We would like to thank anonymous reviewers for the helpful comments and suggestions. In line with the reviewer comments and suggestions, and in line with new publications that were published while this paper was under review, we modify and revised the manuscript. Below are all the comments (in bold) followed by the replies. The parts that are in italic are corrections that are included in the revised version of the paper.

Sincerely,
Karin Ardon-Dryer

Response to Reviewer 1

The work by Ardon-Dryer et al. present a large database of PM2.5 mass concentrations collected by a low-cost sensor, the Purple Air PA-II unit across eight locations in the USA. At each location, there were measurements from multiple PA-II units, and the closest air quality monitoring station (AQMS) data was also acquired. The authors have performed a large amount of analysis on this dataset comparing the PA-II to reference instrumentation. However, it was not clear to me what the scientific novelty of the paper was, as there have been a number of papers already that evaluated the Purple Air sensor, as mentioned in the introduction. The authors state that the aim was to ‘examine how PA-II units perform under atmospheric conditions when exposed to a variety of pollutants and PM2.5 concentrations’, yet this is a rather vague aim, that this dataset may not be suitable to answer. This is a great dataset that could be used to investigate a number of interesting questions regarding low-cost sensors and their calibration and suitability for large scale deployment. I feel that this paper suffers from a lack of focus and could be improved if the authors articulated and addressed more novel, detailed and specific aims and objectives. This leads to another area that could be improved, as most of the analysis is rather descriptive and lacking in depth. In my opinion, it is not enough to just present the regression analysis for all the PA-II units (i.e. r2, slopes etc) to the AQMS instruments. For example, there could have been more analysis on why there was a large range in observed r2 between all the unit and the AQMS? Was there any common factors for units that had a poor or good correlation with the AQMS? Did the actual reference instrument at the AQMS site affect the correlation (e.g. between FRM and FEM instruments)? I would have also like to have seen more focus on the observed slope
between PA-II and the AQMS, as this is a better indicator of the accuracy of the PA-II than the correlation co-efficient (r2).

We appreciate the reviewer’s comment, based on the review comment we clarify our aims in the manuscripts.

The following information was added to the manuscript

This study aims to examine how each PA-II unit performs under atmospheric conditions when exposed to a variety of pollutants and PM$_{2.5}$ concentrations (PM with an aerodynamic diameter smaller than 2.5 μm), when at a distance from the reference sensor. We examine how PA-II units perform in comparison to other PA-II units and Environmental Protection Agency (EPA) Air Quality Monitoring Stations (AQMSs) that are not co-located with them.

This study aims to examine how PA-II units perform under atmospheric conditions when exposed to a variety of pollutants and PM$_{2.5}$ concentrations. For the scope of this study, we chose to focus only on regions that contain at least one pair of co-located PA-II and AQMS units. Corrections of PM$_{2.5}$ values for co-located PA-II and AQMS units, based on MLR, were performed and applied to all the other PA-II units in that region. Comparison of PM$_{2.5}$ measurements taken by all units in each region, AQMSs and PA-II units (when PM$_{2.5}$ values were measured or corrected) are presented. The presented comparisons were done for both the entire study period and for specific events that we wanted to examine in greater detail.

Regarding the comments about factors for units that had a poor or good correlation with the AQMS, In our original manuscript most of the PA-II units that had a low correlation with the AQMS units also suffer from low correlation with the other PA-II units, we decided to remove these units as we believe they are outliers. In the current dataset, there are only two units that had a lower correlation with the AQMSs, but these two units were borderline for our PA-II outlier test. Without these two units, most of the R$^2$ values will be >0.6. The evaluation test between the PA-II units will help identify PA-II units that are not performing well. A reduction in performance can occur over time or due to exposure to events with high PM, as described in Sayahi et al. (2019). This information was added to the manuscript.
The following information was added to the manuscript

*Overall, almost all the PA-II units had high correlation values when compared with the other PA-IIs or AQMSs in their region. Two PA-II units, SL-PA-6 and SL-PA-8 had low R² values with the AQMS, they also had a relatively low correlation with the other PA-II units. It is feasible, that if stricter rules for identifying outlier PA-II units were in use, these two units would have been considered as such and subsequently removed from the data set.*

As for the comment **Did the actual reference instrument at the AQMS site affect the correlation (e.g. between FRM and FEM instruments).**

All the AQMSs that were used in this work were of FEM type, their selection was based on the distance that was used in previous works (e.g. Bi et al., 2020). Therefore, we could not evaluate that difference (FEM vs FRM).

Regarding the reviewer comments **that like to have seen more focus on the observed slope between PA-II and the AQMS, as this is a better indicator of the accuracy of the PA-II than the correlation co-efficient (r²).** Information on the slope was added to the manuscripts for all comparisons as shown in Table 1, Table S3, and Fig. 7.

**One of the key issues with this dataset, as acknowledged by the authors in section 3.3.3, was that the PA-II units were not co-located with each other or the AQMS and could therefore diminish the ability to compare the PA-II to reference instruments. Unfortunately, in my opinion the authors did not adequately address this issue. It would have been interesting if a more in-depth analysis of how the PA-II relationship with reference instrument varied as function of distance, as this would be of great interest to the community.**

The goal of this paper was to observe PM₂.₅ measurements using the PA-II units, these units installed by citizens are for the most found in residential locations across the United States, therefore only a handful are co-located with an AQMS, and in fact many of the regions in which PA-IIs are deployed do not have even a single reference unit. Previous works have examined the efficiency of the PA-II unit by comparing it to a co-located AQMS or in laboratory conditions. For
the purpose of this study we defined a co-located pair as a PA-II that is up to 1.1 km away from an AQMS, the selection of this distance is based on the work of Bi et al. (2020). A major addition to this revision is the implementation of a data correction process that was applied to the PA-II measurements. This correction process was well documented by Bi et al. (2020) and Magi et al. (2020) for both PA-II units and by Malings et al. (2020) for other low cost sensors.

As for the impact of the distance of the units, we did not find that the distance between the units impacted the behavior and comparison of the unit, yet we only evaluate a distance of up to 5 km from an AQMS and up to 10 km between PA-IIs, units which will be far away may have a different impact, but evaluating that would be beyond the scope of this work.

The paper is well written and clearly presented but the large volume of data presented did make it difficult to follow at times. For example, the tables are too big, and could do with either being separated by city, or only the pertinent information being included.

We modified the provided tables and information in the text. Each table now represents a single region and does not include more than two parameters.

In addition to the above, number of more detailed comments are given below Abstract: When you state that the units had good agreement it is important to back this up with numbers, such as giving the slopes, r^2 etc. This generally true throughout the paper.

Information on the comparisons between units was added to the manuscript per region. Values of the R-squared (R^2), root mean square error (RMSE), mean absolute error (MAE) as well as the best fit information, including the slope, are provided in the revised manuscript.

Based on the reviewer comments the following information was added to the manuscript

*In most cases, the AQMSs and the PA-II units were found to be in good agreement (75% of the comparisons had a R^2 > 0.8)*
Page 2, line 63: In addition, the authors could reference Crilley et al 2018 and Di Antonio et al 2018 for possible solutions to the RH effect on low-cost PM sensors.

Both references were added to the manuscript

*Previous studies suggested that part of the problem with the PA-II unit results from the optical particle counter being impacted by changes of RH (Crilley et al., 2018; Malings et al., 2020; Magi et al., 2020). .... deliquescent or hygroscopic growth of particles, mainly under high RH conditions, can lead to higher reported PM concentrations (Di Antonio, 2018; Jayaratne et al., 2018; Bi et al., 2020), which ends as an overestimate of the PM compared to the reference units.*

Page 5, line 152: this paragraph could instead be presented as a table. Furthermore, it may also help the reader if you were to give the AQMS and PA-11 units more accessible names. For example, the Pittsburgh AQMS could be P-AQ-1 and 2, and the PA-II units, P-PA-1, 2, 3 etc

Per the reviewer's comments, the entire paragraph was removed from the manuscript. Also, all unit’s names are now represented by location and instrument code as well as running ID number, as specified in Table S1.

*For simplifications, each region was defined by two letters to represent its name (DE for Denver, SF for San Francisco, VA for Vallejo, and SL for Salt Lake City). Also, each unit type received a two letter code (AQ for AQMS and PA for PA-II). Each unit received a number instead of an ID, as shown in Table S1.*

Section 2.4: more info is needed on the data analysis, what sort of regression analysis did you do? In what computer program? Which AQMS station did you use, the closest or each one for a given city?

In the manuscript, we describe that we used Multivariate linear regression (MLR) models between the PM$_{2.5}$ values of the co-located PA-II and AQMS with T and RH. In addition, all the analyses were performed using Matlab and Excel. This information was added to the manuscript
To evaluate the similarities and differences between the PA-II units and the AQMSs and other PA-II units, a set of calculations and comparisons was performed using Matlab and Excel.

Regarding the AQMS database, we downloaded the entire data set of the hourly PM$_{2.5}$ for all AQMS units that were active during the study period. Using a distance calculation, we were able to identify regions with multiple PA-II units as well as at least one AQMS. We added this information to the manuscript.

Hourly measurements of PM$_{2.5}$ (FRM/FEM Mass code - 88101 file) from all AQMSs collected by the EPA from January 1, 2017, to December 31, 2018, were selected from the EPA website (https://aqs.epa.gov/api).

By using the JSON file for the PA-II units and the 88101 file for the AQMS, we calculated the distances between all the units to identify regions with multiple PA-II units (a minimum of five units) and at least one AQMS. At least one AQMS unit needed to be at a distance of 1.1 km from at least one PA-II unit (defined as a co-located pair, a similar range used by Bi et al., 2020). All the units in these regions needed to be active during the designated time period of January 1, 2017, to December 31, 2018. In each region PA-II units needed to be less than 5 km from at least one AQMS unit and up to 10 km from each other.

Four different regions containing a total of seven different AQMSs (all FEM type) and 46 different PA-II units were identified:

Section 2.6: I do not see the point of calculating the AQI when the point of this article is to compare the measurements between the PA-II and reference instruments. If they report the same concentration, wont they give the same AQI? I think you should just focus on reported concentrations.

AQI information was removed from the manuscript per the reviewer's comment.
Section 3.1.1: If Fig 2 is on page 29, then this is not a distribution but a time series of reported concentrations. A distribution to me implies a histogram, please correct the naming. Also why did the AQMS report higher PM$_{2.5}$ concentrations at Berkley, Ogden, Linden and Salt Lake City compared to all the PA-II units during the first half of 2018? Understanding why the relationship changed is important for knowing the parameters that affect the PA-II measurements.

This plot is now presented as a time series per the reviewer’s comment.

*Time series of daily PM$_{2.5}$ values for each unit at each of the four regions are presented in Fig. 3.*

Some of the regions mentioned in the reviewer comment have been removed from the manuscript as they were missing a co-located AQMS. For the remaining regions, the higher AQMS measurements are attributed to what seems to be a connection with days that have low RH values resulting in lower PM$_{2.5}$ values being measured by the PA-II units. We also believe that chemical analysis during for these times could help understand the difference between the AQMS and PA-II, unfortunately, such analysis will be beyond the scope of this study.

*In some cases, the AQMS measured higher PM$_{2.5}$ daily values compared to the PA-II units, mainly at days with low PM$_{2.5}$ values, as seen in April - June 2018 in Vallejo (Fig. 3C) and Salt Lake City (Fig. 3D). These differences were observed mainly in days with low RH values (Fig. S3).*

![Figure S3: Time series of daily PM$_{2.5}$ measurements from the AQMS and PA-II units in Vallejo (A), and Salt Lake City (B) during April-May 2018. Measurements from AQMS are represented by the green lines and the PA-II units are indicated by purple lines. Relative Humidity values represented by the gray dotted line.](image-url)
Page 8, line 236: the authors state “These high correlation values and relatively low RMSE indicate that although the PA-II units and the AQMS are not co-located, they still tend to behave in a similar way.” Why do think this was the case?

This sentence was removed from the current manuscript.

Page 8, line 242: I do not understand what you mean by instrument efficiency?

This sentence was removed from the current manuscript.

Page 8, line 250: why did you subset the data below 40 ug m-3?

This analysis was removed from the current manuscript. Originally, we set 40 ug m-3 as the maximum point for the study as the work of Sayahi et al. (2019) suggests that above it the PA-II measurements are impacted by the high PM concentrations. Meaning at lower PM concentration we will find a better correlation between the PA-II and the AQMS.

Page 12, line 364. In the previous paragraph you state that RH is a more important parameter than temperature when considering potential artefacts for the PA-II, so why compare to temperature?

This sentence was removed from the current manuscript.

The original sentence was based on findings from several papers. We originally compared the temperature in order to prove our theory that temperature is not as important. However, during the time that our original manuscript was under review several new papers were published which in turn made us make extensive changes in our manuscript regarding the impact of RH and T. We added a hunidogram and a plot that investigate the impact of T in each of the co-located units (AQMS with PA-II, Fig S4). Some of the PA-II units might be impacted by both T and RH, this information was added to the manuscript.
Calculations of the ratio between the measured PM$_{2.5}$ from the PA-II to the AQMS as a function of $T$ and $RH$, known as a hunidogram, were performed (Fig S4). Some of the PA-II units seem to be impacted by $T$ and RH more than others; these units also had relatively low $R^2$ values with the AQMS unit, as in the case of DE-PA-6 in Denver (Fig. S4A).

**Figure S4:** Ratio between measured PM$_{2.5}$ from PA-II to the AQMS, as a function of temperature and relative humidity (hunidogram) for all collocated PA-II and AQMS pairs. Information on the distance and $R^2$ values between the two presented in each plot.

We also added a multivariate linear regression (MLR). The MLR takes into account changes of $T$ and RH. All the PA-II units’ measurements were corrected based on the MLR of the co-located PA-II and AQMS, this information was added to the manuscript:

*Based on the MLR, the multivariable linear dependence of PA-II PM$_{2.5}$ on AQMS, RH and $T$ created the predictors of PA-II as:*
\[ PA - II(PM_{2.5}) = A_1 + A_2AQMS(PM_{2.5}) + A_3T + A_4RH \]

(1)

where \( A_1, A_2, A_3, \) and \( A_4 \) fit coefficients received from the MLR, \( PA-II \) \( (PM_{2.5}) \) and \( AQMS(PM_{2.5}) \) are in units of \( \mu g \text{ m}^{-3} \), \( T \) is in Celsius, and \( RH \) is in percentage. Based on these parameters and fit coefficients, a calculation of the corrected \( PA-II \) \( PM_{2.5} \) hourly values for each \( PA-II \) was performed using the following:

\[ PA - II(PM_{2.5}), \text{corrected} = \frac{PA-III(PM_{2.5}), \text{uncorrected} - A_1 - A_3T - A_4RH}{A_2} \]

(2)

Details of the coefficients received in the MLR as well as the regression output including \( R^2 \), \( RMSE \), \( MAE \), and slope for each correction of \( PM_{2.5} \) values in the \( PA-II \) units, for each region, can be found Table 1. Figure 3 presents a comparison of the \( PM_{2.5} \) values from the uncorrected \( PA-II \) unit to the AQMS as well as the \( PA-II \) \( PM_{2.5} \) values hourly after correction, per region.

Page 12. Line 381: I do not agree with this statement as you have not able to test the precision of the \( PA-II \) as they were not co-located. The precision of the \( PA-II \) units would be tested by how well each \( PA \) unit agree with each other at a given \( RH \), but you have looked for correlation between \( RH \) and \( PA-II \) reported \( PM2.5 \). this does not indicate the precision of the \( PA-II \) only if there was a relationship between \( RH \) and reported \( PM2.5 \) concentrations.

This sentence was removed from the current manuscript.

Page 13, line 418: where the slopes between the \( PA-II \) and AQMS instruments affected by distance?

We did not find an impact of the distance on the slop., No impact was observed when the \( PA-II \) units were compared to the nearest AQMS in all regions, and no effect was found when comparing the \( PA-II \) to each other.

This information was added to the manuscript as text and figure:

*Because the AQMS and the PA-II units were not co-located, we wanted to verify whether the distance between all the units affected the \( R^2 \), RMSE, MAE and slope values. We compared the \( R^2 \).*
RMSE, MAE and slope values received from the comparisons of hourly PM$_{2.5}$ measurements with the corresponding distances between the units (Fig. 7). There was no correlation between the two. Not when the PA-II units were compared to the nearest AQMS units (Fig. 7A), or between the PA-II units (Fig. 7B), before or after the corrections of the PA-II PM$_{2.5}$ values. Therefore, the distance between the units did not impact the comparison.

![Figure 7](image.png)

**Figure 7**: Comparison of distance (km) between PA-II to its nearest AQMS in all regions (A) and between each PA-II unit to all other PA-II units per region (B) to $R^2$. RMSE, MAE and slope values received from the PM$_{2.5}$ hourly measurements comparison.

Section 3.4: I think that section could be improved by including some recommendations based on your findings from this study.

The original section was removed from the manuscript, instead we did implement it as part of section 3.5 “Underlying Differences and Future Implications”. As suggested by the reviewer we added several recommendations, this includes but not limited to how to use the measurements of the PA-II units, the necessity of Temperature and Relative Humidity measurements, steps for assuring the unit integrity and more.

This information was added to the manuscript:

**3.5. Underlying Differences and Future Implications**

While appropriate PA-II PM$_{2.5}$ value corrections can improve the comparison between the PA-IIIs with reference units, there are other differences between PA-IIIs and AQMS units that can influence
the comparison results, including the underlying technology and the manner in which units are placed. The PM$_{2.5}$ sensors in the AQMSs perform gravimetric measurements using the mass of the particle; by contrast, the PA-II units use a laser particle counter to count electric pulses generated as particles crossing through a laser beam. The method used by the PA-II might impact the count of particles during high humidity conditions or when a majority of the particles are volatile. Another difference is the physical location of the units; whereas AQMSs are meticulously positioned in an open area, the location of a PA-II is determined by its owner. Although PurpleAir recommends positioning the PA-II in an open area, ultimately, it is the owner’s decision. In practice, most of the PA-II units are located in residential areas with low-rise housing. Furthermore, the height at which the sensor is located could affect the measurements. The height of the AQMS inlet is regulated and kept constant at each location; on the other hand, the owner of a PA-II unit can freely place it near the ground or higher up. The location of the PA-II units in residential areas can provide both an advantage and a disadvantage. For example, a single PA-II unit might be exposed to more localized PM sources such as a barbecue, lawn mower, or car, making it report different results compared with other units in its area. Therefore, an increase of PM by a single PA-II unit should be taken into account. When the PA-II is used as a network, as suggested by Ford et al. (2019), comparison of the PM values measured by all PA-II units will help identify such a localized source. Maintenance and calibration are other possible causes of differences between the two. The PM$_{2.5}$ sensors in the AQMSs have strict rules for the monthly evaluation of sensor performance, including through flow calibration or calibration based on minimum value threshold (which, in some cases, causes the recording of negative PM values). By contrast, PA-II units do not have any quality control other than that done by the company for each sensor before shipment to the customer (PurpleAir personal communication, 2019). Another point that should be taken into account is the lifetime of the PA-II units. The manufacturer of the PMS5003 sensor used in the PA-II units states that it has a lifetime expectancy of ~3 years (Yong, 2016). Bi et al. (2020) found that the PA-II unit’s efficiency is affected even after only two years of being operational.

Based on the findings from this work, we believe that there are several needed steps that will allow the usage of the PA-II units in air quality and health related research. First, users should identify regions with multiple PA-II units, where at least one PA-II is co-located with an FRM or FEM
Ideally the same location will also contain measurements of T and RH, or at least T and RH measurements will be nearby. Keep in mind that it is not recommended to use the PA-II internal sensors for T and RH values, as they are not representing the atmospheric measurements (Malinges et al., 2020; PurpleAir personal communication, 2019). However, we have found that in many regions there is no meteorological station that can serve as a reference for the correction process. It would be useful then, to devise a way in which the PA-II internal T and RH sensors can be used. To achieve this, an extensive study is necessary, to gain a better understanding of the issues related to the usage of the PA-II internal sensors and to formulate a calibration equation that then can be applied to the desired PA-II units.

Comparison of all PA-II units in each region will help to identify and remove outlier PA-II units from future analysis. Exposure to high PM concentration might affect the PA-II efficiency, as suggested by Sayahi et al. (2019), and therefore, its measurements will differ substantially from those of the AQMSs and other PA-II units. Ideally PurpleAir should monitor all active PA-II units and identify units that behave differently from surrounding PA-II units or identify PA-II units whose internal sensors (A and B) report different values, flag them on the online map, and communicate instructions to the unit owners on how to fix or replace the unit.

After PA-II units have been identified, users should conduct MLR between the co-located PA-II and AQMS units, including measurements of T and RH. For the MLR to be efficient it is important have a wide range of PM$_{2.5}$, T and RH measurements. This MLR will provide a coefficient that will be used to correct all the remaining PM$_{2.5}$ values of all PA-II units in that region. Evaluation of the PA-II PM$_{2.5}$ value corrections should be made for the duration of the study but also for specific events with spatial impact such as inversion, dust storms, biomass burning, and more. Such events should impact a larger area and therefore will allow detection of the PM changes in all PA-II units as a whole (network). Correction of PA-II PM$_{2.5}$ values should be performed per region, as they represent specific PM values as well as changes of T and RH values that the PA-II units were exposed to. This will help the public obtain information on the spatial and temporal distribution of PM concentrations in their area (Gupta et al., 2018; Morawska et al., 2018), which will enable them to monitor local air-quality conditions (Williams et al., 2018) and help make decisions related to events with high PM exposure.
In this study, we evaluated PA-II units that were up to 5 km away from an AQMS unit, as well as up to 10 km from each other. This raises the question of maximum effective distance. What is the maximum distance between an AQMS and PA-II units that will still allow for the MLR to successfully correct the measurement taken by PA-II units; a distance greater than this would carry the potential of introducing additional factors that might impact the comparisons. Another situation that requires further investigation is that of regions that include multiple PA-II units but do not have a co-located pair or completely lack a reference monitoring station. The question in mind, if and how we might use neighboring regions in which measurements were successfully corrected to compensate in the case of such problematic areas. For example, could we have used Vallejo and San Francisco, two regions that were included in this study to correct the measurements of the PA-II units in the region of Berkeley - Oakland that resides between the two?

Page 14, line 433: please call it instrument drift, as instrument efficiency is meaningless in this context.

This sentence was removed from the manuscript.

References


PurpleAir, PurpleAir Map, air quality Map: http://map.purpleair.org/, last access: 1 August 2019.

