Response to Anonymous Referee #1

The manuscript is well-written and clearly structured. Furthermore, the spectral analysis method used to evaluate the data is novel in the context of low-cost sensors. To criticize, the study does not yield scientifically significant new findings. The main conclusion is that the sensor response is source dependent and that without proper calibration, there is a high risk of data misinterpretation. This is the same conclusion that has been made in most, if not all, studies investigating low-cost sensors.

I recommend publication of this study because I consider the approach used to evaluate sensor data valuable. Furthermore, I encourage the authors to consider the following points to strengthen the impact of the research.

We sincerely appreciate the reviewer for dedicating their time to provide detailed feedback. Below, we have addressed each of the specific comments in blue text.

1. The manuscript lacks a clear statement of the limitations of the study. This would be useful for readers to contextualize the findings.

Response: We agree with this suggestion and have included the following paragraph in the manuscript to address the study's limitations.

This study has a few limitations. Firstly, the study is limited to one city, and the low-cost air quality sensor network used in the study is not perfectly co-located with the EPA monitoring sites. This can introduce uncertainties in the analysis due to differences in local air properties and pollution sources for the two data sets. Secondly, the placement of the low-cost sensors relative to local built-structures could affect its measurement performance and increase data uncertainty, but this information is not available to us. Thirdly, we did not have access to local traffic-related information or industrial activity, restricting our ability to strongly relate frequency components to specific emission sources. The likely variability of the local emission sources at the different Purple Air and EPA sites adds uncertainty in quantifying the differences in the short-term responses of the two networks.

2. The authors suggest that the results of their analysis will provide guidance in devising new approaches to calibrate data from low-cost sensors, but it is unclear what specific recommendations are being made . A more explicit discussion of the implications of the study's findings for future research and policy decisions would have strengthened the overall impact of the article.

Response: We have included the following paragraph in the conclusions of the manuscript to discuss the implications of the study's findings for future research and policy decisions.

This study clearly demonstrates that low-cost sensor PM data has non-uniform contribution of different PM sources. Any field calibration of these sensors using simple regression models cannot correct
for this non-uniform contribution. As best practice, it is recommended that calibration models from field data should report, at a minimum, the distribution of different PM emission sources at that location, and ideally also, the particle size distributions.
Given the periodic signatures of many sources, frequency-based scaling approach should be explored towards the development of more robust calibration models that account for the wide range of emission sources common in urban environments. Accuracy of such models will scale with time periods of calibration. Considering the source-dependent response of low-cost sensors, calibration models developed using land-use data might be an advance over simple regression models.

3. Line 89 foe;d typo?

Response: We thank reviewer for pointing out this typo, we have fixed it in the manuscript.

4. 1 Consider adding a scale for the map and units for the population density.

Response: We have added the scale and units of population density in figure.

5. Local correction model; justify the use of both temperature and relative humidity in multiple linear regression. These variables are correlated with each other which can be problematic as the independent variables in MLR should be independent .

Response: We have included the following justification of using relative humidity and temperature in the model to manuscript.

Typically in MLR models, we would only consider independent variables and it could be argued that temperature and relative humidity are not entirely independent. But from a particulate matter perspective, the differing impact of these parameters make them independent of each other. Relative humidity directly affects particle size and hence measurements by low-cost sensors, such as PA. Temperature, however, has a more complex connection to particle properties. Temperature directly affects particle size and composition by modulating condensation/evaporation, which can affect PM measurements by both EPA and low-cost sensors. Temperature also indirectly affects PM properties at a location through its relation to local meteorology, especially wind direction, and hence the distribution of sources at the measurement
location. To establish the independence of these parameters, we calculated the Variance Inflation Factors (VIFs) for temperature and relative humidity and these were found to be below 5. These small VIF values indicate a low level of multicollinearity for the two parameters(Ros-Gálvez, 2017) and permit their inclusion in the MLR model.

6. Line 294 “Our analysis clearly demonstrates for the first time that the PA network’s very different sensitivity to different sources.” I suggest you remove the “first time” part here.

Response: We have removed the “first time” from our manuscript.

Response to Anonymous Referee #2

This paper uses spectral analysis to assess the accuracy of low cost air quality sensor networks. This approach is sensible because as the authors note “Sources with short time periods, relative to the calibration period, are averaged out and inadequately accounted for in the calibration. Thus long time scale events are completely lost in the calibration process.” The new calibration approach is interesting and appears to have potential.

Response: We sincerely appreciate the reviewer for their valuable time in providing constructive feedback. Our responses to the specific comments are presented below in blue text.

The paper highlights that both the regulatory and low cost networks show distinct peaks in the power spectrum at 4, 8, 12 and 24 hours. It is stated that the 24 hour peak “likely represents harmonics of the 8 hour and 12 hour signals, and hence represents a combination of both (8 and 12 hours) sources”. Isn’t it more likely that the same sources occur at the same times daily, i.e. morning rush hour, evening rush hour, etc. all start at the same time each day and so you would expect 24 hours to turn up in the power spectrum very strongly?

Response: Yes, we expect the 24 hour peak to be stronger than the 8 and 12 hour peak as is observed in Figure 6. Small differences in the times of rush hour between weekend and weekdays and possibly also within weekdays, could slightly moderate the magnitude of the 24 hour peak.

The 8 hour peak being traffic seems very plausible and the analysis on the days of the week is good evidence due to the preponderance of 8 hour working days. The argument put forward for the 12 hour peak would seem to make more sense to me to be at 24 hour peak, i.e. peak sunlight follows a 24 hour period not a 12 hour period. Could the 12 hour period be representative of 12 hour work shifts, in addition to the 8 hour shifts? Whereas 8 hour shifts represents a standardised “9-5 working day through Mon-Fri”, a 12 hour shift is more likely for shift work where workers have a certain number of days on and a certain number of days off, which would not show so significantly in the day of the week effect? The peaks at 4 and 6 hours are also intriguing, what could be causing these?

Response: We agree that the 12 hour peak could also be traffic related. In addition, the 4 and 6 hour peaks also likely have a relation to traffic patterns, as shown in (Lu et at., 2014).

The lack of concrete understanding of the periods should be stated more clearly in the conclusions and abstract. The results are interesting, but far from being definitive as yet.

Response: Yes, we have added the sentence in abstract as well as a paragraph in the conclusions to state lack of concrete understanding of the periods.

With the rationale of calibration periods missing short term signals through the averaging out of signals, it is unclearly why outliers (L100) were removed. These signals could very possible be true, for example due to construction dust etc.

Response: The reviewer raises a good point. Some of the outliers could indeed be real data. Our rationale in removing the outliers was based on following the approach of previous studies, such as the work by (Barkjohn et al., 2021). We believe the major results of our study will not be affected by the treatment of the outliers.

It's not clear why a linear RH correction is used (L118) when the RH effect has been shown to be due particle hygroscopicity, which is non-linear with respect to RH. The k-Kohler approximation well in many low cost sensors, e.g. Crilley et al. 2018 - https://amt.copernicus.org/articles/11/709/2018/ The lack of diurnal RH effect in the correction might explain some of the difference seen in the power spectra of the EPA and low cost data.

Response: It is true that particle hygroscopicity is influenced by RH in a non-linear manner. However, in our study, we intentionally avoided making assumptions about particle properties and their behavior.

The linear RH correction approach was chosen to maintain simplicity and minimize assumptions. This approach has been used in other studies, including the work by (Barkjohn et al., 2021), which we followed for consistency and comparability. By using a linear correction, we aimed to provide a practical and straightforward method for correcting low-cost air quality sensor data without making complex assumptions about particle hygroscopicity.

The authors should provide some rationale why they think the low cost sensors are missing the high frequency outputs. For example, why would the low cost sensors be blind to traffic data for example? Are the low cost sensors not measuring the smallest particles that the EPA sites are measuring? If so, then the low cost measurements are blind to a subset of particles. What is the implication of this?

Response: We have added following paragraph in conclusions.

The reason why low-cost sensors may be missing high-frequency components from sources such as traffic can be attributed to several factors. One factor is the minimum detection size limit of the sensors, which is ~ 300nm. Sources, such as traffic, with PM emissions predominantly in the sub-300nm size range will, thus, be under-detected in low-cost sensors. EPA measurements do not have this limitation. Additionally, low-cost sensor response depends on the composition and shape of particles, resulting in PM measurement accuracy varying with emission sources.

The implication of these limitations is that the measurements provided by low-cost sensors, such as those in PurpleAir, will be underestimated with respect to certain pollutants, including those associated with traffic emissions, and overestimated related to others. Consequently, relying solely on low-cost sensor measurements without considering the limitations in particle detection and composition could result in an incomplete understanding of air quality, especially in relation to specific pollutant sources or components.

Minor comments

L23-24 the literature around links between air pollution and COVID-19 outcomes are still contested. I would insert a 'maybe' within "…exposure to PM2.5 [might] also impact responses to acute diseases such as COVID-19".

Response: We have fixed this in our manuscript.

L29 – define FRM and FEM and indicate how they differ.

Response: We have added this paragraph to the manuscript.

FRM refers to the specific monitoring methods that have been designated by the EPA as the reference standard for measuring air pollutants, while FEM refers to alternative monitoring methods that have been deemed equivalent to the FRM methods by the EPA. The two methods may utilize different instruments or measurement techniques but have demonstrated comparability in accuracy and reliability. The strict maintenance and calibration routines followed in these stations ensure high-quality data and comparability between different location.

L44 "Using two sensors… allows for the robustness of data collection". Spell out why two sensors improve robustness.

Response: We have added this updated sentences in the manuscript.

Using two sensors that measure the exact same PM measurements allows for the robustness of data collection by minimizing any data noise, loss of data due to sensor failure, or measurement error due to sensor electronics issues.

L80 provide some rationale why Chicago PM2.5 have near doubled from 2017 to 2019.

Response: We have added this into our manuscript.

The likely reason for increase in PM2.5 levels is the associated increase in emissions from mobile sources in recent years (Milando et al., 2016).

L89 typo 'foe;d'

Response: We thank reviewer for pointing out this typo, we have fixed it in the manuscript.

L110 the thermal conditioning within EPA measurements can impact upon particle mass through the removal of volatiles and this should be highlighted.

Response: We have added this information into manuscript

Additionally, temperature and relative humidity can alter particle physical and optical properties that PA measurements are sensitive to. While EPA measurements will also be affected by these air properties, the impact is lower because of thermal and humidity conditioning of samples prior to measurements.

L162 define 'stationarity'.

Response: We have defined the stationarity into manuscript.

Figure 1 – provide units on population density.

Response: We have added the units of population density in figure as well as in the caption.

Figure 2 – define the coloured lines on top of the E2 and P6 time plots.

Response: We have added colored line on top of figure.