

Atmos. Meas. Tech. Discuss., author comment AC2
<https://doi.org/10.5194/amt-2021-154-AC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.



Reply on RC2

Jeffrey Bean

Author comment on "Evaluation methods for low-cost particulate matter sensors" by
Jeffrey K. Bean, Atmos. Meas. Tech. Discuss.,
<https://doi.org/10.5194/amt-2021-154-AC2>, 2021

The author thanks the reviewer for the comments. The manuscript has been updated and more information on the updates is provided below.

General Comments:

In this work, the author examines the impact of controllable factors, such as averaging time and type of reference instrument, on performance metrics used to evaluate low-cost sensors. In addition, the author demonstrates that pairing two sensors together with a data agreement requirement can act as an easy quality assurance check. Also, the author proposes and tests using a prediction interval as a method to evaluate low-cost sensor performance. Both of which would enhance our current methods of evaluating and comparing low-cost particulate matter sensors. However, the methods used in this paper are not well described and further work could be done to strengthen some of the conclusions.

Major:

Line 31: The EPA recently published a report on performance testing protocols, metrics, and target values for PM_{2.5} low-cost air sensors. In this report they recommend using various indicators to evaluate sensor performance and offer performance target values for those indicators. This report may better address the concerns of using R² and RMSE to evaluate sensor performance in addition to other performance metrics.

Duvall, R., A. Clements, G. Hagler, A. Kamal, Vasu Kilaru, L. Goodman, S. Frederick, K. Johnson Barkjohn, I. VonWald, D. Greene, AND T. Dye. Performance Testing Protocols, Metrics, and Target Values for Fine Particulate Matter Air Sensors: Use in Ambient, Outdoor, Fixed Site, Non-Regulatory Supplemental and Informational Monitoring Applications. U.S. EPA Office of Research and Development, Washington, DC, EPA/600/R-20/280, 2021.

Response: This newer work from the EPA has now been included in the discussion. This newer work helps to address some of the issues with standardization of evaluations, though there is a continued need to discuss whether R² and RMSE are the best metrics for evaluations.

Line 66: Multiple times in this paper agricultural burning emissions are listed as the cause of the high PM events, what other evidence can be provided to confirm these burning events? Do particles generated by these events have different optical properties than those present in ambient air?

Response: These agricultural burning events are very common in this area and plumes of smoke can often be seen rising from fields while driving around in the area. Further downwind, the plumes become less distinct and become a broad haze. The particles likely have unique properties, but identification properties were not measured in this work. While it would be nice to have a way to express more certainty about the origins of these high concentration events, the origin of these events is not crucial to the analysis in this manuscript.

Line 72: "...while the T640 uses an optical counting method that is more similar to the method used by the low-cost particulate matter sensors." Please provide a little more detail on the similarities and differences between the T640 and the low-cost PM sensors used in this study.

Response: More details have been added to clarify the key similarities and differences: "The T640 is an optical particle counter, which dries and then counts individual particles. It differs from the evaluated low-cost sensors, which take a nephelometric measurement undried, bulk particle concentrations. However, the T640 is still an optical measurement that and is more similar to the method used by low-cost particulate matter sensors."

Line 73: "The BAM was used throughout the entire period of evaluation but often struggled to maintain sample relative humidity below 35%, which is a FEM requirement. Any data which did not meet this criterion was removed prior to analysis." Please clarify which data was removed, is this the cause of the gaps in Figure 2E?

Response: Any data for which the internal RH measurement of the BAM exceeded 35% was removed. This is the cause of all gaps in Figure 2E and a note explaining this has been added to the caption for that figure.

Line 78: More information should be provided on the initial testing of the 4 low-cost PM sensors of different brands. Why is this one-month test indicative of how the sensors will perform over the ten-month test? Why was only R2 used to determine the best-performing brand?

Response: It has now been clarified in the manuscript that a definitive identification of the "best" sensor from this one-month trial was not crucial to the study. The purpose of the one-month trial was to identify a useful sensor that could be explored with additional analysis, as described in the rest of the manuscript.

Line 79: It would be helpful to state the brands of the low-cost PM sensors in addition to the OEM of the optical sensors inside the devices. Additionally, more information such as sampling/averaging time of these sensors should be provided.

Response: The sensor models are withheld in part to allow focus on evaluation methods, rather than specific sensors, but models are also withheld to avoid giving the impression that the author endorses or disparages one sensor company over another. Additional details on sensor measurements have been added to the manuscript, as requested, but the specific makes and models are still withheld for the reasons stated.

Line 113: "It is noteworthy that sensor measurements correlated so closely from one sensor to another (Figure 2A) despite such a large range of variation from reference measurements." Figure 2A only shows the correlation for 2 sensors, is this same trend

seen when comparing all 8 sensors? Did all sensors have the same response to the environmental conditions?

Response: For simplicity only the comparison between 2 sensors is shown but the result was similar for other combinations of sensors. This has been clarified in the manuscript: "Comparison between only two sensors is shown in Fig. 2A for simplicity, but similarly strong correlation was observed for other pair of sensors."

Line 218: The paper states that the prediction interval can be used to evaluate/compare multiple sensors, however data is only provided for 1 sensor when 8 were tested. It would be interesting to examine the PI between all 8 of the same brand and even compare the prediction intervals of the 4 brands initially tested. Including the results of these tests would strengthen the argument to include prediction interval as a performance evaluation metric for low-cost PM sensors.

Response: This is a good suggestion. Unfortunately, there was an insufficient range of concentrations during the 1 month of initial testing to produce a meaningful prediction interval. However a comparison of uncertainty between other sensor replicas at 35 $\mu\text{g}/\text{m}^3$ has now been included (line 260).

Line 280: "The question remains on how much distance can be allowed between the reference sensor and the network sensors before this method fails." The EPA recommends mounting sensors within 20m horizontal from the FRM/FEM monitor. See above citation which also includes recommendations on setting up low-cost sensors at test sites.

Response: This statement was referring to sensors that are deployed in the field (away from a reference) but still calibrated using a sensor that is collocated with the reference. It has been clarified in the manuscript to indicate this.

Minor:

Line 98: "As more stringent data agreement requirements are put in place (moving to the right in Fig. 1) there are not significant improvements in correlation." Change not to no.

Line 107: "The bottom figure shows how R2 between the sensor pair (average of the pair) and reference measurement changes as when these disagreeing data points are removed." Reword this sentence in the Figure 1 caption. Suggestion: "The bottom figure shows how the R2 between the sensor pair (average of the pair) and reference measurements changes when disagreeing data points are removed."

Line 273: "Figure 6 suggests that this method will have mixed results if calibrating over short time periods but is reliable if enough time is allowed to capture all variations in slope, concentration, and residual standard error." Reword second part of this sentence. Suggestion: "...but can be reliable given enough time to capture all variations in slope, concentration, and residual standard error."

Response: These minor suggestions have all been implemented and the author thanks the reviewer for pointing them out.