corrections for the training period of May 24 to June 30, 2024. Each LEAPS unit was corrected based on the reference sensor it was collocated with. The results of these corrections and comparisons can be found in Table S6. Between BAM-1022 and LEAPS02, there were 788 hours of comparison. Liu et al. (2022) was the only one that produced a reasonable R^2 value of 0.5. The other two corrections had $R^2 \leq 0.3$ and much higher RMSE and MAE values. Between the EDM-180 and LEAPS01, none of these corrections were able to produce a good comparison, and much lower R^2 values were found ($R^2 \leq 0.31$). Even with these low corrections, these coefficients (in Table S6) were used to correct the entire data set (July 2024 to February 2025). Hourly values from each LEAPS unit were then compared to the reference sensor it was collocated with. Regardless of the period examined (with or without July-August), R^2 values for both cases were $\leq$ 0.28. Highlighting the issue of using corrections made in a different location that had different meteorological and pollution types. Nilson et al. (2022) also stated that correction models do not perform the same at different locations and should be examined and/or developed per location.

Since the correcting of LEAP02 based on BAM-1022, and correction from the literature, did not produce any good corrections or agreement between the LEAP02 and BAM-1022, it was decided to examine the LEAP02 from a different perspective. Since the comparison between the BAM-1022 and EDM-180 shows that the areas experience similar air quality conditions (except for very few cases, shown in Fig. 9). A comparison was made between the EDM-180 in AEROS to the corrected LEAPS02 (based on TTU-calibration) in the TCEQ site from September to February (since July and August were problematic). High comparison between the EDM-180 to the corrected LEAPS02 was found, with R^2 of 0.89, RMSE and MAE of 1.7 and 1.1 $\mu g\ m^{-3}$, and a slope of 0.9 (based on 4332 hours of comparison). Even with July and August, high R^2 values were found (R^2 of 0.85; based on 5792 hours of comparison). Observations comparison based on daily $PM_{2.5}$ values between the EDM-180 and the corrected LEAPS02 (based on TTU-calibration) from September to February had even better correlation values (R^2 of 0.93, RMSE and MAE of 0.9 and 0.7 $\mu g\ m^{-3}$;

181 days). Even the three corrected LEAPS units in AEROS (LEAPS01, LEAPS41, and LEAPS42) had a good agreement with the corrected LEAPS02 based on hourly $PM_{2.5}$ values. LEAPS01 and LEAPS41 had a better agreement (R^2 of 0.94, RMSE of 1.3, MAE of 0.68) with LEAPS02 than LEAPS42 had with LEAPS02 (R^2 of 0.79, RMSE of 2.3, MAE of 1.16). If a good agreement is found between the LEAPS units and the EDM-180 across the two locations, perhaps the issue is with the BAM-1022 unit. It should be noted that we did not have any control over the BAM-1022 unit, as it was operated and calibrated by the TCEQ.

It was speculated that the numerous negative values produced by the BAM-1022 unit, coupled with the calibration for LEAPS02 being developed based on the EDM-180, led to the low comparison between the BAM-1022 and the corrected (TTU-calibration) LEAPS02. Perhaps the lowered regression values between the BAM-1022 and the corrected LEAPS02 resulted from the subzero $PM_{2.5}$ measurements, as BAM-1022 can measure concentrations down to -15 µg m^{-3} (Met One, 2025). During the examined period, May 2024 to February 2025, there were 765 hours when the BAM-1022 unit reported $PM_{2.5}$ concentrations < 0 µg m^{-3} (11.6% of measured time). Additional studies reported negative $PM_{2.5}$ concentration values from the BAM units (Khreis et al., 2022; Jiang et al., 2023). According to Jiang et al. (2023), there is very limited documentation on handling negative $PM_{2.5}$ data in the literature. Ambient air always contains certain amounts of particles, and negative $PM_{2.5}$ concentrations should never occur, yet some instruments, like BAM-1022, can record negative $PM_{2.5}$ values.

Some studies converted the negative values to 0 µg m^{-3}, while others used a lower limit of detection threshold for the $PM_{2.5}$ concentrations (Magi et al., 2020; Khreis et al., 2022). Multiple attempts were made, including removal of all the reported negative values, converting the negative values to 0 µg m^{-3} as suggested by Khreis et al. (2022), as well as using a limit of detection threshold (2.4 µg m^{-3}; based on Magi et al., 2019). Yet none of these attempts improved the regression between LEAPS02 (TTU-calibration) to the TCEQ BAM-1022 unit; R^2 values remained below 0.4. To examine the BAM-1022 negative values in depth, all the minimum daily values reported by TCEQ since the site became operational (on August 13, 2016) were observed. From August 13, 2016, to July 11, 2018, the site hosted a TEOM unit, and on July 11, 2018, the unit was replaced with the BAM-1022. None of the 667 days operated by the TEOM had negative hourly $PM_{2.5}$ concentrations. Out of the 2278 days examined since the BAM-1022 became operational (July 11, 2018, to December 31, 2024), more than half (53.2%) had negative $PM_{2.5}$ concentrations daily minimum. It is known that some negative readings are caused by instrument faults or procedural errors, meaning they can be invalid and excluded from air quality reporting towards the public domain (Jiang et al., 2023). But in the case of the unit in Lubbock, they are reported. Perhaps since the air quality system database for the USA (USEPA, 2014) treats negative data from $PM_{2.5}$ continuous monitors as valid, and only values below a threshold of -10 µg m^{-3} are removed. Yet, the USEPA. (2016) indicated that it is generally agreed that negative data should be excluded from public reporting.

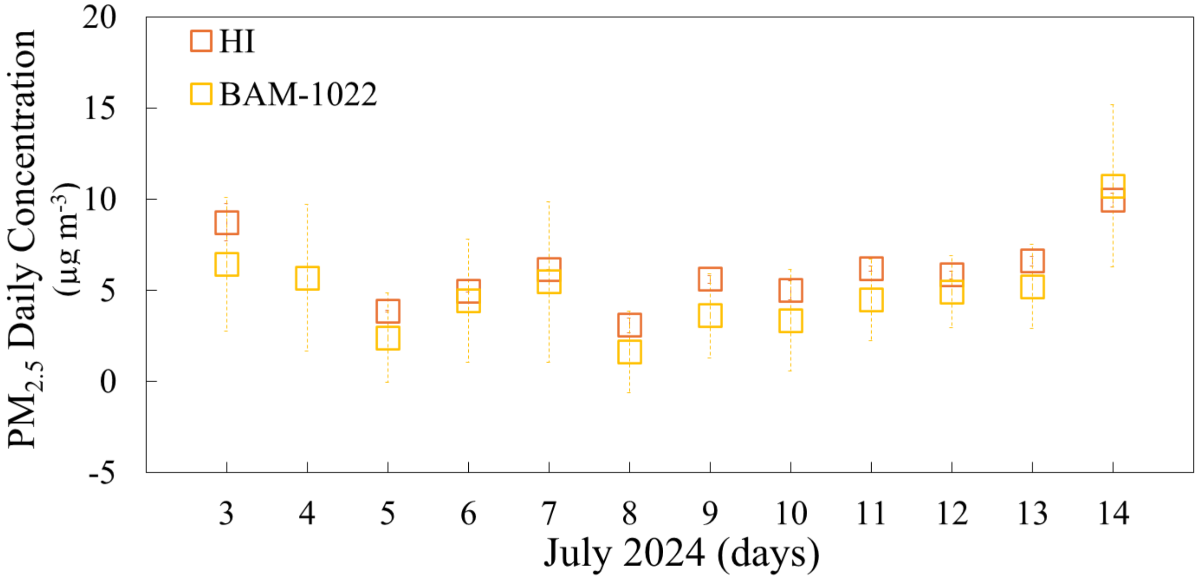

Figure S9: Comparison of daily $PM_{2.5}$ concentrations between Harvard impactor (dark orange square) and BAM-1022 (light orange square) for July 2024.