Reviewer comments for:

Evaluation of Calibration Performance of a Low-cost Particulate Matter Sensor Using Colocated and Distant NO_2

This review is for the above manuscript submitted for publication in Atmospheric Measurement Techniques. The manuscript uses co-located and regulatory monitoring-based measurements to build calibration models for low-cost $PM_{2.5}$ sensors. The authors test several combinations of four variables such as PM2.5 measurements, temperature, relative humidity, and NO_2, where the NO_2 measurements are from either collocated or nearby instrumentation. The authors conclude that collocated or nearby NO_2 measurements should be used for calibration models since model performance improves in terms of statistical measures such as RMSE, MAE, and R^2. However, in its current form, the authors' comparisons are very limited in scope (single sensor), lack appropriate choice of performance parameters (e.g., adjusted R^2), show very limited performance improvement (~5%), and are marred by lack of uncertainty analysis and poor presentation of methods and results. **Reviewer 1 has discussed the limitations of a single sensor so I will focus on other aspects. I recommend that the authors significantly revise and resubmit this manuscript for further consideration.**

1. **Very limited in scope and performance improvement:** Given that the authors focused the comparison on calibration of a single sensor, the weight of "substantial contribution" of this manuscript falls on performance improvements of calibration models associated with that sensor. Unfortunately, the improvements on inclusion of NO_2 are quite minimal. For example, in Tables 3 and 4, the best performances of models with and without NO_2 are ~5% of each other. Does that qualify this work as "represent(ing) a substantial contribution to scientific progress" as is required by AMT? I disagree. **I suggest that the authors conduct the analysis for the excluded sensor (sensor #8) that otherwise passes all checks, but was not included in the analysis for an unknown reason, as also pointed by reviewer 1**.

2. Choice of performance parameters and lack of uncertainty analysis: While the authors include three performance measures, despite considering models with multiple and changing number of variables, the authors fail to include the most important one: adjusted R^2. The authors have clearly used the multiple R^2 squared value to compare model fits; however, multiple R^2 will increase on addition of even poorly correlated variables. I suggest that the authors report adjusted R^2 results. Additionally, presentation of such calibration results would also benefit from an uncertainty analysis, and a key manuscript cited by the authors uses bootstrapping to do just that (Hua et al., 2021). I strongly recommend uncertainty (in terms of standard deviation) be considered when presenting performance metrics associated with such comparisons. **The authors can then answer the**

question: are the distributions of performance parameters statistically significantly different with or without NO_2? I would consider answering that question as a significant contribution.

3. Poor presentation: Large sections of the manuscript are unnecessarily detailed, and could be moved into tabular form whether in the main manuscript or the supplement. These include large portion of the lines 198-222 and 233-246 which are two representative examples. Additional examples include lists of variables shown in text format, which is laborious to read or keep track of (e.g., Lines 355-364). Additionally, key details of the authors' methodology such as performance metrics and intercomparison exercises are dispersed throughout the Results section (Sect. 3.1 to 3.6). **I suggest that authors separate the methods portions of these results and discuss them in a separate subsection under Methods called "Instrument intercomparisons".**

Minor comments

1. Lines 123-139 The authors start off with a large dataset but remove data points using some filters. I suggest that the authors add a supplementary table showing how many data points were removed at each step.

2. Lines 412-413 and Lines 287-290 The language used by the authors is unclear. I suggest either expanding on these sentences or rephrasing them so that the point is made clearly.

1. Hua, J., Zhang, Y., de Foy, B., Mei, X., Shang, J., Zhang, Y., Sulaymon, I. D., & Zhou, D. (2021). Improved PM2.5 concentration estimates from low-cost sensors using calibration models categorized by relative humidity. *Aerosol Science and Technology*, *55*(5), 600–613. https://doi.org/10.1080/02786826.2021.1873911